IRM | institut du risk management BY Institut.

Mémoire présenté pour la validation de la Formation
« Certificat d'Expertise Actuarielle »
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le

Par : Camille VIGNON et Dmitri BORISSOV

Titre : Etude ALM d'un portefeuille retraite avec une approche par reinforcement learning

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1an ☐ 2 ans)
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de l'Institut des actuaires :

Membres présents du jury de l'Institut du Risk Management :

Secrétariat :

Bibliothèque :

Entreprise : AXA France

Nom :

Signature et Cachet :

Directeur de mémoire en entreprise :

Nom : Karim CHELLALI

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

Signature(s) du candidat(s)

Résumé

Les réglementations récentes, comme Solvabilité 2 ou bientôt IFRS 17, exigent des assureurs un suivi prospectif de leurs portefeuilles en vision économique. Pour cela, les modèles utilisés intègrent des stratégies de gestion actif/passif leur permettant d'optimiser les bénéfices futurs sous contrainte de solvabilité. Ces stratégies sont généralement figées et issues de jugements d'experts ou bien calibrées dans un model *ad-hoc*. Le *reinforcement learning* est un domaine de l'apprentissage automatique s'attaquant à des problèmes de prise de décision en interaction avec un environnement. Un agent y apprend de son expérience grâce à un conditionnement par des récompenses et des pénalités. Cette approche est particulièrement efficace dans le domaine des jeux ou de la conduite autonome. Mais elle peut également être appliquée à des modèles assurantiels pour mettre en place une stratégie d'allocation d'actifs performante. Dans ce mémoire, nous proposons de mettre en place et étudier portefeuille de contrats de retraite en run-off dans un cadre de *deep learning*. L'objectif est d'entrainer un agent prenant des décisions d'allocation d'actifs grâce à un réseau de neurones. Les résultats montrent qu'un tel agent obtient des taux de rendement sur capitaux propres plus importants et plus stables que des stratégies classiques. Ainsi, le *reinforcement learning* offre un champ large d'application pour les assureurs, allant de la réalisation d'études ALM à l'audit de modèle.

Mots-clés : *reinforcement learning*, *réseau de neurones*, *assurance retraite collective*, *ALM*.

Abstract

Recent regulations in the field of insurance such as Solvency II or IFRS 17 require a prospective monitoring of insurance portfolios with an economic view. It is achieved by using complex models integrating decision-making modules to implement ALM strategies that maximize the return on equity under solvency constraints. In general, these strategies as based on expert judgement or calibrated in *ad-hoc* models. Reinforcement learning is a machine learning approach tackling problems with interactions between decisions and an environment. An agent learns from its experience thanks to conditioning with rewards and penalties. This approach is particularly successful in training AI agents playing video games or for autonomous driving. But it can also be applied to create strategies for efficient asset allocation in insurance. Our paper aims to study a run-off group pension portfolio in a deep learning framework. The goal is to train a neural-network-based agent to take investment decisions to optimize shareholder's return on equity. Results show that such an agent is able to over-perform traditional ALM strategies in terms of both risk and return. This approach can be extended to models involved in business decisions and offer a vast range of applications from ALM studies to model validation.

Keywords : *reinforcement learning*, *neural network*, *group pension insurance*, *ALM*.

Note de Synthèse

La réglementation Solvabilité 2 puis l'entrée en vigueur très prochaine de la norme IFRS 17 imposent aux assureurs une valorisation régulière de leur portefeuille dans des délais généralement très courts et avec un besoin de justification toujours plus accru. Dans ce contexte, les métriques de rentabilité future ou de coût en capital (respectivement VIF (Value of Inforce) et SCR (Solvency Capital Requirement)) sont devenues incontournables dans les études ALM pour piloter le niveau de risque acceptable dans les portefeuilles d'actifs. Le calibrage des allocations d'actifs futures pour le plan d'investissement de l'entreprise, puis exploité sur 60 ans dans les modèles de projection actuariels vie s'effectue généralement via une méthode nécessitant un nombre important de simulations pour la recherche d'une frontière efficiente du couple (rendement, risque) du portefeuille. Ce mémoire propose une approche d'apprentissage automatique sur un modèle simplifié de retraite collective : le modèle devient capable après la phase d'apprentissage de choisir directement l'allocation à chaque scénario et chaque pas de temps en fonction de l'état des lieux du portefeuille.

Le *reinforcement learning* (RL) est un domaine de l'apprentissage automatique s'appliquant à des problèmes de prise de décision d'un agent en interaction avec un environnement dans le but de maximiser un objectif long-terme. La stratégie de l'agent s'améliore au fil des itérations grâce à un conditionnement par des récompenses et des pénalités déterminées en fonction des conséquences des actions prises. Par exemple, il est possible d'entrainer dans ce cadre une intelligence artificielle à jouer à maximiser son score dans un jeu comme Tetris. Le *deep reinforcement learning* est une sous-branche du RL où l'agent s'appuie sur un réseau de neurones pour attribuer un score à chacune des actions possibles et l'utilise dans sa décision. Ce domaine a acquis une notoriété auprès du grand public en 2015 lorsque l'algorithme AlphaGo développé par Google/DeepMind est devenu le premier algorithme à battre un joueur de Go professionnel. La gestion d'un portefeuille de retraite en *run-off* peut être formalisée comme un problème de RL. Le modèle projetant le bilan et le résultat de l'assureur est l'environnement dans lequel un agent allocataire d'actifs effectue des transactions. Ces actions sont choisies pour maximiser la création de valeur pour l'actionnaire et son comportement s'adapte au portefeuille ainsi qu'au contexte économique.

Nous avons donc choisi d'étudier un portefeuille simplifié de 13Md€ de passif réparti entre des contrats en phase de constitution (9Md€) et des contrats en phase de restitution (4Md€). En face, l'actif est composé d'OAT (10Md€), d'actions (1Md€), d'immobilier (1Md€) et de cash (1Md€) avec un taux de plus-values global de 49%. Un SCR brut total de 860M€ est associé à ce portefeuille. Les scénarios économiques *real world* du modèle interne d'AXA France ont été utilisés pour projeter l'actif. L'agent prend en entrée un vecteur de dimension 13 contenant l'allocation d'actifs actuelle, la duration, les différentes composantes du SCR, le taux de richesse, le taux de produits financier et le taux de marge du dernier exercice. A partir de ces éléments, un réseau de neurones donne un score à chacune des 25 actions possibles permettant d'acheter ou vendre un ou plusieurs actifs. Voir

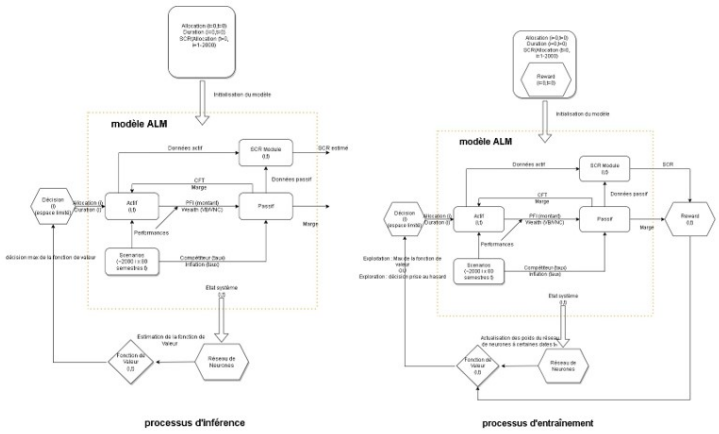


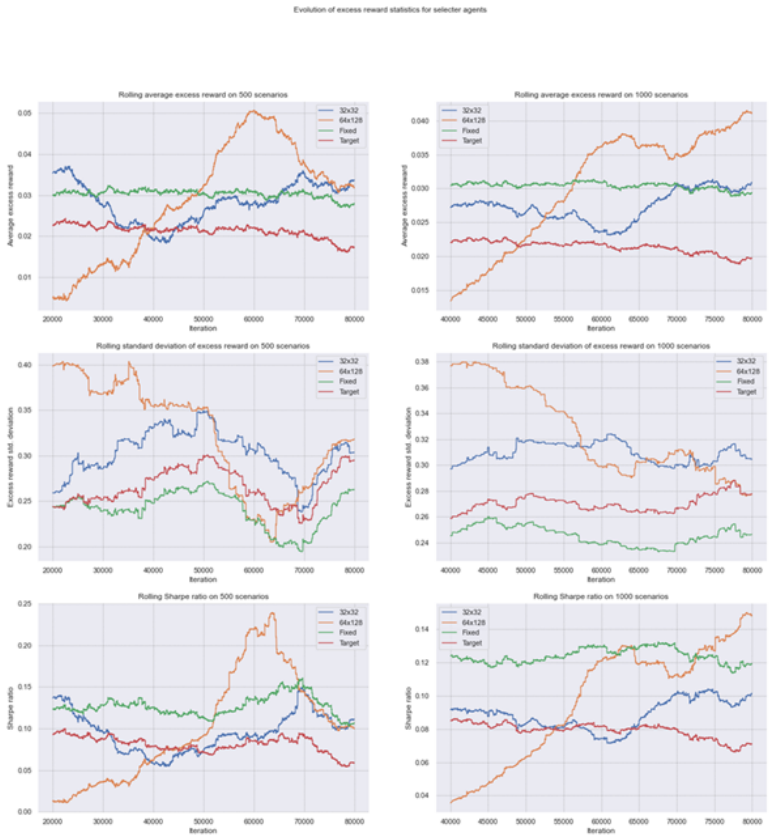
FIGURE 1 – Deep ALM - Schéma fonctionnel en phases d'entraînement et d'inférence

figure 1.

En nous appuyant sur des résultats théoriques et des heuristiques, nous avons choisi d'explorer des architectures de réseaux de neurones à deux couches cachées avec chacune un nombre de neurones allant de 16 à 128. Le réseau 64x128 s'est avéré être le plus performant en étant sur la phase d'entrainement meilleur que la stratégie d'allocation cible comme le montrent les résultats du graphe 2. Ces résultats ont été confirmés en phase de test : nous avons entrainé l'agent sur 90% des scénarios et mesuré sa performance sur les 10% restants sans entrainement. On observe une surperformance par rapport à la stratégie de référence et par rapport à l'agent en phase de test à niveau d'écart-type stable. Sur les scénarios de test, l'agent profite des performances importantes des actifs risqués pour réaliser des plus-values qui augmentent les produits financiers et par conséquent la marge, via la rétention, et la satisfaction client qui est une source de marge future. Ces résultats sont très encourageants et peuvent être encore largement améliorés. Premièrement en calibrant mieux les hyperparamètres de l'agent tels que le taux d'actualisation ou la stratégie de *dropout*. Le *reward* peut être enrichi grâce à des pénalités en cas d'insolvabilité voir de faillite de l'assureur. Enfin, l'environnement doit être complexifié avec l'ajout de provisions permettant d'effectuer du pilotage plus fin des taux servis ou encore grâce à des classes d'actifs supplémentaires.

Cette étude possède plusieurs champs d'applications dans le domaine de l'assurance. Il est possible d'utiliser un modèle de RL pour produire des études ALM en s'abstrayant des préconceptions sur les stratégies d'allocation d'actif et en explorant l'ensemble des actions possibles. Il serait également intéressant de brancher un tel agent apprenant à un modèle interne. D'une part cela permettrait potentiellement d'améliorer la valeur prospective du portefeuille de l'assureur. D'autre part, il est possible d'utiliser l'agent pour auditer le modèle. En effet, avec un nombre suffisant d'itérations, l'agent est capable de tirer profit d'opportunités d'arbitrage qui pourraient exister et les mettrait donc en évidence.

Malgré les perspectives de performance qu'offrent les modèles de *deep reinforcement learning*, leur manque de transparence est un frein à leur adoption pour produire de indicateurs réglementaires tels que le ratio de solvabilité. La validation de ce type de modèles sera un enjeu majeur pour les régulateurs qui doivent explorer de nouveaux paradigmes de contrôle.



Comparatif des performances d'entrainement pour les agents suivants :

- "Fixed" : allocation d'actif fixe
- "Target" : actions prises pour s'approcher d'une allocation d'actifs cible
- "32x32" : réseau de neurones 32x32 avec 20% de taux d'exploration
- "64x128" " : réseau de neurones 32x32 avec 20% de taux d'exploration

Les indicateurs de moyenne, d'écart-type de ratio de Sharpe sont calculés pour la surperformance par rapport à des décisions aléatoires sur une fenêtre glissante de 500 et 1000 itérations.

FIGURE 2 – Déroulé d'une simulation

Synthesis note

The Solvency 2 regulations and the forthcoming implementation of IFRS 17 require insurers to regularly value their portfolios within very short timeframes and with an ever-increasing need for justification. In this context, future profitability or cost of capitl metrics (respectively VIF (Value of Inforce) and SCR (Solvency Capital Requirement)) have become essential in ALM studies for steering the acceptable level of risk in asset portfolios. The calibration of future asset allocations for the company's investment plan, then used over 60 years in life actuarial projection models, is generally carried out via a method requiring a large number of simulations to find an efficient frontier for the portfolio's risk/return trade-off. This thesis proposes a deep learning approach on a simplified group pension model : after the learning phase, the model is able to directly choose the allocation for each scenario and each time step according to the state of the portfolio.

Reinforcement learning (RL) is a field of machine learning applied to decision-making problems where an agent interacts with an environment in order to maximize a long-term objective. The agent's strategy improves with each iteration thanks to conditioning by rewards and penalties determined according to the consequences previous actions. For instance, it is possible to train in this framework an artificial intelligence to play and maximize its score in a game such as Tetris. Deep reinforcement learning is a sub-branch of RL where the agent relies on a neural network to score each possible action which is used in the decision. This field gained mainstream notoriety in 2015 when AlphaGo, an algorithm developed by Google DeepMind, became the first algorithm to beat a professional Go player.

The management of a run-off pension portfolio can be seen from the RL perspective. The model projecting the balance sheet and the income statement of the insurer is an environment in which an asset allocating agent carries out transactions. Its actions are chosen to maximize value creation for the shareholder through a behavior adapted to the portfolio and the economic context.

We have therefore chosen to study a simplified portfolio of €13 billion in liabilities divided between contracts in the accumulation phase (€9 billion) and contracts in annuity phase (€4 billion). The assets are a mix of French govies (€10bn), equity (€1bn), real estate (€1bn) and cash (€1bn) with an overall unrealized capital gains rate of 49% . A total gross SCR of €860m is associated with this portfolio. Real world economic scenarios from AXA France's internal model were used to project the asset. The agent takes as input a vector of size 13 containing the current asset allocation, duration, SCR components, URGL rate, GA rates and the margin rate for the previous financial year. Based on this data, a neural network gives a score to each of the 25 possible actions allowing trade one or several assets.Voir figure 3.

Based on theoretical results and heuristics, we chose to explore neural network architectures

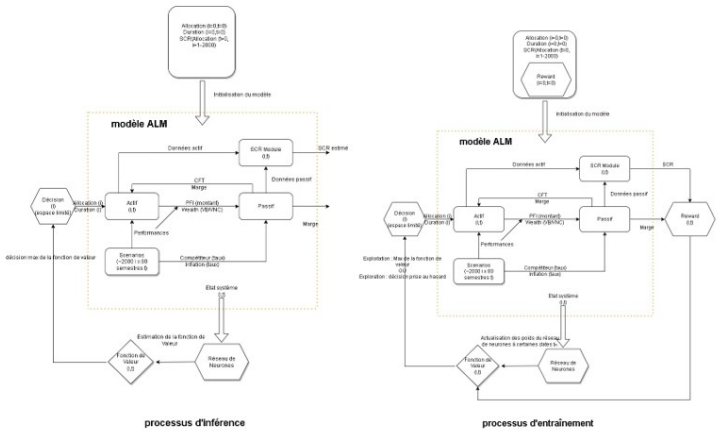
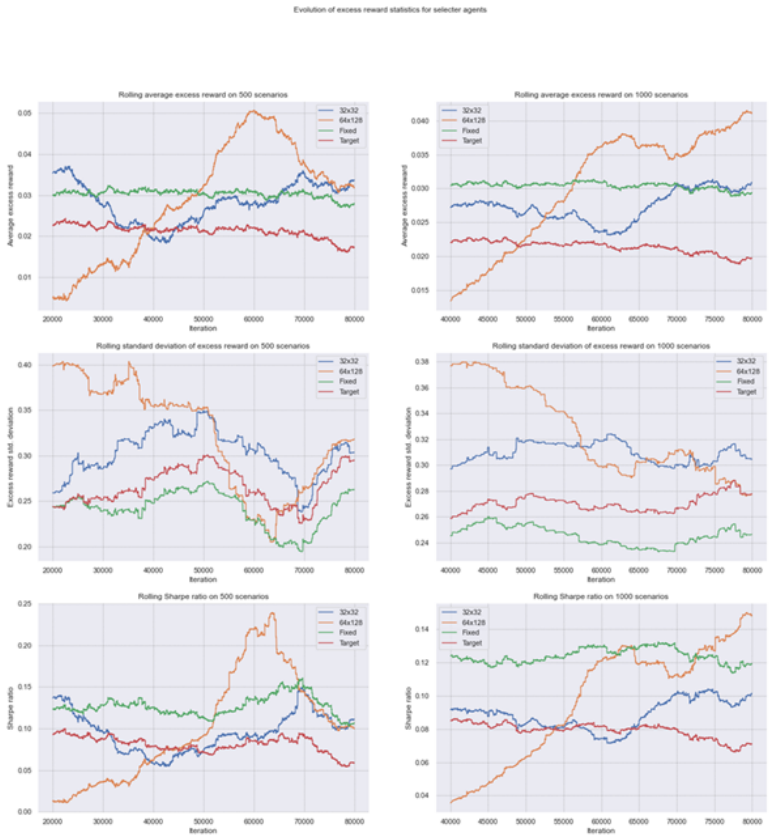


Figure 3 – Deep ALM - Schéma fonctionnel en phases d'entraînement et d'inférence

with two hidden layers containing between 16 and 128 neurons. The 64x128 network proved to be the most efficient on the training phase outperforming our benchmark (target allocation strategy) as shown on the figure 4. These results were confirmed during the test phase: we trained the agent on 90% of the scenarios and measured its performance on the remaining 10% without training. We observed an out-performance compared to the reference strategy and compared to the agent in the test phase at similar standard deviation levels. In the test scenarios, the agent takes advantage of the significant performance of the risky assets to realize capital gains which increase the GA rates and consequently the margin, via retention, and customer satisfaction which is a source of future margins. These results are very encouraging but can still be greatly improved. First, by optimizing hyper-parameters such as the discount rate or the dropout strategy. The reward can be enriched thanks to penalties in the event of insolvency or even bankruptcy of the insurer. Finally, the environment must complexified by adding reserves allowing GA rates steering or by adding asset classes.

This study has multiple applications across the field of insurance. It is possible to use an RL model to produce ALM studies without preconceptions about asset allocation strategies. It would also be interesting to connect a deep learning agent to an internal model. On the one hand, it would potentially improve the prospective value for the shareholder. On the other hand, it is possible to use the agent to audit the model. Indeed, with enough iterations, the agent can take advantage of arbitrage opportunities that may exist and therefore highlight them. Despite the prospects offered by deep reinforcement learning models, their lack of transparency is a hindrance for their adoption to produce regulated indicators such as the solvency ratio. Reviewing this kind of models will be a major challenge for regulators who must explore new validation paradigms.



Comparatif des performances d'entrainement pour les agents suivants :

- "Fixed" : allocation d'actif fixe
- "Target" : actions prises pour s'approcher d'une allocation d'actifs cible
- "32x32" : réseau de neurones 32x32 avec 20% de taux d'exploration
- "64x128" " : réseau de neurones 32x32 avec 20% de taux d'exploration

Les indicateurs de moyenne, d'écart-type de ratio de Sharpe sont calculés pour la surperformance par rapport à des décisions aléatoires sur une fenêtre glissante de 500 et 1000 itérations.

Figure 4 – Déroulé d'une simulation

Remerciements

Nous tenons à remercier toutes les personnes qui nous ont permis de faire aboutir ce projet qui nous a passionné. Plus particulièrement les personnes du Risk Management ou de la Direction des Investissements d'AXA France avec qui nous avons pu échanger des conseils techniques ou qui ont pris le temps de relire attentivement ces pages : Karim Chellali, Nicolas Eyrolle, Arnaud At, Julio Thelusca, Arthur Maçon, Jean Bergot et Alban Davand.

Nous remercions également nos familles pour leur soutien incondionnel grâce à qui nous avons consacré le temps nécessaire à la réalisation de ce mémoire.

Table des matières

Introduction

La réglementation Solvabilité 2 puis l'entrée en vigueur très prochaine de la norme IFRS 17 imposent aux assureurs une valorisation régulière de leur portefeuille dans des délais généralement très courts et avec un besoin de justification toujours plus accru. Les modèles de projection actuariels vie doivent être capables de simuler sur le long-terme les résultats des portefeuilles de passif avec les interactions ALM liées, afin de capturer tous les risques afférents aux garanties techniques ou financières sous-jacentes aux contrats.

Traditionnellement, ces modèles fonctionnent avec un nombre important de paramètres prédéfinis et justifiés par des études préalables. En particulier, à l'actif, le modèle doit prendre des décisions de réallocation à chaque pas de temps, en tenant compte du vieillissement des portefeuilles actif et passif, mais aussi du pilotage financier souhaité par l'entreprise. Des algorithmes existent dans les modèles pour réaliser les décisions estimées du management sur une cible de produits financiers sous contrainte de marge ou de taux de participation aux bénéfices à servir pouvant générer un comportement dynamique des assurés. En complément, ces modèles prévoient également des cibles d'allocations d'actifs figées dans le temps pour le choix des différentes classes d'actifs qui composent le portefeuille au fur et à mesure du temps, ce qui fait que le modèle ne s'adapte pas vraiment aux aléas futurs de marché, mais uniquement à une cible souhaitée à l'instant 0. Par ailleurs, ils font l'objet d'études ALM préalables coûteuses en temps homme et machine car nécessitant un nombre important de simulations.

Dans le domaine de l'intelligence artificielle, les algorithmes de *machine learning* se sont révélés performants pour aider à la prise de décision, de manière semi-automatique ou automatique dans des situations complexes. Nous avons voulu explorer ce domaine pour dynamiser et automatiser le choix des allocations d'actif optimales. Pour cela, nous avons implanté dans notre modèle un algorithme de *deep learning* entraîné par une méthode d'apprentissage par renforcement.

Nous avons choisi de réaliser cette étude sur un portefeuille de retraite simplifié, ce qui nous a permis d'explorer les possibilités offertes par cette nouvelle approche. Pour entraîner le système, nous valorisons le portefeuille pendant 60 ans en partant d'une allocation aléatoire et en définissant un stimulus, le *reward*, qui permet de récompenser ou au contraire pénaliser le système si l'allocation choisie est favorable ou non. Une fois entraîné sur un nombre élevé de scenarios d'actifs aléatoires, le système permet à notre modèle de réaliser lui-même les choix d'allocation d'actif en fonction de l'état du système qu'il observe à chaque pas de temps et chaque scenario.

Nous allons décrire par la suite de ce mémoire le contexte de la retraite en assurance vie et le cadre de modélisation associé, puis la mise en œuvre concrète de notre modèle avec les simulations réalisés et résultats observés et enfin les possibilités d'application résultantes.

Chapitre 1

La Retraite Collective en France sous Solvabilité II

1.1 Enjeux de la retraite en France

La France compte 16,7M de retraités de droit direct à fin 2019, d'après le rapport de la DREES dans son édition 2021 sur les retraités et les retraites.

Ce nombre reste depuis des années en augmentation, avec un taux d'évolution de 1,8% entre 2018 et 2019 du fait d'un nombre de nouveaux retraités plus important que celui des retraités décédés comme le présente la figure 1.1 issue d'un rapport de la DREES sur les retraites, dirigé par F. Arnaud (2021).

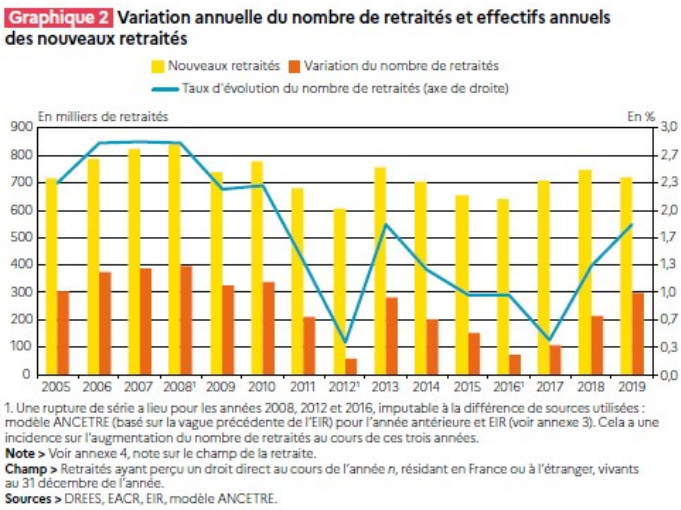


FIGURE 1.1 – Variation annuelle du nombre de retraités – Source : Les retraités et les retraites - Edition 2021 - DREES

Le contexte de la crise du Covid de 2020-2021 qui a touché de manière plus forte les personnes âgées pourrait ralentir ce constat mais son impact devrait rester modéré même si la crise perdurait encore compte tenu de l'ensemble des mesures de protection mises en place par l'état comme

le montrent les chiffres provisoires disponibles. L'espérance de vie à 65 ans mesurée de manière provisoire (pour les années 2019-2021) par l'Ined montre une légère inflexion en 2020 avec un niveau qui revient sur celui de 2015 pour les femmes et 2012 pour les hommes (voir figure 1.2). Il s'agit pour les femmes de 65 ans de 23 ans d'espérance de vie en 2015 et en 2020, après avoir atteint un pic à 23,4 ans en 2019. Le constat est semblable pour les hommes de 65 ans dont l'espérance de vie en 2020 est estimée à 18,9 ans contre 18,8 ans en 2012, après avoir atteint un pic à 19,6 ans en 2019.

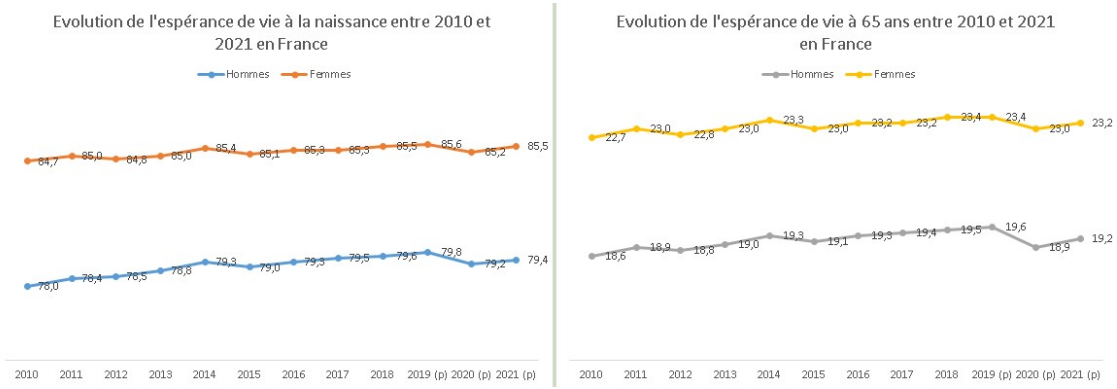


FIGURE 1.2 – Espérance de vie en France entre 2010 et 2021 – (p) résultats provisoires à fin 2021 - Source : Insee, statistiques de l'état civil et estimations de population - https ://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-cause-deces/esperance-vie/

Les enjeux actuels et futurs autour de ce sujet restant élevés, le gouvernement en place a posé fin 2021 les premières réflexions d'une nouvelle réforme des retraites pour préparer l'opinion publique à de nouvelles mesures.

En l'état actuel des choses, le rapport entre le nombre de personnes actives et les personnes retraitées est passé de 2,02 en 2004 à 1,7 en 2019 (voir figure 1.3). Dans un régime de retraite par répartition, qui est le principal régime de retraite en France, cette baisse a une influence négative sur l'équilibre budgétaire.

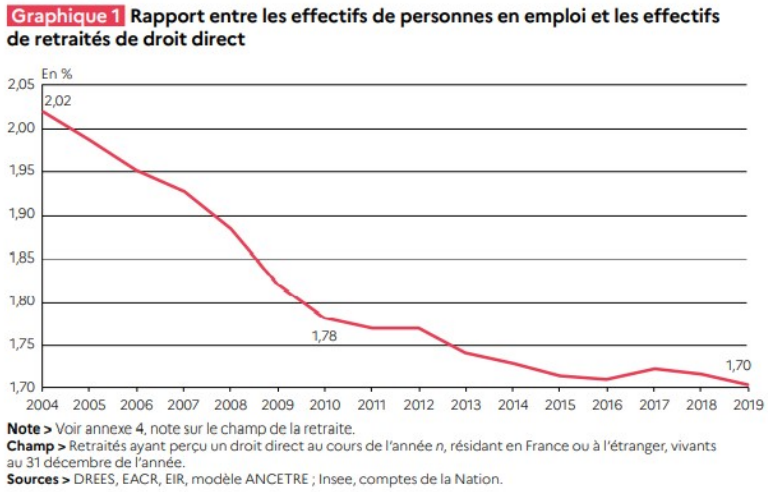


FIGURE 1.3 – Ratio entre le nombre d'actifs et de retraités - Source : Les retraités et les retraites - Edition 2021 - DREES

Il est clair que la place de la retraite dans l'économie française, importante actuellement, ne devrait cesser d'augmenter.

1.2 Le mécanisme général français de retraite

1.2.1 Principes généraux

Le système de retraite en France est constitué de 3 niveaux, mis en place progressivement depuis 1945 :

- 1er Pilier : Régime obligatoire de base (par répartition)

- 2ème Pilier : Régime complémentaire obligatoire (par répartition)

- 3ème Pilier : Régimes supplémentaires d'entreprise ou individuels (par capitalisation).

Voir figure 1.4

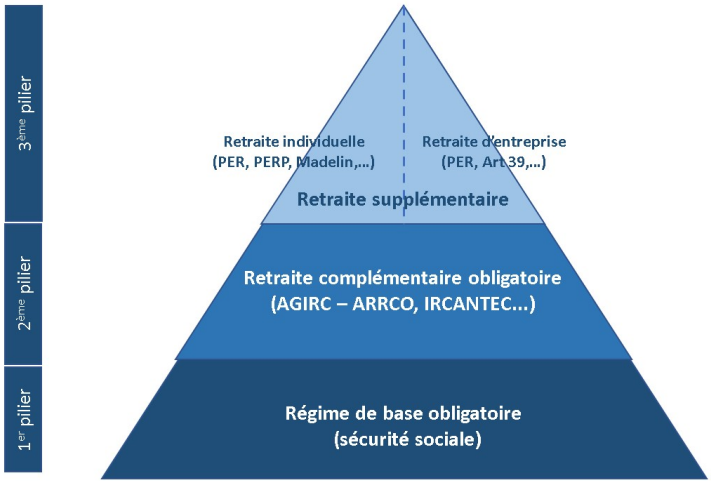


FIGURE 1.4 – Les 3 piliers schématiques du système de retraite française

D'autres placements peuvent aussi constituer des sources de financement des personnes retraitées, comme l'assurance-vie (épargne), l'épargne salariale, ou des investissements immobiliers (pierre ou pierre-papier). Ces différents montages financiers permettent soit des revenus réguliers (investissements immobiliers par exemple avec des loyers à percevoir), soit des mécanismes de rachats partiels réguliers ou des options de sortie en rentes (assurance vie ou épargne salariale). L'investisseur intègre dans son choix les dispositions fiscales associées à chaque investissement et chaque mécanisme de sortie du montage pour optimiser son placement : le fisc français prévoit en effet un niveau d'imposition ou des avantages fiscaux à l'entrée ou à la sortie du placement qui sont spécifiques pour chaque type d'investissement.

1.2.2 Les régimes de base obligatoires

Régulés par l'Etat, les régimes de base obligatoires sont multiples (42 régimes différents en vigueur) et ont été mis en place progressivement pour des catégories ciblées de travailleurs. On distingue par exemple :

- Le régime général : salariés du secteur privé

- Les régimes des indépendants pour les non-salariés : commerçants, artisans, professions libérales

- Les régimes spéciaux : salariés du secteur public.

Ces régimes fonctionnent par répartition : les cotisations sont collectées chaque année auprès des actifs ; ces dernières permettent de financer les prestations délivrées la même année aux retraités.

L'équilibre de tels régimes est donc particulièrement dépendant du ratio entre le nombre d'actifs et de retraités déjà évoqué dans la section 1.1, mais également des facteurs suivants :

- la masse salariale des actifs et les taux de cotisation associés (par exemple le chômage vient diminuer les assiettes de cotisation)

- le nombre de retraités et le niveau des rentes.

Pour piloter l'équilibre de ces régimes, l'état joue en particulier sur l'âge de départ à la retraite, la durée minimale de cotisation, le niveau des rentes. Dans l'une des récentes réformes en 2010, l'âge légal de départ à la retraite a été fixé à 62 ans et le retraité ne peut bénéficier d'une retraite à taux plein qu'en partant à 67 ans. Elle a été complétée en 2014 par une durée minimale de cotisations de 172 trimestres (mise en place progressive).

1.2.3 Les régimes de retraite complémentaire obligatoires

Le second niveau regroupe les régimes complémentaires obligatoires, ayant vocation à compléter les régimes de base afin d'atteindre un niveau de retraite plus cohérent avec les revenus d'activité. Ceux-ci sont, comme les régimes de base, financés par répartition et légalement obligatoires. Leur particularité est qu'ils sont gérés en points de retraite et négociés avec les partenaires sociaux des conventions collectives de chaque secteur.

On parle de taux de remplacement pour désigner le ratio entre le montant annuel de la rente à la liquidation et le montant du dernier salaire annuel net perçu. Les régimes obligatoires complémentaires ont permis de rehausser ce taux. Toutefois ce dernier baisse au fil des générations.

Voir figure 1.5

Ce taux de remplacement est passé de 80,1% à 73,4% dans le secteur public entre la génération de 1938 et celle de 1950. Cette baisse a été légèrement moins marquée pour le secteur privé qui est passé de 78,3% à 74,7% pour les mêmes générations.

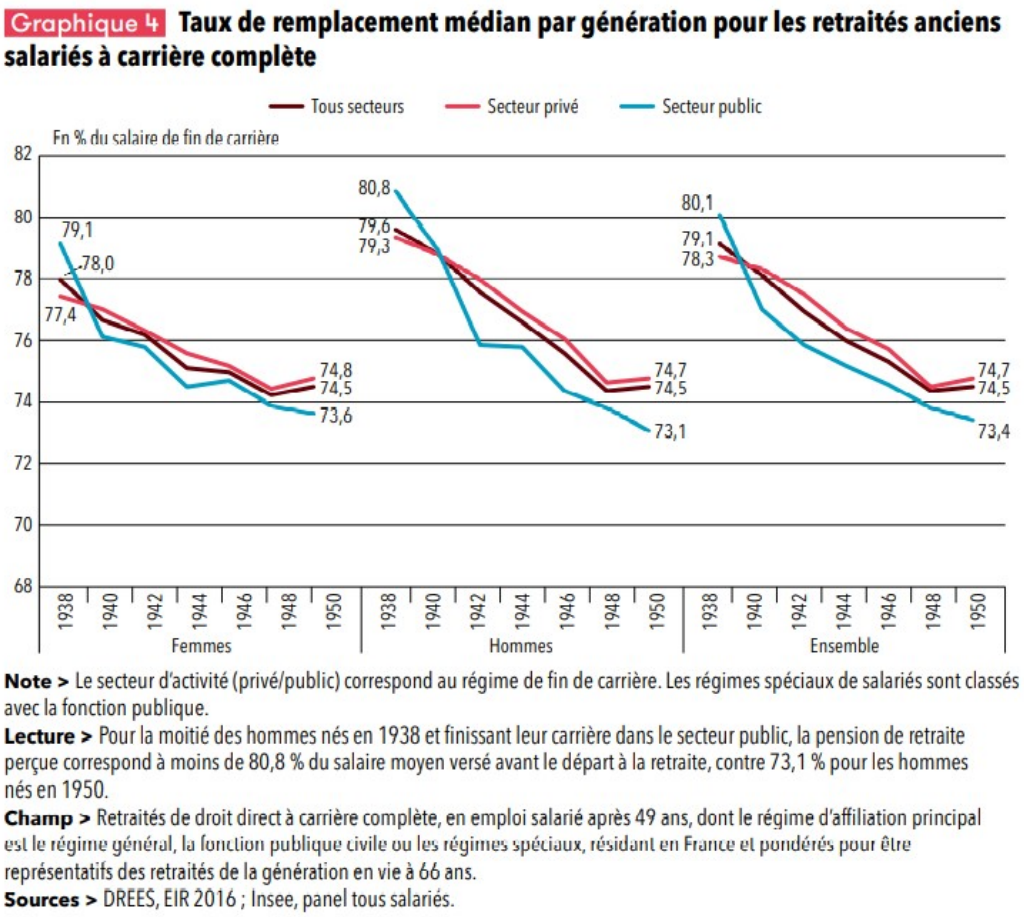


FIGURE 1.5 – Taux de remplacement par génération de femmes et d'hommes – Source : Les retraités et les retraites - Edition 2021 - DREES

En 2019, les deux caisses de retraite complémentaire des salariés du secteur privé, AGIRC et ARRCO, ont fusionné afin de mettre en commun leur réserves et d'harmoniser les règles de gestion. L'AGIRC (Association Générale des Institutions de Retraite Complémentaire des Cadres) était anciennement réservé aux cadres, tandis que l'ARRCO (Association pour le Régime de Retraite Complémentaire des Salariés) régissait la retraite de tous les salariés du privé.

1.2.4 Les dispositifs de retraite supplémentaire

Le troisième niveau regroupe deux grands types de contrats :

- Les couvertures facultatives et supplémentaires d'assurance retraite mises en place dans un cadre d'entreprise,

- Les couvertures facultatives et supplémentaires d'assurance vie individuelle dédiées aux Travailleurs Non Salariés (anciens dispositifs Madelin) ou à des personnes salariées (anciens dispositifs Perp, nouveaux dispositifs PER).

Les dispositifs de retraite supplémentaire reposent sur le principe de capitalisation et permettent aux personnes ayant la capacité d'épargner de se constituer un capital en vue de combler la baisse

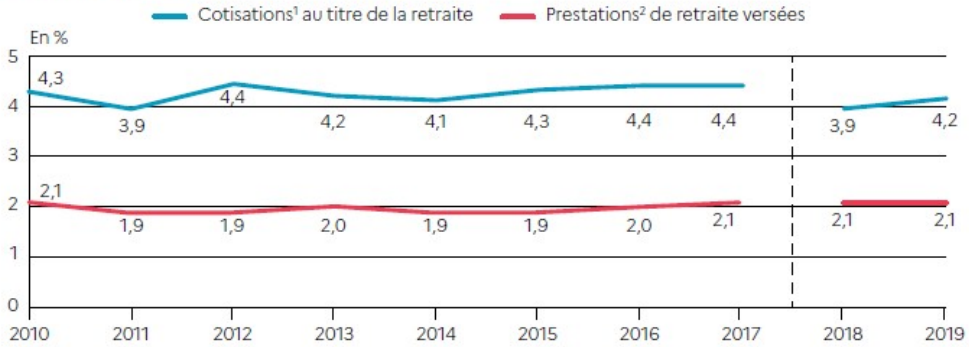
de leurs revenus au moment du passage à la retraite.

Ces dispositifs sont amenés à se développer, dans la mesure où le régime de retraite obligatoire ne pourra pas financer, au même niveau que les générations précédentes, la retraite des futures générations. En effet, la réglementation évolue régulièrement sur ce sujet, amenant les assureurs à proposer de nouveaux produits adaptés aux nouvelles dispositions réglementaires. Par exemple, le PER commercialisable à partir d'octobre 2019 est venu remplacer les différents produits usuels : PERP, Madelin et Article 83, qui ne sont plus commercialisables depuis octobre 2020.

Voir figure 1.6

Malgré les efforts réguliers de l'état pour mettre en avant ces régimes de retraite supplémentaire, ce graphique met en exergue que leur part dans le financement des retraites est encore très faible et reste stable depuis au moins 2010 avec seulement 4,2% des cotisations de l'ensemble des régimes de retraite affectées à une retraite supplémentaire en 2019, et 2,1% des prestations.

Graphique 2 **Part de la retraite supplémentaire dans l'ensemble des régimes de retraite (obligatoire et facultative)**



1. Cotisations sociales à la charge des employeurs et des salariés, contributions publiques, transferts pris en charge par le FSV (Fonds de solidarité vieillesse) rentrant dans le financement de la retraite. Données révisées en 2020.
2. Dans les prestations sont intégrées les pensions de retraite versées au titre des droits directs et dérivés, ainsi que les allocations du minimum vieillesse.
Note > Le champ de l'enquête retraite supplémentaire de la DREES étant exhaustif, les résultats ne sont plus calés sur les données des fédérations. Ceci conduit à une rupture de série entre 2017 et 2018, ainsi qu'à une révision de la masse des prestations versées pour l'année 2018, et donc de la part des prestations de retraite supplémentaire de 0,3 point (de 2,4 % à 2,1 %) [voir encadré 1 de la fiche 28].
Par rapport à l'édition précédente, la part des cotisations est révisée à la baisse, en raison du changement de source de cotisations totales : les données du Conseil d'orientation des retraites (diffusées dans le rapport de novembre 2020) remplacent une estimation réalisée à partir du rapport de la commission des comptes de la Sécurité sociale. Cette modification induit une révision à la hausse des cotisations sociales totales, de sorte que la part des cotisations de retraite supplémentaire dans le total est abaissée, de l'ordre de 0,4 point.
Champ > Ensemble des contrats en cours de constitution et de liquidation.
Sources > DREES, enquêtes Retraite supplémentaire de 2010 à 2019 ; rapport du Conseil d'orientation des retraites, novembre 2020.

FIGURE 1.6 – Part de la retraite supplémentaire dans les cotisations et les prestations – Source : Les retraités et les retraites - Edition 2021 - DREES

Plan d'Epargne Retraite individuel ou collectif

Avant d'aboutir au PER Loi PACTE de 2019, de nombreuses générations de contrats de retraite individuelle ont été commercialisées comme les plus récentes : les retraites *« Madelin »* à destination des Travailleurs Non Salariés (depuis 1994), les *« PERP »* à destination des salariés (depuis 2003). De même, plusieurs générations de contrats de retraite à cotisations définies ont été commercialisées auprès d'entreprises : les *« articles 82 »* du Code Général des Impôts, *« articles 83 »* du Code Général des Impôts, *« PERE »* (fusionnés en 2010 avec les articles 83), *« PERCO »* .

Le PER a pour objet l'acquisition et la jouissance de droits viagers personnels ou le versement d'un capital, payables au titulaire à compter, au plus tôt, de la date de liquidation de sa pension dans un régime obligatoire d'assurance vieillesse ou à l'âge légal de départ à la retraite.

Ce produit a été défini avec un but d'harmonisation des solutions de retraite pour tout type d'actif (salarié d'entreprise ou indépendant), afin de permettre au futur retraité un meilleur suivi de sa situation financière globale.

Selon l'origine des sommes versées, on distingue 3 compartiments, ayant chacun des spécificités au niveau de la fiscalité et modalités de sorties :

- Compartiment 1 (PER individuel) : Versements volontaires du titulaire (épargne individuelle et facultative),

- Compartiment 2 (PER d'entreprise Collectif) : Sommes versées au titre de l'épargne salariale (la participation, l'intéressement, l'abondement de l'employeur, droits inscrits sur le Compte Epargne Temps),

- Compartiment 3 (PER d'entreprise obligatoire) : Versements obligatoires de l'employeur et du salarié s'agissant des PER d'entreprise auxquels le salarié est affilié à titre obligatoire.

Les sommes versées ne peuvent être rachetées avant la date de liquidation (date de la retraite) sauf cas exceptionnels de rachats prévus par la règlementation :

- Expiration des droits de l'assuré aux allocations chômage,

- Cessation d'activité non salariée à la suite d'un jugement de liquidation judiciaire,

- Invalidité de l'assuré de deuxième ou troisième catégorie,

- Décès du conjoint ou du partenaire lié par un pacte civil de solidarité,

- Situation de surendettement de l'assuré.

Dans le cas du PER individuel et du PER d'entreprise Collectif, l'achat de la résidence principale est également un motif de rachat autorisé. En outre, pour ces 2 compartiments, la liquidation du contrat est possible aussi bien en rente qu'en capital (ou un mix des 2). Au sein du PER d'entreprise obligatoire, et pour le capital constitué des versements obligatoires, la rente viagère est la seule possibilité de sortie, sauf si le capital constitutif de la rente est inférieur à un seuil fixé par la réglementation. Ce seuil a récemment été réhaussé à 100€ par mois par arrêté du 7 juin 2021 (Art A. 160-2-1 du Code des Assurances).

Du point de vue technique, la réglementation impose un taux technique au plus égal à 0% pour le PER (A142-1 du Code des Assurances), contrairement aux produits de retraite des générations précédentes dont le taux technique pouvait être positif en respectant l'article A132-1 du Code des Assurances (article ayant subi des évolutions à plusieurs reprises au fur et à mesure des années et du contexte économique). Le produit doit faire l'objet d'un cantonnement réglementaire (avant le 01/01/2023). Il peut être libellé en points (régimes L441 du Code des Assurances).

Par ailleurs, le contrat PER doit prévoir obligatoirement une grille d'investissement pilotée qui sécurise progressivement l'investissement à l'approche de la retraite, mais avec toutefois la liberté de proposer d'autres modes de gestion au souscripteur. Du fait que la loi prévoie que des gestionnaires d'actif puissent proposer des contrats PER, dits PER Bancaire (avec une phase de rente assurée par un assureur externe), il n'y a pas d'obligation de garantir le capital pendant la phase de constitution.

Du point de vue fiscal, le dispositif ressemble dans ses principes généraux à ce qui existait déjà sur les produits de retraite en assurance vie avec toutefois des spécificités à prendre en compte selon le compartiment :

- Compartiment 1 : lorsque le détenteur d'un PER a bénéficié de versements déductibles de l'assiette d'impôt sur le revenu (dans la limite d'un plafond réglementaire de 10% des revenus du foyer), le capital racheté est soumis à l'impôt sur le revenu sans avantage fiscal et les plus-values en cas de rachat du capital sont également soumises au prélèvement forfaitaire unique (PFU). En cas de sortie en rente, l'imposition à l'impôt sur le revenu s'effectue après un abattement forfaitaire de 10% (fiscalité des rentes viagères à titre gratuit). L'assuré peut également choisir de ne pas bénéficier de déduction fiscale au moment des versements ; le cas échéant, le rachat en capital est exonéré d'impôt avec les plus-values soumises au PFU, et la rente viagère suit le régime des rentes à titre onéreux, c'est à dire avec un abattement croissant en fonction de l'âge de déclenchement de la rente.

- Compartiment 2 : les versements issus de l'épargne salariale en entreprise sont les plus avantageux fiscalement car ils ont l'imposition la plus légère à l'entrée et à la sortie. Les versements ne sont pas soumis à l'impôt excepté la CSG, le capital racheté non plus, les plus-values au moment du rachat en capital sont soumises aux seuls prélèvements sociaux, et la rente bénéficie du régime des rentes à titre onéreux.

- Compartiment 3 : les versements obligatoires ont la même imposition à l'entrée que ceux du compartiment 2, mais la rente en sortie est soumise au régime des rentes à titre gratuit comme pour les versements déductibles du compartiment 1.

Autres contrats d'Epargne Retraite Collective

L'Epargne Retraite Collective rassemble les produits d'épargne ou de retraite supplémentaire pouvant être constitués par les entreprises pour leurs employés. D'une manière générale, de tels dispositifs sont mis en place dans la branche professionnelle ou dans l'entreprise via un accord entre l'employeur et les représentants des salariés. Le caractère collectif et obligatoire garantit une équité de traitement pour les catégories de salariés concernés et donne droit à des exonérations sociales et fiscales pour l'entreprise. Cette dernière participe à l'effort d'épargne des salariés en contribuant potentiellement aux cotisations.

Il existe deux grands types de régimes de retraite supplémentaire :

- Le régime de retraite à cotisations définies : ce sont les régimes comme le PER d'Entreprise décrit au paragraphe précédent, mais aussi comme les « articles 83 » ou « articles 82 » dont le schéma général est semblable au PER, mais qui peut comporter des garanties de taux supérieures à 0% (taux minimum garanti sur le fonds EURO pendant la phase de constitution, taux technique positif ou nul pendant la phase de restitution)

- Le régime de retraite à prestations définies : Articles 39, IFC.

Articles 39

Les régimes « Articles 39 » reposent sur un engagement sur le montant de la prestation au moment de la souscription du contrat. Le niveau de prime du contrat s'adapte au fur et à mesure que l'engagement de prestation évolue.

Les contrats Article 39 ont été transformés par l'ordonnance n° 2019-697 du 3 juillet 2019 relative aux régimes professionnels de retraite supplémentaire. Le principal changement pour ces régimes concerne l'individualisation des droits à la retraite pour le salarié bénéficiaire du contrat. Dans le

mécanisme précédent, les droits étaient attachés au contrat de l'entreprise et non pas à une tête au sein de cette dernière. En particulier, un salarié qui quittait prématurément l'entreprise perdait tout droit sur le contrat.

Les prestations doivent être exprimées et garanties sous forme de rente annuelle, en euros ou en pourcentage de la rémunération annuelle du salarié. Les droits à prestation s'acquièrent progressivement pour le salarié bénéficiaire et le droit annuel est plafonné à 3% du salaire de l'année. Une fois attribués, ces droits sont définitivement acquis, même en cas de départ de l'entreprise. La revalorisation de la rente est définie au contrat dans la limite de la revalorisation du Plafond Annuel de la Sécurité Sociale (PASS).

Les cotisations payées par l'entreprise permettent de couvrir chaque année au moins 80% des engagements futurs de rentes de l'entreprise calculées avec les données techniques en vigueur, et 100% de l'engagement pour les rentes en cours.

IFC

Sous certaines conditions, un salarié quittant son entreprise au moment de son départ à la retraite a le droit à des indemnités de fin de carrière (IFC) dont le montant est défini par la loi, la convention collective applicable à son entreprise ou par le contrat de travail du salarié concerné (généralement, le montant des indemnités dépend de l'ancienneté du salarié à la date de son départ à la retraite).

Ce passif social de l'entreprise peut être externalisé dans un contrat d'assurance IFC en échange d'avantages fiscaux.

Pour tous les contrats IFC et Article 39, l'assureur est engagé à partir des montants des sommes investies sur le contrat (elles-mêmes calculées à partir de l'engagement de prestation), et peut éventuellement garantir un rendement annuel minimum. Si le montant de prestations que doit verser l'entreprise à son salarié est supérieur aux engagements de prestation du contrat, c'est à l'entreprise de compléter la prestation et non pas à l'assureur. L'entreprise n'a pas obligation de couvrir ses engagements sociaux par un contrat d'assurance.

1.3 Evaluation des risques en norme Solvabilité II

La Directive Solvabilité II est entrée en vigueur depuis le 1er janvier 2016. Elle définit pour tous les assureurs européens le cadre de gouvernance, de communication financière et de gestion de la solvabilité. Cette norme a été mise en place pour répondre à plusieurs objectifs :

- intégrer dans le bilan des assureurs une vision économique plutôt que comptable.

- prévoir un niveau de solvabilité ajusté du niveau de risque de l'entreprise afin de mieux protéger les assureurs et le système financier en général en cas de crise ; en effet, sous la norme Solvabilité I, le niveau du capital minimal exigé pour couvrir les activités d'assurance était défini par un pourcentage fixe du niveau des provisions techniques.

- pousser les entreprises d'assurance à piloter leur activité avec la prise en compte d'un niveau de risque adéquat via une gouvernance et des études de risque adaptées. Dans la notice « Solvabilité II - Modèles Internes » délivrée par l'ACPR, l'orientation 14 précise « 23.L'entreprise veille à ce que le modèle interne soit utilisé pour la prise de décision et elle est en mesure de démontrer cet usage ».

- homogénéiser les pratiques et la lisibilité des indicateurs en matière de communication financière.

1.3.1 Les 3 Piliers de la Directive

La Directive repose sur 3 piliers distincts :

Pilier I

Ce pilier de la Directive couvre l'approche quantitative du bilan fondée sur l'évaluation des fonds propres disponibles de la compagnie (EOF, *Eligible Own Funds*), ainsi que le calcul de l'exigence de capital de solvabilité requis. Le rapport entre le niveau des fonds propres disponibles (EOF) et le capital de solvabilité requis (SCR, *Solvency Capital Requirement*) s'appelle le ratio de solvabilité de l'entreprise.

Si ce dernier est évalué en deçà de 100%, la directive prévoit en particulier les dispositions suivantes à l'article 136 : *3. L'autorité de contrôle exige de l'entreprise d'assurance ou de réassurance concernée qu'elle prenne les mesures nécessaires pour rétablir, dans un délai de six mois après la constatation de la non-conformité du capital de solvabilité requis, le niveau de fonds propres éligibles couvrant le capital de solvabilité requis ou réduire son profil de risque afin de garantir la conformité du capital de solvabilité requis.*

La Directive prévoit également un niveau minimal de capital requis (MCR, *Minimum Capital Requirement*).

Le bilan en normes SII se présente schématiquement ainsi :

Voir figure 1.7

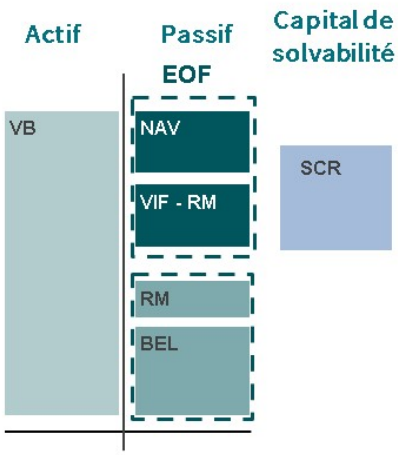


FIGURE 1.7 – Schéma général du bilan S2

L'actif est évalué en valeur de marché (VB, Valeur Boursière). Le passif peut se décomposer en plusieurs briques :

- La *« Net Asset Value »* qui correspond aux fonds propres de l'assureur en valeur de marché.

- La *« Value of Inforce »* (VIF) qui correspond aux profits futurs de l'assureur, évaluée comme la

valeur actuelle des résultats futurs distribuables à l'actionnaire générés par le portefeuille de contrats qui évolue en run off et avec le principe de continuité d'exploitation. Elle peut être exprimée brute ou nette d'impôts selon le besoin.

- Le *« Best Estimate Liabilities »* (BEL) qui correspond à la meilleure estimation des flux futurs liés aux engagements vis-à-vis des intervenants au contrat (assurés, distributeurs).

Il faut également évaluer une marge pour risque (RM, *Risk Margin*) qui représente le coût du capital que devrait lever le cessionnaire pour couvrir son exigence de capital jusqu'à l'extinction des passifs (se déduit de la projection du capital requis sous Solvabilité 2). Une fois ces éléments quantifiés, les fonds propres éligibles (EOF) peuvent être calculés comme la somme de la Net Asset Value et de la VIF, auxquelles on retranche la Marge pour Risque (RM). Les provisions techniques en Solvabilité 2 représente la somme du BEL et de la RM.

A noter que depuis le 24 décembre 2019, l'article A132-16-1 du code des Assurance permet à titre dérogatoire dans des situations exceptionnelles et sous certaines conditions une reprise de la provision pour participation aux bénéfices (PPB). Cette dernière peut donc depuis cette date être comptabilisée entièrement ou en partie en tant que fonds propres éligibles à la couverture du capital de solvabilité, c'est à dire comptabilisée comme une composante supplémentaire de l'EOF. Une note technique de l'ACPR encadre le niveau de PPB éligible.

Pilier II

Ce pilier de la Directive vise à décrire les exigences en matière de gouvernance et de gestion des risques au sein de l'entreprise. Cette dernière doit en particulier identifier, cartographier et suivre les risques inhérents à l'activité, se fixer des limites d'appétence au risque révisables annuellement, se doter de politiques formalisées. Ces politiques doivent décrire l'ensemble des règles à appliquer au sein de l'entreprise, des différents organes de décision et des niveaux de contrôle interne.

La liste ci-dessous dénomme quelques-unes de ces politiques :

- politique générale de gestion des risques,

- politique de gestion du capital,

- politique de contrôle interne,

- politique de provisionnement,...

mais il en existe une vingtaine dans les textes règlementaires.

Quatre fonctions clés doivent être prévues dans l'organisation de l'entreprise : la fonction Actuarielle, la fonction de Gestion des Risques, la fonction de Conformité et la fonction d'Audit Interne. Pour chacune de ces fonctions et comme pour les dirigeants effectifs de l'entreprise, la nomination du responsable de la fonction clé est notifiée au régulateur via un dossier formel (l'ACPR pour la France). La Directive définit des critères d'honorabilité, de compétence et d'expérience pour le choix du responsable et l'ACPR demande à ce que soient documentés les pouvoirs liés à cette fonction délégués par les dirigeants effectifs et le lien d'autorité entre les dirigeants effectifs et le responsable de la fonction clé.

La Directive définit également une exigence de formation et de diffusion de la culture du risque au sein de l'entreprise.

Au coeur du pillier II de la directive, l'ORSA ou *« Own Risk and Solvency Assessment »*, est un processus qui doit permettre aux dirigeants de s'assurer de l'adéquation du calcul des fonds propres

au profil de risque de l'entreprise. Un rapport interne annuel ORSA permet de rendre compte de ce processus. Il vise à traiter l'ensemble des risques, que ce soit ceux quantifiés selon les règles de la directive, ou ceux non quantifiés (comme par exemple les risques stratégiques ou les risques de liquidité).

Pilier III

Ce pilier de la Directive organise et définit les exigences en matière de publication financière, aussi bien auprès du superviseur qu'auprès du public.

L'entreprise se doit de produire régulièrement des reportings :

- reportings trimestriels quantitatifs : les QRT (certains QRT sont uniquement annuels) pour publier en interne et au régulateur l'état des comptes détaillés de l'entreprise (activité et résultats).

- reportings annuels : RSR (*Regular Supervisory Report* à destination du régulateur) ou le SFCR (*Solvency and Financial Condition Report* à destination du public, tous les 3 ans sauf si le superviseur souhaite une publication plus régulière). Ces reportings sont à la fois quantitatifs et qualitatifs car ils doivent décrire le système de gouvernance, le profil de risque de l'entreprise, mais également la valorisation du bilan et la gestion du capital de l'entreprise.

1.3.2 L'évaluation du SCR en Formule Standard

Le Bilan S2 est fondé sur des valeurs économiques à l'actif et au passif, sous l'hypothèse de continuité d'exploitation. Les flux futurs servant au calcul du Best Estimate sont actualisés avec la courbe des taux sans risque publiée par EIOPA. Afin de capturer les effets non linéaires de certains risques (comme le risque de taux garanti), une approche par simulations est plus appropriée, ce qui conduit à utiliser des scenarios risque-neutre stochastiques pour faire évoluer les actifs produits par un Générateur de Scénarios Economiques (ESG).

Le principe sous-jacent du SCR correspond au capital à immobiliser pour couvrir une perte bicentenaire à horizon 1 an, soit pour que la probabilité de ruine soit inférieure à 0,5%. La norme Solvabilité II prévoit une approche par brique de risques et des niveaux de chocs prédéfinis appelés Formule Standard. Chaque brique de risque donne lieu à un SCR calculé après un choc bicentenaire propre à ce risque. Le niveau de ce choc est prévu dans la réglementation pour la Formule Standard, mais il peut aussi être calibré avec les données de l'entreprise pour un assureur qui développe un Modèle Interne en recherchant le quantile à 99,5% des lois statistiques du portefeuille modélisé.

Plusieurs étapes de calcul permettent d'obtenir le SCR final :

- Etape 1 : Tous les SCR rattachés à un même risque principal sont agrégés à l'aide de matrices de corrélation définies dans la norme.

- Etape 2 : L'ensemble des risques principaux sont également agrégés à leur tour avec une matrice de corrélation qui leur est propre afin d'obtenir le SCR de base.

- Etape 3 : Le SCR final est la somme du SCR et de deux autres composantes décrites ci-après.

Du point de vue calculatoire, si on définit CorrSCR (i,j) le coefficient de corrélation entre les risques i et j, on calcule le SCR du risque principal k de la manière suivante :

$$SCR_k = \sqrt{\sum_{i,j} CorrSCR(i,j) \times SCR_i \times SCR_j}.$$

Ce même principe de calcul est répété avec les risques principaux pour obtenir le BSCR *Basis Solvency Capital Requirement* :

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

Le calcul de la dernière étape se présente comme suit :

$$SCR = BSCR - Adj + SCR_{operationnel}.$$

Le terme d'ajustement est la somme de deux éléments au minimum positifs contenant la capacité d'absorption des chocs par la participation aux bénéfices ainsi que celle des impôts différés, tandis que le BSCR évalue le coût de chaque risque sans intégrer ces absorptions.

Schématiquement, on trouve les briques de risques suivantes comme composantes du SCR d'une compagnie en Formule Standard :

Voir figure 1.8

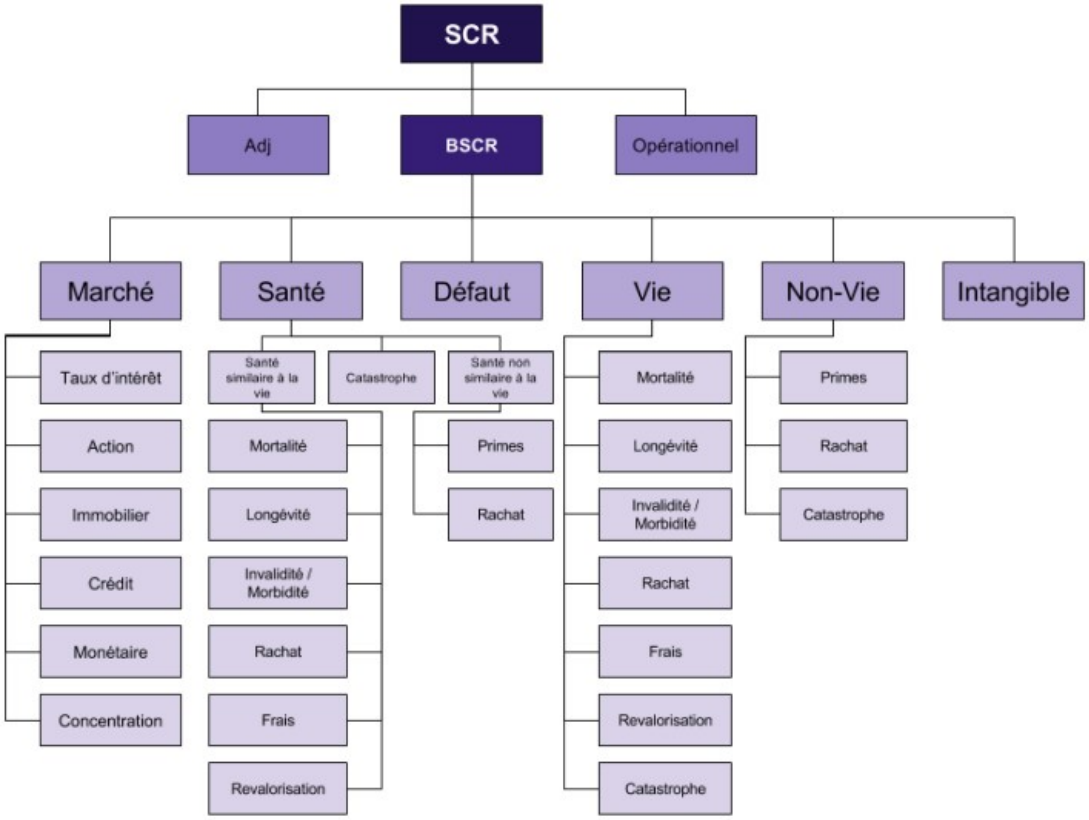


FIGURE 1.8 – Composantes du SCR en Formule Standard

Pour chaque choc, le SCR vaut :

$$SCR_{risquei} = max(0, \delta FondsPropres) = max(0, \delta Actif - \delta BestEstimate).$$

Le risque Opérationnel en formule standard est une composante de 3 éléments, le BSCR, le BORC *« Basic operational risk charge »* et le niveau d'expense pour les unités de compte (noté ici UL_{exp}), définis de la manière suivante :

Le BORC est le maximum entre des facteurs appliqués aux différents volumes de chiffres d'affaires euro et uc des 12 derniers mois, et de facteurs appliqués aux différents volumes de provisions techniques (hors marge pour risque) vie et non vie.

$$SCR_{operationnel} = min(30\% \times BSCR, BORC) + 25\% \times UL_{exp}.$$

AXA a développé un modèle interne qui lui permet de mesurer son besoin en capital en tenant compte des risques calibrés sur son propre portefeuille.

Dans le cadre de ce mémoire, le portefeuille étudié étant un portefeuille de retraite fictif, la Formule Standard décrite dans ce chapitre est plus adéquate pour la mesure du besoin en capital. Les risques principaux portés par ce portefeuille fictif ont été chiffrés et présentés dans le chapitre 4 :

- risques techniques : le portefeuille comporte une phase de rente qui est le mode de sortie principal, mais également une phase de constitution qui est soumise à un risque de transfert des contrats vers d'autres compagnies (faculté offerte aux entreprises ou aux assurés dans la réglementation) ou de rachat individuel (rachats exceptionnels prévus par la réglementation) ; les risques modélisés sont donc le risque de longévité, de frais généraux, et de rachat.

- risques financiers : seules certaines classes d'actif ont été retenues pour ce portefeuille fictif (actions, immobilier, obligations d'état) ce qui reste réaliste par rapport à la composition de portefeuilles réels composés en grande partie d'obligations d'état et avec des poches d'actif diversifiées permettant un revenu régulier et avec un potentiel de rendement plus élevé. Une composante financière en obligations corporate pourrait compléter ce portefeuille mais cette dernière a été négligée dans cette étude par simplification ; les risques modélisés sont donc le risque de taux, action et immobilier.

- risque opérationnel.

1.3.3 L'évaluation de la Risk Margin en Formule Standard

D'après le règlement délégué (UE) 2015/35, l'évaluation de la marge pour risque s'effectue à partir d'une approche coût du capital, c'est-à-dire qu'elle est définie par la valeur actuelle de l'immobilisation du capital pour chaque année future rémunérée au coût du capital. Le capital à immobiliser exclut le risque de marché. Le niveau du coût en capital (CoC) a été fixé par l'EIOPA à 6 %, mais pourrait être revu prochainement dans le cadre de la révision Solvabilité 2 étudiée par l'EIOPA depuis 2020.

Du point de vue calculatoire, la *Risk Margin* s'obtient en projetant tous les SCR techniques, crédit et opérationnel, ce qui donne la formule ci-dessous si r est le taux d'intérêt technique risque neutre dépendant de chaque maturité :

$$RM = CoC \times \sum_{t=0}^{+\infty} \frac{SCR(t)}{\big(1 + r(t+1)\big)^{(t+1)}}.$$

En pratique, le calcul exact à chaque pas de temps du capital réglementaire n'est pas possible car nécessiterait trop de calculs (toutes les briques de risque à calculer à chaque pas de temps avant agrégation dans le SCR). La réglementation autorise dans sa Notice Solvabilité 2 sur les provisions techniques (2015) d'utiliser une méthode simplifiée, sous conditions de vérifier de la robustesse de cette dernière en tenant compte du profil de risque du portefeuille. Les compagnies d'assurance optent donc généralement pour une projection par *driver*. Pour chaque risque, il est possible de calibrer un *driver*, puis de projeter chaque SCR avec ce *driver* et obtenir après agrégation les SCR de chaque pas de temps.

Chapitre 2

Modélisation et gestion d'un portefeuille retraite

Comme vu dans la section précédente, le calcul des différents indicateurs sous S2, (BEL, VIF, SCR, RM) passe par une projection sur les années futures des engagements pris par l'assureur à une date donnée. Un modèle de projection est donc nécessaire pour le calcul de chacun de ces indicateurs.

Cette étude s'attache exclusivement à un portefeuille composé de contrats de type retraite, investis 100% en fonds euros. On suppose que ce sont des « article 83 », c'est-à-dire des contrats de retraite avec des droits individualisés, soumis à des contraintes strictes concernant leur rachat (en particulier transferts vers une autre compagnie) ou sortie en capital, et portant des garanties en capital : taux minimum garanti pour la phase de constitution, taux technique pour la phase de restitution. Une vision simplifiée du modèle de projection est décrite ci-après.

A chaque pas de temps dans la projection, les différents éléments du compte de résultat sont calculés.

Ce modèle peut-être utilisé avec des scénarios générés en univers risque-neutre (supposant ainsi que la performance en moyenne de tous les actifs est le taux sans risque) pour des fins de valorisation du bilan économique ou avec des scénarios générés en « univers réel » pour assurer la gestion de risque ou la détermination d'allocation d'actifs.

Dans cette étude, les deux environnements vont être exploités :

- Des scénarios risque-neutre afin de calculer l'exigence de capital Solvabilité II et de quantifier le bilan économique.

- Des scénarios monde réel afin de permettre au modèle d'adopter la décision d'allocation d'actif optimale intégrant à la fois des contraintes de rendements effectifs, et de solvabilité.

Les sections suivantes décrivent le cadre de modélisation ALM qui a été mis en place pour cette étude. Comme déjà précisé, il s'agit d'un portefeuille simplifié du point de vue du passif et de l'actif.

2.1 Modèle de projection du passif

Au cours de la projection, il est nécessaire de calculer les cash-flows de passif entrants et sortants, ce qui implique d'écouler à chaque pas de temps les provisions techniques de chaque assuré ou groupe d'assuré.

2.1.1 Les provisions techniques

Phase de Constitution

On suppose ici que le contrat se présente sous forme d'une capitalisation financière pour sa phase de constitution, comme une phase d'épargne.

On introduit le taux de chargement sur prime prévu au contrat noté g_1, et la prime versée par le souscripteur du contrat, appelée $Prime_{brute}$.

Les primes versées sont investies sur un fonds en euros (avec garantie financière de l'assureur) ou des supports en UC, ce qui donne la relation suivante à la date de versement de la prime :

$$PM = Prime_{brute} \times (1 - g_1) = Prime_{nette}.$$

A chaque pas de temps t, pour un assuré donné d'âge x, il faut prendre en compte le taux minimum garanti, les lois biométriques et les loi de comportement client pour faire évoluer la provision mathématique. On note TMG le taux minimum garanti de capitalisation du contrat pendant la phase de constitution ; ce dernier est défini contractuellement et dans le respect de la réglementation. $q_x(t)$ est le taux de décès entre l'âge x et x+1 de la table de mortalité utilisée. $r(t)$ est le taux de rachat en capital ou transfert du contrat de retraite (tels que prévus par la réglementation) entre la date t et t+1. La provision mathématique s'exprime avec la formule suivante :

$$PM(t+1) = PM(t) \times \big(1 - q_x(t) - r(t)\big) \times \big(1 + TMG\big).$$

L'équation ci-dessus suppose :

- Que les sorties (décès, transferts et rachats exceptionnels tels que prévus par la réglementation) interviennent en milieu d'année (hypothèse de non-saisonnalité des sorties).

- La PM est capitalisée au taux minimum garanti (celui-ci pouvant éventuellement être nul).

- Il est à noter que pour ce contrat, si l'assuré décédait pendant la phase de constitution, alors la PM serait due à ses héritiers.

Les contrats de retraite peuvent également se présenter sous forme de rente viagère différée, avec une capitalisation viagère en phase de constitution. Dans ce cas, l'assureur est déjà engagé sur un niveau de taux technique et de table de mortalité pendant la phase de constitution, et l'assuré ne bénéficie pas de contre-assurance en cas de décès. La provision mathématique du contrat pendant la phase de constitution se calcule alors de manière similaire au calcul présenté ci-après pour la phase de restitution, mais en tenant compte du différé de la rente.

Phase de Restitution

Au moment de la liquidation en rente du contrat au temps $t = T_{retraite}$ avec un assuré d'âge x, plusieurs paramètres contractuels entrent en jeu :

- Le capital constitutif de la rente (CC) donné directement par la Provision Mathématique disponible à cette date en fin de phase de constitution.

- Les caractéristiques prévues au contrat pour le versement de la rente : à terme à échoir ou échu, de mensuelle à annuelle, avec ou sans réversion, avec ou sans indexation au fil du temps.

- La table de mortalité contractuelle (qui peut être définie comme celle en vigueur à la date de liquidation de la rente), pour laquelle le nombre de personnes en vie à chaque âge est défini par lx. Dans le cas des rentes viagères, les tables exigées par le code des assurances actuellement sont les tables TGH05 et TGF05. Ces dernières définissent des niveaux de mortalité différents selon la génération de naissance et le sexe. On note ω l'âge maximal de la table (généralement 120 ans).

- le taux technique r (qui peut être défini dans le contrat comme le taux réglementaire en vigueur à la date de liquidation de la rente).

- le chargement de rente g_2 qui permet à l'assureur de financer les coûts de gestion des rentes.

On calcule alors l'arrérage annuel A grâce à la relation suivante, en supposant une rente annuelle à terme échu sans réversion et sans indexation :

$$CC = PM_{constitution}(T_{retraite}, x) = \sum_{i=1}^{\omega-x} \frac{A}{1+g_2} \times \frac{l_{x+i}}{l_x} \times \frac{1}{(1+r)^i},$$

ce qui donne donc le montant d'arrérage :

$$A(T_{retraite}) = \frac{CC \times (1+g_2)}{\sum_{i=1}^{\omega-x} \times \frac{l_{x+i}}{l_x} \times \frac{1}{(1+r)^i}}.$$

Puis à chaque pas de temps, la PM peut se calculer à partir de l'arrérage servi pour la rente :

$$PM_{restitution}(t, x) = \sum_{i=1}^{\omega-x} \frac{A(T_{retraite})}{1+g_2} \times \frac{l_{x+i}}{l_x} \times \frac{1}{(1+r)^i}.$$

2.1.2 Le compte de participation aux résultats

Les contrats de retraite collective prévoient généralement des clauses de participation aux bénéfices complémentaires aux exigences réglementaires. En effet, le Code des Assurances dans les articles A132-10 et suivants impose pour les fonds en euro qu'au minimum 90% des résultats techniques (lorsque positifs, sinon 100% du solde technique débiteur) et 85% des résultats financiers soient distribués aux assurés. En cas de solde global débiteur, ce dernier peut toutefois être reporté l'année suivante.

La participation aux bénéfices peut prendre plusieurs formes :

- Une participation aux bénéfices financière : une partie des produits financiers nets de marge de gestion est ainsi distribuée aux assurés (si ce montant est positif).

- Une participation aux bénéfices technique : une partie du résultat technique si celui-ci est positif est distribuée aux assurés.

- Une participation aux bénéfices technico-financière : une partie du résultat technico-financier (principe de mutualisation technico-financière) si celui-ci est positif est distribué au client.

Généralement l'assureur prévoit deux comptes de participation aux résultats séparés pour la phase de constitution et la phase de restitution d'un même contrat pour des contrats de rente viagère immédiate avec capitalisation financière préalable. Dans le cas des RVD (rentes viagères différées), il n'est pas possible de séparer ces deux phases du contrat et il y a donc nécessairement un seul compte de participation aux résultats.

Compte de résultat financier

Le contrat définit un niveau de marge sur encours fixe, noté ENC ainsi qu'un pourcentage de produits financiers distribués α pour encadrer le niveau de participation aux bénéfices financiers minimal contractuel.

Dans le contrat, le montant de marge sur encours s'exprime généralement comme un taux appliqué à l'assiette de provision rémunérée. Le taux de distribution des produits financiers α s'élève historiquement au-dessus de 85% pour garantir le minimum réglementaire pour chaque contrat. Plus récemment les assureurs ont baissé ce taux pour les nouveaux contrats, tout en respectant au global de leur portefeuille le niveau minimal réglementaire requis. En réduisant ainsi le niveau de garantie minimale dans chaque contrat, la capacité de pilotage de la participation aux bénéfices de l'ensemble des contrats se trouve augmentée, et potentiellement le niveau d'exigence du capital réglementaire est réduit. En effet, il est possible de modéliser une plus grande absorption des chocs de marché par les assurés en distribuant moins de participation aux bénéfices aux contrats dont la garantie le permet. Dans ce cas, la VIF stochastique est plus élevée, et le SCR marché est également réduit.

En notant PFI le montant total des produits financiers réalisés, en général nets des frais de placement, et IT les intérêts techniques issus du taux minimum garanti (TMG) ou du taux technique de la rente, on obtient une participation aux bénéfices complémentaires aux intérêts techniques :

$$PB_{financière} = max(0, \alpha \times PFI - ENC - IT).$$

Certains contrats prévoient sur la phase d'épargne le prélèvement de la marge sur encours même en cas de produits financiers plus faibles que les intérêts techniques garantis. Cette formulation de clause de participation aux bénéfices a été introduite au fur et à mesure dans les contrats dans le contexte de taux bas afin de faciliter la prise de marge de l'assureur même en cas de produits financiers bas. Dans ce cas, le capital garanti après incorporation de la participation aux bénéfices peut diminuer, et la PB financière se calcule de la manière suivante :

$$PB_{financière} = max(0, \alpha \times PFI - IT) - ENC.$$

Les intérêts techniques peuvent s'exprimer selon la phase de constitution ou la phase de restitution avec les formules ci-dessous :

$$IT_{constitution}(t) = PM_{constitution}(t) \times TMG + (Primes(t) - Prestations(t)) \times \big((1 + TMG)^{\frac{1}{2}} - 1\big),$$

et de la même façon :

$$IT_{restitution}(t) = PM_{restitution}(t) \times r - Prestations(t) \times \big((1 + r)^{\frac{1}{2}} - 1\big).$$

En phase de constitution, l'incorporation à la PM donne le calcul suivant :

$$PM_{constitution}(t+1) = PM_{constitution}(t) \times \big(1 - q_x(t) - r(t)\big) + IT_{constitution}(t) + PB_{financière}(t).$$

En phase de restitution, la $PM_{restitution-avantPB}(t+1)$ correspond à la PM de début de période diminuée des arrérages versés pendant l'année lorsque la personne est encore vivante en fin de période. L'incorporation à la PM donne la relation suivante :

$$PM_{restitution-apresPB}(t+1) = PM_{restitution-avantPB}(t+1) + IT_{restitution}(t) + PB_{financière}(t).$$

Cela donne lieu à une réévaluation de l'arrérage distribué :

$$A(t+1) = \frac{PM_{restitution-apresPB}(t+1) \times (1+g_2)}{\sum_{i=1}^{\omega-x} \frac{l_{x+i}}{l_x} \times \frac{1}{(1+r)^i}}.$$

A noter que le code des Assurances autorise l'assureur à mettre en réserve cette participation aux bénéfices jusqu'à 8 ans (la PPB citée au paragraphe 1.3.1). Dans ce cas, elle n'intègre pas tout de suite la PM et l'arrérage du client n'est pas réévalué. Cependant, en retraite collective, elle est fréquemment distribuée dans l'année de sa constitution.

Solde technique

Le résultat technique du contrat est défini comme suit :

$$Résultat_{technique}(t) = Primes_{nettes} + PM_{ouverture}(t) - PM_{clôture}(t+1) + IT(t) - Prestations_{brutes}(t).$$

La notion de nette et brute désigne ici la prise en compte ou non des chargements d'acquisition sur les primes ou de gestion sur les prestations. Puis la PB technique se calcule ainsi :

$$PB_{technique} = max(0, \beta \times Résultat_{technique}),$$

où β désigne le taux de distribution des produits financiers (historiquement au-dessus de 90% pour garantir le minimum réglementaire pour chaque contrat).

Cette PB technique se déverse dans la PM des assurés de la même manière que la PB financière, sans toutefois pouvoir être mise en réserve dans la PPB.

Mutualisation technico-financière

Dans le cas d'un contrat prévoyant une mutualisation technico-financière, l'assureur prévoit de déduire les éventuels résultats techniques négatifs du montant global de participation aux bénéfices. On peut donc parler d'une PB technico-financière, exprimée de la manière suivante :

$$PB_{technico-financière} = max(0, PB_{financière} + \beta \times Résultat_{technique}),$$

équivalent à la formule suivante :

$$\begin{aligned} PB_{technico-financière} &= PB_{financière} + PB_{technique} \\ &\quad - min\big(PB_{financière}, -min(0, \beta \times Résultat_{technique})\big). \end{aligned}$$

Dans un contrat de retraite, le résultat technique est issu des écarts de mortalité pour les rentes immédiates ou différées entre la mortalité prévue par la table réglementaire et le réel observé sur le portefeuille.

Un mécanisme de report de perte peut être prévu au contrat, ce qui permet à l'assureur de reporter pour le futur des pertes non absorbées par la PB financière une année donnée.

2.1.3 Marge de gestion

Les frais de chargement contractuels permettent d'absorber les coûts de l'assureur. Ainsi, la marge de gestion tient compte des prélèvements assurés nettés des frais généraux de l'entreprise.

On peut donc calculer la marge de gestion avec la relation suivante :

$$marge_{gestion}(t) = g_1 \times Prime(t) + g_2 \times A(t) - frais_{generaux}(t).$$

2.2 Modèle de projection de l'actif

De la même manière que pour le passif, les cash-flows de l'actif sont évalués à chaque pas de temps. Il s'agit de valoriser les instruments financiers selon leurs caractéristiques et en fonction de chaque scénario d'actif, mais également d'avoir la capacité à réinvestir en cas d'arrivée à échéance ou de réajuster les choix d'allocation d'actif. Dans cette partie on s'attachera au descriptif des instruments étudiés et de leur vieillissement dans le temps. On se limitera donc ici aux obligations d'état ainsi qu'aux actions ou investissements immobiliers.

2.2.1 Comptabilisation des actifs en Assurance

Le Code des Assurances distingue quatre grands types de classes d'actif du fait de leur mode de comptabilisation :

- Les actifs *« R343-9 »* qui comprennent généralement les obligations détenues en direct (sans risque ou risquées). Ces actifs sont amortissables : la différence entre le prix d'achat et la valeur de remboursement est amortie au cours de la durée de vie des obligations. Une autre particularité de ces actifs concerne la comptabilisation des plus ou moins-values latentes réalisées. En cas d'achat ou de vente d'un titre « R343-9 », le montant de plus ou moins-value latente est comptabilisé dans la réserve de capitalisation (tant qu'elle reste positive). De ce fait, le revenu de l'année n'est pas impacté de cette opération.

- Les actifs *« R343-10 »* amortissables qui comprennent les prêts ou obligations non cotées. Leur valeur de remboursement et leur arrivée à échéance sont connues à l'avance.

- Les actifs *« R343-10 »* non amortissables qui comprennent les actions cotées ou non cotées, les valeurs immobilières papier ou pierre en direct, les OPCVM (Organismes de Placement Collectif de Valeurs Mobilières). Leur valeur nette comptable est égale à leur prix d'achat éventuellement diminuée de la provision pour dépréciation durable. Cette provision peut être constituée sur un actif dont la dépréciation a un caractère « durable » conformément au règlement ANC (Art 123.6 à 123.19).

- Les actifs *« R343-13 »* : placements en unités de comptes admis en représentation de contrats en UC dont la somme assurée est déterminée par rapport à une valeur de référence. Ces derniers ne sont pas représentés dans le portefeuille étudié.

La réalisation de plus ou moins-values sur des actifs *« R343-10 »* impacte les produits financiers, tandis que celle sur des actifs *« R343-9 »* (sauf obligations à taux variable) impactera la réserve de capitalisation (celle-ci ne pouvant être négative).

2.2.2 Chronologie des événements à l'actif sur une période

Plusieurs étapes vont se succéder au cours d'une même période du fait de la nécessité d'intégrer les évolutions de marché relatives à chaque scenario d'actif, mais également la stratégie d'investissement modélisée. Cette dernière se doit de refléter les réalités suivantes :

- Réinvestissements consécutifs au vieillissement de l'actif ou du passif (arrivées à échéances des titres obligataires, performance de l'année, paiement des prestations, réception des nouvelles primes),

- Pilotage long-terme des investissements par le biais d'allocations cibles décidées par le management,

- Optimisation des produits financiers à court terme avec le pilotage du partage des résultats financiers entre l'assureur et les assurés.

C'est pourquoi le modèle tient compte des caractéristiques de chaque actif qu'il fait vieillir, mais il doit également intégrer des paramètres lui permettant d'atteindre une allocation cible définie, ou un algorithme ALM d'optimisation de la valeur.

Voir figure 2.1

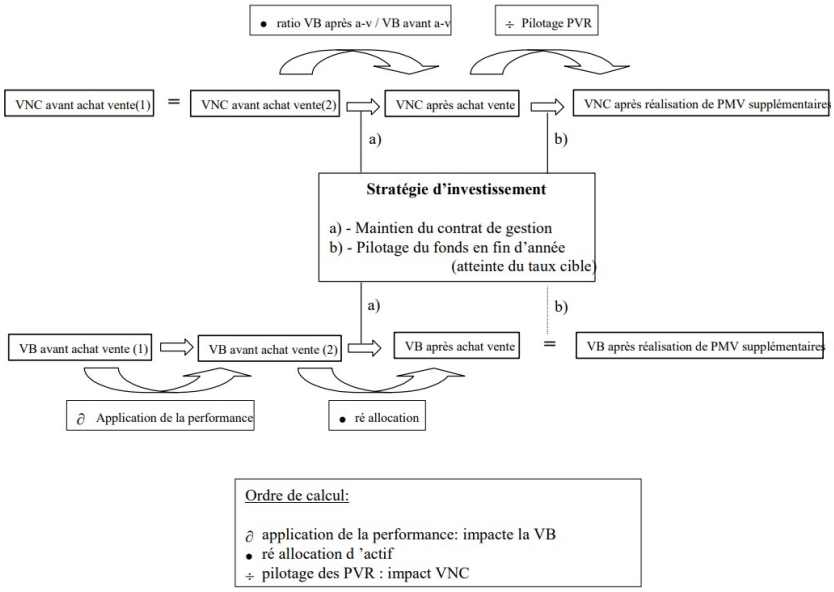


FIGURE 2.1 – chronologie des opérations sur les actifs entre deux pas de temps

2.2.3 Actions et Immobilier

Les actions et les investissements immobiliers peuvent être modélisés de la même façon, via une évolution de leur valeur de marché qui induit une modification de la plus-value latente, ainsi qu'un revenu annuel, le dividende ou les loyers perçus, qui intègre la performance de l'année.

Les scénarios économiques pour ces deux types d'instruments sont différents, paramétrés avec une volatilité et un niveau de dividende propres à chacun d'eux. Ils sont ainsi maintenus dans le

modèle séparément, ce qui permet également un pilotage des plus-values latentes plus adéquat. A contrario, le modèle regroupe toutes les lignes actions ou toutes les lignes immobilières en une seule ligne d'actif, ce qui peut *in fine* limiter la capacité de pilotage à la hausse ou à la baisse en fonction du besoin ou limiter la comptabilisation adéquate de provision pour dépréciation durable (PDD) pour chaque ligne d'actif.

Au départ de la projection, le modèle prend en compte la Valeur Boursière (VB), la Valeur Nette Comptable (VNC) de ces actifs, le niveau de PDD, puis les fait vieillir dans le temps en fonction des scenarios.

Application de la performance

Au cours de l'étape d'application de la performance, la VNC reste inchangée :

$$VNC_{avant-strategie-inv}(t+1) = VNC(t).$$

Le scenario de marché donne une performance brute *Total Return* et un niveau de dividende inclus dans la performance brute *Income Return*.

Ainsi la VB et le revenus évoluent de la manière suivante :

$$\begin{aligned} VB_{avant-strategie-inv}(t+1) &= VB(t) \times \big(1 + TotalReturn(t)\big). \\ Revenu_{avant-strategie-inv}(t) &= VB(t) \times IncomeReturn(t). \end{aligned}$$

Le revenu va intégrer la poche d'actif de cash. De la même manière, les cash-flows de passif impactent la poche de cash :

$$Cash_{avant-strategie-inv}(t+1) = Cash(t) + Revenu_{avant-strategie-inv}(t) + CashFlow_{passif}(t).$$

Application de la stratégie d'investissement

Du fait de la stratégie d'investissement, deux types d'événements peuvent se produire :

- Achat / vente de l'instrument financier via la nouvelle allocation cible

- Réalisation de plus-value latente.

Nouvelle allocation cible : rebalancement du portefeuille

Après chaque étape du modèle, la composition du portefeuille est observée. Si une distribution cible par classe d'actif est définie dans le modèle, alors il faut ajuster par achat/vente le niveau détenu pour chaque instrument après prise en compte de la performance et de l'évolution du passif. Si on appelle $alloc(t+1)$ le niveau d'allocation cible souhaitée pour l'actif concerné, on calcule d'abord un écart d'allocation entre l'allocation en cours et l'allocation cible :

$$Delta_{alloc}(t+1) = \frac{VB_{avant-strategie-inv}(t+1)}{VB^{\text{totale}}{}_{avant-strategie-inv}(t+1)} - alloc(t+1).$$

Puis en cas de Delta négatif (nécessité de vendre), la vente s'effectue sans déformation du taux de plus-value latente ; a contrario, en cas de Delta positif (nécessité d'acheter), l'achat introduit une

nouvelle VNC sans plus-value latente ajoutée dans la VB :

$$\begin{aligned} VB_{avant-PVRsuppl}(t+1) &= VB_{avant-strategie-inv}(t+1) \times \Big(1 + min\big(0, Delta_{alloc}(t+1)\big)\Big) \\ &\quad + VNC_{avant-strategie-inv}(t+1) \times max\big(0, Delta_{alloc}(t+1)\big), \end{aligned}$$

$$VNC_{avant-PVRsuppl}(t+1) = VNC_{avant-strategie-inv}(t+1) \times \Big(1 + Delta_{alloc}(t+1)\Big),$$

$$\begin{aligned} Cash_{avant-PVRsuppl}(t+1) &= Cash_{avant-strategie-inv}(t+1) \\ &\quad + VB_{avant-strategie-inv}(t+1) \times min(0, Delta_{alloc}(t+1)). \end{aligned}$$

Réalisation supplémentaire de plus-values latentes

L'algorithme ALM peut être amené à choisir la réalisation de plus-values ou moins-values latentes afin de piloter le taux de marge ou de participation aux bénéfices. Dans ce cas la réalisation de ces plus-values latentes est considérée comme une opération de vente puis de rachat successives qui n'impacte donc pas la valeur de marché globale sur la classe d'actif considérée.

On note $PVRsuppl(t+1)$ le montant de plus-values à réaliser. Puis on obtient :

$$VB_{apres-PVRsuppl}(t+1) = VB(t+1) = VB_{avant-PVRsuppl}(t+1),$$

$$VNC_{apres-PVRsuppl}(t+1) = VNC(t+1) = VNC_{avant-PVRsuppl}(t+1) + PVRsuppl(t+1).$$

Dans le modèle interne AXA France, plusieurs types d'algorithmes sont implémentés selon la nature du portefeuille de passif sous-jacent, afin de refléter les *management actions* prévues dans le futur sur le taux à servir aux assurés ou la marge minimale à dégager.

Rachats dynamiques

Le risque de rachat dynamique est un risque surveillé de près dans les compagnies d'assurance en épargne, mais également en retraite. Le comportement de rachat des assurés observé de manière structurelle est généralement lié à des paramètres endogènes aux contrats (structure fiscale par exemple en épargne, contraintes commerciales et réglementaires sur les sorties en retraite). Pourtant des paramètres exogènes de marché pourraient aussi générer des augmentations de sorties (transferts) en cas de difficulté à générer de la performance financière par rapport à des concurrents. Un comportement dynamique des assurés a donc été modélisé pour matérialiser la satisfaction client selon l'évolution financière du fonds EURO ; en cas de produits financiers trop faibles ou de richesse résiduelle insuffisante dans le fonds euros, le client transfère son contrat de retraite vers une autre compagnie.

La loi choisie pour ce modèle se base sur l'idée que l'entreprise cliente réagit à un score de satisfaction mesuré à chaque pas de temps à partir de 4 facteurs selon la formule suivante :

$$satisfaction(w, x, y, z) = \mu_1 \times (w - w_0) + \mu_2 \times (x + x_0) + \mu_3 \times (y - y_0) - \mu_4 \times (z - z_0),$$

où $\mu_1, \mu_2, \mu_3, \mu_4, w_0, x_0, y_0, z_0$ sont des paramètres positifs fixés dans le modèle, et :

- w : niveau de richesse globale du fonds (VB totale / VNC totale). Ce niveau de richesse peut être communiqué à certaines entreprises clientes et donc une situation de moins-value latente met en risque la confiance du client.

- x : $taux_{servi}(t-1) - \frac{1}{2} \times (taux_{inflation} + taux_{concurrent})$, où le concurrent est modélisé comme un nouvel entrant du marché avec un taux de rendement à 80% de taux 10 ans et 20% d'equity additionnés d'un taux de surperformance fixe. Un taux effectivement servi l'année précédente inférieur à celui d'un concurrent génère un risque de sortie.

- y : TMG, taux minimum garanti (phase de constitution). Le taux garanti sur la phase de constitution représente un facteur de rétention du client dans la mesure où il assure un rendement certain pour l'assuré sur une longue durée.

- z : ENC, marge fixe sur encours du contrat. Le niveau de marge définie au contrat influe sur le niveau de satisfaction puisqu'il affecte le taux de rendement financier net final du contrat.

Si le score est positif, le client est satisfait. Au bout de 3 ans d'insatisfaction (score négatif), il y a une probabilité de rachat de 50% sur le groupe de contrats du model point considéré. Cet effet mémoire se justifie notamment par le fait que le réseau commercial est capable de se mobiliser en cas d'insatisfaction pour ajuster l'offre client et que la décision finale de l'entreprise après un appel d'offre sur le marché et les discussions avec les instances sociales peut prendre un temps non négligeable pour un contrat de retraite collective (estimé à 3 ans ici).

2.2.4 Obligations à taux fixes

Il existe dans un portefeuille réel plusieurs types d'obligations, notamment des obligations à taux fixes, variables, indexées à l'inflation. Dans cette étude, nous nous sommes limités aux obligations à taux fixe.

A l'inverse des autres instruments, le modèle récupère la composition ligne à ligne des obligations détenues afin de garder le bon niveau de sensibilité aux taux du fait de la diversité réelle des niveaux de coupons et de maturité dans le portefeuille.

Ainsi, au départ de la projection, on considère les caractéristiques suivantes pour l'obligation :

- Valeur Boursière (VB)

- Valeur Nette Comptable (VNC)

- Maturité (T)

- Coupon (C)

- Durée courue (D)

- Face value (FV) : valeur du nominal

- ZC(t,i) est le Zéro Coupon de maturité t à la date i.

A chaque pas de temps, la valeur boursière s'écrit alors :

$$VB(t) = \sum_{i=1}^{T-t} \frac{FV \times C}{(1 + ZC(t,i))^i} + \frac{FV}{(1 + ZC(t,T))^{(T-i)}}.$$

En début de projection, on calcule le taux de rendement actuariel de l'obligation qui permet d'égaliser la valeur comptable (y compris coupon couru) avec la valeur actualisée des flux.

$$VB(0) = \sum_{i=1}^{T} \frac{FV \times C}{(1 + TRA)^i} + \frac{FV}{(1 + TRA)^T}.$$

Puis la valeur comptable de l'obligation est déterminée à chaque pas de temps en actualisant les flux résiduels avec le TRA.

$$VNC(t) = \sum_{i=1}^{T-t} \frac{FV \times C}{(1+TRA)^i} + \frac{FV}{(1+TRA)^{(T-t)}}.$$

Par ailleurs, les revenus sont calculés de la manière suivante :

$$Revenu(t) = VB(t) - VB(t-1) + CF(t),$$

où CF(t) représente les cash-flow de l'obligation, soient les coupons perçus dans l'année ou bien le nominal récupéré.

Comme ce qui a été décrit pour les instruments financiers de la section précédente, le revenu affecte la poche de Cash, puis le portefeuille doit subir des réallocations liées à l'allocation d'actif cible.

2.3 Gestion ALM au sein d'Axa France

2.3.1 Enjeux de la gestion ALM d'un portefeuille assurantiel en fonds euro

La gestion du portefeuille d'actif d'une compagnie d'assurance est une activité sensible non seulement du fait du besoin de maîtrise des risques financiers associés aux investissements, mais également au vu de la capacité de la compagnie à rester concurrentielle, que ce soit vis-à-vis de ses clients avec des taux de produits financiers attractifs ou bien de ses actionnaires avec des retours sur investissement élevés.

Le pilotage de la gestion d'actifs passe par la décision mesurée d'allocations tactiques (court-terme) ou stratégiques (long-terme). Pour être en mesure de décider d'une allocation optimale, la théorie de Markowitz préconise de déterminer au préalable les allocations efficientes, c'est à dire qui permettent de maximiser le rendement pour un niveau de risque pris, et de minimiser le risque pour un niveau de rendement donné. Puis l'identification d'une fonction d'utilité propre au gestionnaire permet d'effectuer le choix final d'allocation le long de la frontière efficiente.

Pour cela, il est nécessaire de mener des études ALM capables d'intégrer l'ensemble des contraintes de l'assureur, c'est à dire capables de quantifier les risques de marché, mais également les interactions que les investissements opèrent vis-à-vis des engagements au passif.

Ainsi, l'assureur souhaite réaliser les investissements qui minimisent ses risques présents et futurs, mais aussi qui pourraient maximiser le profit distribuable à l'actionnaire et la satisfaction des assurés si le rendement financier de leur contrat est élevé.

2.3.2 Choix des métriques à optimiser

Le couple (rendement, risque) naturellement étudié par tout gestionnaire de portefeuille correspond à l'espérance de rendement futur en environnement monde réel (c'est à dire en tenant compte des primes de risques des instruments financiers) ainsi que la volatilité de celui-ci. Ce sont ces métriques impliquées dans la théorie initiale de Markowitz. Cependant, il est possible de s'appuyer sur cette théorie en la transformant avec des métriques plus adapatées au besoin de l'entreprise d'assurance.

Dans une étude long-terme, les assureurs pourraient ainsi déterminer la frontière efficiente de leur portefeuille en calculant la somme actualisée de toutes les marges futures du portefeuille dans un monde réel projeté et moyennées sur un nombre important de scenarios, et en observant leur volatilité moyenne.

Cependant, depuis l'introduction de Solvabilité II, les métriques de risque choisies pour les études ALM ont évolué : au lieu de s'attacher à la volatilité du rendement espéré, elles tiennent compte généralement du niveau de capital requis pour chaque portefeuille. En effet, le SCR est bâti comme une mesure de risque. Il synthétise à la fois tous les risques directement liés au passif (risques biométriques (longévité), risques de comportement assuré (rachat), risques de dérive des coûts), mais également les risques de marché (évolutions des taux, de la volatilité, des prix de marché). Toute variation de ce dernier pèse dans les comptes de l'assureur puisqu'il agit directement sur la capacité de rendement des actionnaires. Le besoin minimal en capital est donc surveillé de près par les assureurs.

On remarquera qu'en minimisant le SCR, on contribue également à l'optimisation du retour sur investissement de l'actionnaire qui immobilise moins de capital, et donc à l'amélioration du rendement final. Ainsi, une maximisation d'un ROE (Return on Equity) qui serait le rapport entre le rendement futur et le niveau de capital requis pourrait également directement être exploité pour une étude ALM.

Le ratio de solvabilité, qui est le rapport entre l'actif disponible et le capital requis peut également être maximisé.

Dans le cas d'une étude ALM visant à définir une allocation d'actif stratégique optimale, s'intéresser au SCR marché seul peut suffire car on peut supposer que les SCR techniques seront peu sensibles au niveau de l'allocation.

Pour le portefeuille de retraite simplifié décrit en section III et simulé à l'aide du modèle interne d'AXA France nous avons estimé le SCR pour chaque brique de risque dans différentes allocations de portefeuille. En partant d'une allocation centrale *Base Case* (BC), on peut définir par exemple pour les actions d'autres allocations cibles à horizon dix ans allant de -6% à $+8\%$, compensé par une allocation allant de $+6\%$ à -8% sur la cible obligataire. Puis en comparant le SCR obtenu avec le SCR du cas central, on obtient le graphique suivant :

Voir figure 2.2

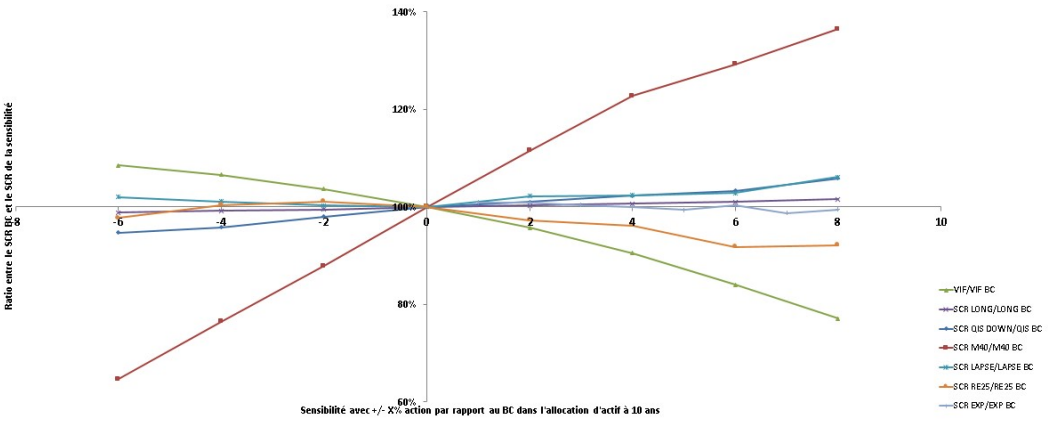


FIGURE 2.2 – Evolution des SCR par brique de risque en fonction de l'allocation d'actif (sensibilité actions)

Les SCR modélisés ici sont le longévité (LONG), le rachat (Lapse), le coûts (EXP), l'action (M40), l'immobilier (RE25) et le taux (QIS DOWN). La VIF en environnement risque neutre est également présentée. Le graphe montre que le SCR action est celui qui varie le plus fortement avec les changements d'allocation (à +/- 40% du SCR Base Case dans cette simulation). Les SCR coûts et longévité restent quasiment stables, très proches de 100%. Les SCR rachat et taux s'écartent légèrement du cas central mais de manière peu significative (<6%). Le SCR immobilier varie peu à la baisse mais plus significativement à la hausse.

En réalisant cette même étude pour des variations d'allocations cibles similaires sur la ligne d'actif immobilière, on observe les mêmes résultats.

Voir figure 2.3

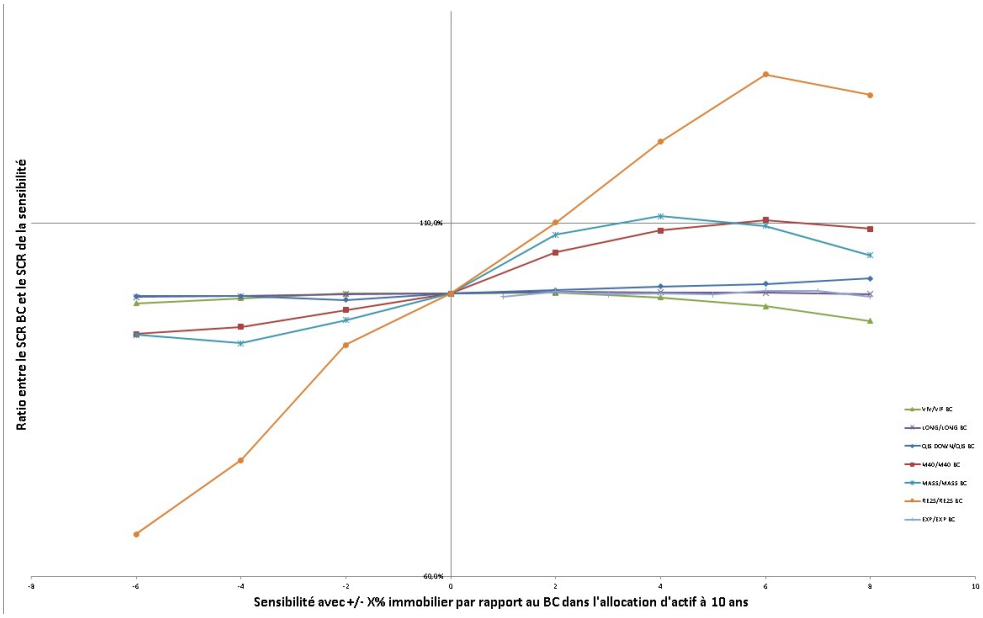


FIGURE 2.3 – Evolution des SCR par brique de risque en fonction de l'allocation d'actif (sensi immobilier)

Dans la mesure où le SCR longévité sera prépondérant dans le SCR technique Vie du portefeuille de retraite étudié, on peut conclure de cette étude que le SCR technique peut être considéré comme constant dans l'étude ALM, et nous préconisons de ne s'intéresser qu'au SCR marché.

2.3.3 Méthodologie usuelle du choix de l'allocation optimale chez Axa France

Lors d'une décision d'allocation d'actifs stratégique pour un portefeuille Axa France, une modélisation des passifs concernés et de l'actif associé, projetée sur 60 ans est réalisée. L'utilisation de scenarios Monde Réel calibrés sur les données de marché les plus récentes permet de simuler les rendements futurs dans différentes configurations d'allocations fixées en hypothèse d'entrée du modèle.

Puis les couples (rendement futur moyen, SCR marché) sont calculés dans chaque situation, ce qui permet de tracer une frontière efficiente d'allocation.

Pour le portefeuille de retraite simplifié décrit au chapitre 4.1 et simulé à l'aide du modèle interne

d'AXA France, l'étude précédente a permis de tracer également les couples (rendements, risque) observés pour les différentes allocations d'actifs. En ajoutant également des variations d'allocations d'actif autour de l'immobilier, cela permet d'obtenir les couples (rendement (VIF monde réel), risque (SCR SII)) de la figure 2.4.

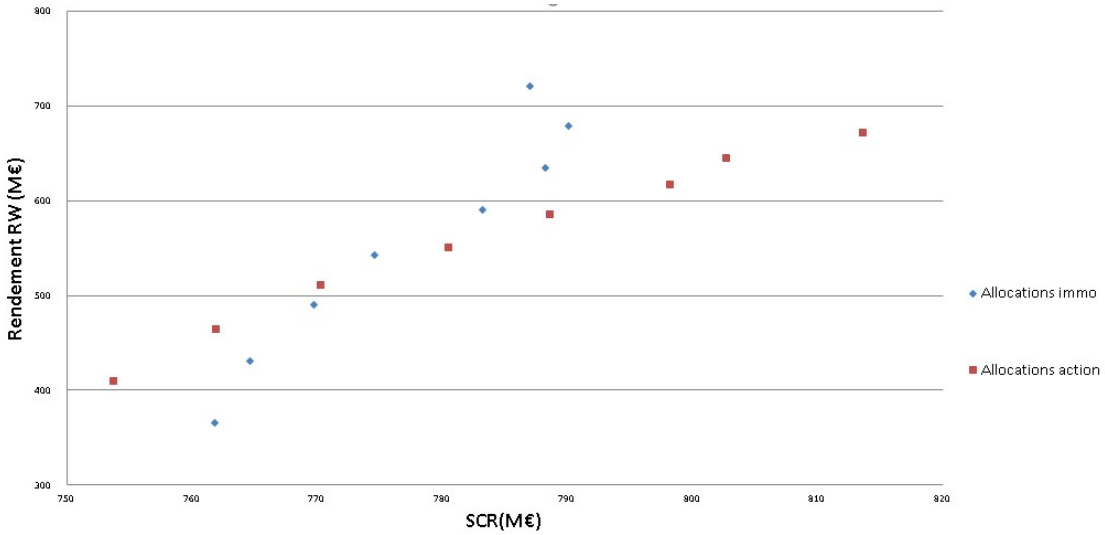


FIGURE 2.4 – Couples (rendement, SCR) en fonction de plusieurs allocations d'actif

Le nuage rouge est celui des sensibilités d'allocation au pourcentage d'actions cible ; le nuage bleu est celui relatif au pourcentage d'immobilier cible. Les points du graphe sont ordonnés en fonction du niveau d'allocation cible sur l'actif étudié. Cette étude simplifiée montre déjà une tendance à accroître simultanément le rendement et le SCR avec une cible croissante d'actions ou immobilier par rapport à un niveau cible pour ces instruments financiers croissant. Cela est bien cohérent avec l'idée générale que le niveau de capital requis est croissant en fonction du niveau de risque des actifs financiers.

Pour pouvoir réaliser une étude complète, il faut tester un jeu important d'allocations, situées pour chaque type d'actif autour de la position courante de ce dernier, afin de pouvoir observer dans chaque situation les métriques de risque choisies (rendement futur, niveau du SCR marché). La métrique de rendement s'obtient à l'aide d'une simulation en univers Monde Réel ; la métrique de SCR marché s'obtient avec une simulation non choquée, puis plusieurs simulations choquées, en univers Risk Neutre.

Si on considère trois classes d'actif, et un intervalle de 20% possible autour de chaque allocation définies à 0,5% près, on va donc simuler environ $(20 \times 2)^2 = 1600$ allocations différentes (dans le respect de la contrainte d'une répartition à 100% pour les trois classes d'actif). En cas de calcul exact de SCR, cinq simulations différentes (cas central dans deux environnements financiers et trois cas choqués) sont nécessaires pour un calcul exact, soit en tout 8000 simulations. Si chaque simulation s'effectue sur 2000 scenarios stochastiques, on arrive en tout à 16 millions de projections à réaliser.

Dans la pratique, des simplifications sont nécessaires pour les calculs pour limiter le temps de run du modèle. Les équipes ALM recourent à plusieurs procédés :

- limitation du nombre d'allocations testées par jugement d'expert : il s'agit d'identifier les allocations réalistes et d'exclure les allocations qui ne semblent pas implémentables dans un horizon court terme (entre 3 et 5 ans) du fait de la capacité à acheter ou vendre sur le marché des volumes

importants.

- limitation du nombre d'allocations testées par méthode itérative : il s'agit de commencer par tester des allocations avec un pas élevé entre deux tests, puis une fois les résultats obtenus, tester de nouvelles allocations avec un pas plus faibles, mais dans une zone ciblée.

- limitation du nombre de runs par estimation du niveau de SCR : à partir du niveau de SCR du cas central, il s'agit d'estimer l'évolution du niveau de SCR pour chaque nouvelle allocation d'actif à l'aide de drivers prédéfinis.

Dans cette étude, des drivers ont pu être définis grâce à l'étude réalisée précédemment :

- prise en compte du niveau de SCR marché calculé à t=0 avec l'allocation réelle,

- pour la projection, intègration d'un facteur d'évolution du portefeuille (projection en *run-off*).

- pour chaque brique de risque, le SCR évolue également en fonction d'un niveau d'allocation sur la classe d'actif. D'après les résultats présentés à la figure 2.2, on peut estimer un coefficient propre à la classe d'actif sur le portefeuille étudié à appliquer à une variation d'allocation sur l'actif considéré, ce qui donne les formules ci-après :

$$SCR_{action}(t) = SCR_{action}(t-1) \times (1 + Delta_{allocation_BC} \times \theta) \times \frac{PM(t)}{PM(0)},$$
$$SCR_{immobilier}(t) = SCR_{immobilier}(t-1) \times (1 + Delta_{allocation_BC} \times \lambda) \times \frac{PM(t)}{PM(0)}.$$

Par exemple sur le portefeuille étudié, on mesure le coefficient θ à 5,1 et le coefficient λ à 4,4.

D'après l'étude précédente, le coefficient sur les taux est évalué à $-1/2$, mais il paraît pertinent d'intégrer également un facteur lié au gap de duration du portefeuille, défini comme l'écart de duration actif-passif ce qui donne la formule suivante :

$$SCR_{taux}(t) = SCR_{taux}(t-1) \times \gamma \times \frac{gap - duration(t)}{gap - duration(0)} \times \frac{PM(t)}{PM(0)},$$

avec $\gamma = -0,5$.

Dans le modèle retenu, ces drivers sont utilisés pour estimer le niveau du SCR marché à chaque pas de temps.

Chapitre 3

Du *reinforcement learning* à la gestion ALM

3.1 Principes du *reinforcement learning*

3.1.1 Définitions et généralités

Le *reinforcement learning* (RL) ou apprentissage par renforcement est un domaine de l'apprentissage automatique qui vise à résoudre des problèmes de prises de décision d'un agent en interaction avec un environnement dans le but de maximiser des objectifs long-terme. Un problème de *reinforcement learning* est constitué d'un environnement, généralement incertain, et d'un agent qui évolue dans celui-ci en cherchant à atteindre un objectif défini explicitement. L'agent n'a pas de connaissance *a priori* de l'environnement et le découvre au fil des interactions qu'il a avec lui. Il ne connaît pas non plus la stratégie gagnante.

Les principaux éléments d'un problème de *reinforcement learning* sont les suivants :

— **Etat du système** (*state*) : vecteur de données représentant l'état du système. Il peut donner une information complète sur l'environnement (ex : jeu de puissance 4) ou bien le résumer à travers une sélection de variables.

— **Politique** (*policy*) : comportement de l'agent à un instant donné qui associe une action à l'état de l'environnement qu'il observe. Il peut s'agir d'une simple table de correspondance ou de fonctions déterministes ou stochastiques.

— **Fonction de récompense** (*reward*) : impact instantané du changement d'état de l'environnement sur l'agent. Celui-ci provient directement de l'environnement et est subi par l'agent. Il s'agit d'un stimulus semblable à la sensation biologique de plaisir ou de douleur.

— **Fonction de valeur** (*value*) : valeur de l'état de l'environnement pour l'agent. Il s'agit de l'évaluation par l'agent du bénéfice futur qu'il a de l'état de l'environnement actuel.

Ce cadre théorique peut s'appliquer à de nombreux problèmes, par exemple un robot cherchant à sortir d'un labyrinthe en minimisant le nombre de ses déplacements, un enfant apprenant à marcher, ou encore le jeu de Tetris détaillé dans la section 3.1.3. Le *reinforcement learning* est une branche de l'apprentissage automatique distincte de l'apprentissage supervisé et de l'apprentissage non-supervisé. En effet, dans un cadre d'apprentissage supervisé, nous devons posséder en amont une connaissance des décisions passées optimales que nous essayons de généraliser à des états du système que nous n'avons pas encore vus. Ce n'est pas le cas dans un problème de *reinforcement*

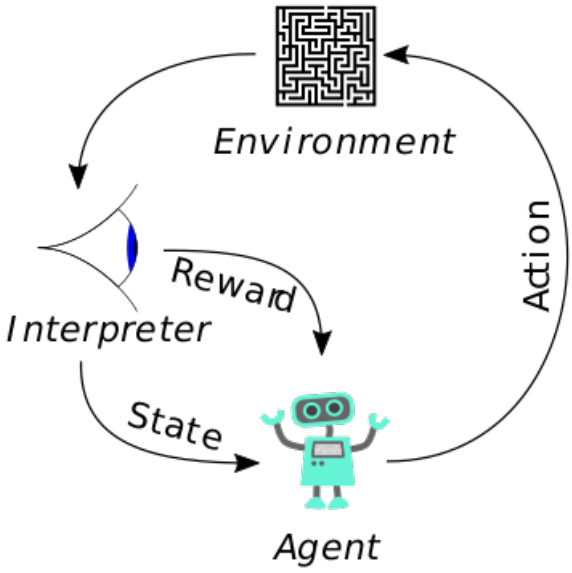


FIGURE 3.1 – Illustration du cadre général de *reinforcement learning*

	Apprentissage supervisé	Apprentissage non supervisé	Reinforcement learning
Données nécessaires	Données avec une variable cible	Données sans variable cible	Données synthétiques créés via les interactions d'un agent avec un environnement
Objectif	Prédire la variable cible pour de nouvelles données sur la base de ce qui a été appris à travers les données d'entrainement	Identifier des groupes de données similaires	Identifier une séquence d'actions optimales pour atteindre un objectif donné
Flexibilité	Il est nécessaire d'entrainer à nouveau le modèle lorsque les données changent	Il est nécessaire d'entrainer à nouveau le modèle lorsque les données changent	L'agent peut s'adapter à un léger changement d'environnement

TABLE 3.1 – Différentes approches d'apprentissage automatique

learning où l'agent n'a aucune connaissance de l'environnement au début de l'entraînement et cherche à atteindre un objectif long-terme au lieu d'optimiser chaque action individuellement. Dans un problème d'apprentissage non supervisé, l'algorithme cherche à trouver des structures au sein d'un espace de données pour les classifier. Bien que la résolution d'un problème de *reinforcement learning* s'appuie sur la recherche de structures au sein de l'environnement, la simple connaissance de celles-ci n'est pas suffisante pour définir une stratégie permettant d'atteindre l'objectif.

Le *reinforcement learning* a une double origine avec d'un côté les problèmes de contrôle optimal et de programmation dynamique et l'apprentissage automatique de l'autre. Le terme de renforcement est issu du cadre de l'apprentissage animal et via une traduction en anglais des travaux de Pavlov sur le conditionnement. A partir des années 1980, le développement des réseaux de neurones artificiels et leur rapprochement avec le *reinforcement learning* a créé le sous-domaine du *Deep Reinforcement Learning* dont le premier succès fut TD-Gammon, un algorithme développé en 1992 par Gerald Tesaura utilisant l'approche de différence temporelle et un réseau de neurones pour jouer au backgammon. Le *Deep Reinforcement Learning* a acquis une notoriété auprès du grand public en 2015 lorsque l'algorithme AlphaGo développé par Google/DeepMind est devenu le premier algorithme à battre un joueur de Go professionnel. On constate depuis une poursuite du développement de cette branche avec de nombreuses applications naissantes par exemple dans le domaine de la conduite autonome.

FIGURE 3.2 – Carde du RL appliqué au jeu Tetris

3.1.2 L'équilibre exploration - exploitation

L'un des principaux problèmes soulevés par le cadre de *reinforcement learning* est l'équilibre entre exploration et exploitation. Lors de l'entraînement d'un agent, il est d'un côté nécessaire que celui-ci applique des actions qui génèrent des récompenses pour renforcer la place de celles-ci dans sa stratégie. Il s'agit de l'exploitation des connaissances déjà acquises. Toutefois, il faut également s'assurer que l'agent parcoure tout l'espace des actions disponibles dans différentes situations afin de tester de nouvelles actions qui pourraient s'avérer plus rémunératrices que la meilleure action connue à ce stade. C'est l'exploration de nouvelles options dont la valeur sera augmentée si elles aboutissent à des issues positives ou diminuée dans le cas inverse. Pour illustrer ce concept, on peut imaginer qu'il faut aller d'un point A à un point B tous les jours et qu'un chemin convenable en un temps t est déjà connu. Il est intéressant de tester un autre chemin qui n'a jamais été emprunté. Si celui-ci est plus rapide, il permettra de gagner tous les jours le temps économisé. Si celui-ci est plus long, le temps perdu sur ce trajet ne se produira qu'une fois et le chemin initial pourra être emprunté à nouveau dès la fois suivante. D'un point de vue pratique, il est important de forcer un algorithme de *reinforcement learning* à explorer en phase d'apprentissage car il risque de s'enfermer dans des solutions suboptimales (optimums locaux) déterminées par l'état initial (aléatoire) de ses paramètres.

3.1.3 Exemple d'application - le jeu Tetris

Tetris un célèbre jeu vidéo développé par l'ingénieur soviétique Alexey Pajitnov en 1984 où le joueur doit placer des blocs (tetrominoes) de façon à former des lignes complètes dans le champ de jeu.

Ce jeu se prête bien au cadre de *reinforcement learning* défini dans la partie précédente. Il s'agit en effet d'un problème où la fonction objectif est clairement définie, où le nombre d'actions est limité et où l'environnement est observable. Pour chacun des éléments du problèmes, plusieurs implémentations sont possibles pour sa résolution.

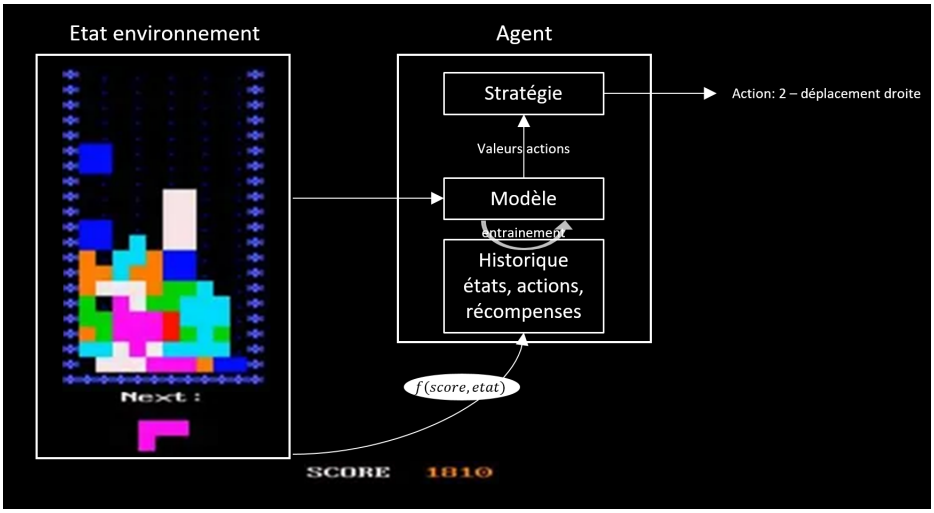


FIGURE 3.3 – Déroulement d'une prise de décision par l'agent lors d'une partie de Tetris

Voici quelques exemples :

Environnement :
- Implémentation 1 : image en noir et blanc de l'espace de jeu. L'environnement est une matrice de $\{0,1\}^{k*l}$ dont les éléments valent 0 si le pixel est noir et 1 sinon.
- Implémentation 2 :
  - Type de bloc en déplacement, position et orientation de celui-ci ; vecteur d'entiers de dimension 4 ou vecteur booléen de dimension supérieure
  - Type du bloc suivant : entier de 1 à 7
  - Matrice donnant l'occupation de l'espace de jeu par des blocs immobiliers : matrice de booléens de taille 10×20

Fonction de récompense : $reward = \Delta score - \lambda \delta_{game-over}$
- Impact positif de l'évolution du score suite à l'action
- Pénalisation si l'action entraîne la fin de la partie

Actions : entier entre 1 et 4 ; en supposant que l'accélération de la chute n'a pas d'impact sur le score :

1. aucune action
2. déplacement à gauche
3. déplacement à droite
4. rotation du bloc

Une fois ces éléments mis en place, l'idée est d'entraîner un agent qui à chaque instant observe l'environnement, calcule la valeur de chacune des actions possibles et en sélectionne une. La stratégie sera différente en phase d'entraînement et en phase de test de l'algorithme. En phase de test, deux stratégies peuvent être envisagées : séléctionner l'action avec la valeur la plus importante ou bien tirer aléatoirement une des actions avec des probabilités proportionelles aux valeurs des actions. En phase d'entraînement, il est nécessaire d'ajouter de l'exploration, en prévoyant un certain pourcentage des décisions aléatoires au cours de l'apprentissage.

3.1.4 Formalisation d'un problème de *reinforcement learning*

Les principes du *reinforcement learning* sont décrits dans l'ouvrage Sutton & Barto (2018). Les auteurs présentent les bases de la modélisation, qui s'appuient sur des processus de décision markoviens. Ces derniers donnent un cadre cohérent aux briques de bases définies précédemment (état du système, politique, fonction de récompense, fonction de valeur).

Un processus de décision markovien est un quadruplet $\{S, A, T, R\}$ définissant :

— un **ensemble d'états** $\mathcal{S}$, qui peut être fini, dénombrable ou continu ; cet ensemble définit l'environnement tel que perçu par l'agent.

— un **ensemble d'actions** $\mathcal{A}$, qui peut être fini, dénombrable ou continu et dans lequel l'agent choisit les interactions qu'il effectue avec l'environnement.

— une **fonction de transition** $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0; 1]$. cette fonction définit l'effet des actions de l'agent sur l'environnement. En particulier, $\mathcal{P}^a_{ss'} = \mathbb{P}(S_{t+1} = s' | S_t = s, A_t = a)$ et caractérise la probabilité de se retrouver dans l'état s' après avoir effectué l'action a dans l'état s.

— une **fonction de récompense** $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$. Elle définit la récompense (positive ou négative) reçue par l'agent. En particulier, $R(s, a, s)$ est la récompense obtenue en $t+1$ pour être passé de l'état s à s' en ayant effectué l'action a.

Les processus de décision markoviens caractérisent parfaitement les modèles à informations parfaites pour lesquelles l'information de l'état actuel résume parfaitement l'historique des actions antérieures. Ces derniers vérifient :

$$\mathbb{P}(R_{t+1} | S_1, A_1, S_2, ..., A_{t-1}, S_t) = \mathbb{P}(R_{t+1} | S_t).$$

Sutton et Barto affirment que si cette propriété n'est pas parfaitement vérifiée pour le problème étudié, le modèle peut toutefois partir de l'information contenue dans l'état courant et en tirer son apprentissage (Chapitre 3.5). Un problème se rapprochant de cette hypothèse obtiendra donc de meilleures performances. Dans le cas présent, il est possible de penser qu'on peut appliquer la théorie markovienne à ce problème dans la mesure où cette hypothèse est couramment utilisée pour la valorisation financière.

Formellement on peut définir le cadre général de la manière suivante : à chaque étape de décision $t \in [1, N]$, l'agent se trouve à un état précis $S_t \in \mathcal{S}$ où lui est permis un ensemble d'actions $A_t \in \mathcal{A}(S_t)$. En fonction de l'action choisie en t, l'agent perçoit à l'étape d'après (en $t + 1$), une récompense $R_t \in \mathbb{R}$ et se retrouve dans un nouvel état du jeu S_{t+1}.

Une politique décrit les choix des actions à jouer par l'agent dans chaque état. Formellement, il s'agit donc d'une fonction $\pi : \mathcal{S} \to \mathcal{A}$ dans le cas d'une politique déterministe ou $\pi : \mathcal{S} \times \mathcal{A} \to [0; 1]$ dans le cas stochastique. On note $\pi(a|s)$ la probabilité de jouer a dans l'état s à l'instant t, i.e. $P[A_t = a | S_t = s]$.

L'agent choisit une politique à l'aide de la fonction de récompense R. Notons $R_t = R(S_t, A^\pi_t, S_{t+1})$, la récompense effective obtenue après avoir effectué l'action a par l'agent qui suit la politique π. Nous rappelons que l'agent cherche à maximiser son gain sur le long terme. Voici plusieurs critères d'intérêts que l'agent peut chercher à maximiser :

— $\mathbb{E}_\pi \left[\sum_{t=0}^h r_t \right]$: espérance de la somme des récompenses à un horizon fini fixé h,

— $\liminf_{h \to +\infty} \mathbb{E}_\pi \left[\frac{1}{h} \sum_{t=0}^h r_t \right]$ ou $\limsup_{h \to +\infty} \mathbb{E}_\pi \left[\frac{1}{h} \sum_{t=0}^h r_t \right]$: récompense moyenne à long terme,

— $\mathbb{E}_\pi\left[\sum_{t=0}^{\infty}\gamma^t r_t\right]$: récompense escomptée (ou amortie) à horizon infini où $0 \leq \gamma < 1$.

Dans la suite nous maximiserons la récompense escomptée à horizon infini où $0 < \gamma < 1$. γ traduit la préférence de l'agent pour le présent. Ainsi si $\gamma \longrightarrow 0$ cela signifie que l'agent manifeste une préférence pour le présent. Inversement si $\gamma \longrightarrow 1$, l'agent accorde presque autant d'importance au futur qu'au présent.

Lorsqu'une politique et un critère sont déterminés, deux fonctions centrales peuvent être définies :

— $V^\pi : S \to \mathbb{R}$: c'est la fonction de valeur des états ; $V^\pi(s)$ représente le gain (selon le critère adopté) engrangé par l'agent s'il démarre à l'état s et applique ensuite la politique π ad infinitum. Dans le cas de gains escomptés on a $V^\pi(s) = \mathbb{E}\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}|S_t = s\right]$.

— $Q^\pi : S \times A \to \mathbb{R}$: c'est la fonction de valeur des états-actions ; $Q^\pi(s,a)$ représente le gain engrangé par l'agent s'il démarre à l'état s et commence par effectuer l'action a, avant d'appliquer ensuite la politique π ad infinitum. Dans le cas de gains escomptés on a : $Q^\pi(s,a) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}|S_t = s, A_t = a\right]$.

V^π et Q^π sont deux fonctions intimement liées. Elles vérifient la relation suivante :

$$\forall s \in \mathcal{S},\ V^\pi(s) = Q^\pi(s,\pi(s)) := \sum_{a\in\mathcal{A}(s)} \pi(a|s)Q^\pi(s,a) \tag{3.1}$$

Equations de Bellman :
L'utilisation des processus de Markov pour la modélisation de comportements rationnels réside dans les équations de Bellman qui exhibent une relation de récurrence entre $V^\pi(S_t)$, $V^\pi(S_{t+1})$ et $Q^\pi(S_t, A_t)$, $Q^\pi(S_{t+1}, A_{t+1})$ comme suit :

Equation de Bellman

$\forall s \in \mathcal{S}, \forall a \in \mathcal{A}$, on a :

$$Q^\pi(s,a) = R(s,a) + \gamma \sum_{s'\in\mathcal{S}} P^a_{ss'} \sum_{a'\in\mathcal{A}} \pi(a'|s')Q^\pi(s',a') \tag{3.2}$$

$$V^\pi(s) = \sum_{a\in\mathcal{A}} \pi(a|s)\big(R(s,a) + \gamma \sum_{s'\in\mathcal{S}} P^a_{ss'} V^\pi(s')\big) \tag{3.3}$$

En effet $\forall s \in \mathcal{S}, \forall a \in \mathcal{A}$, on a :

$$\begin{aligned} Q^\pi(s,a) &= \sum_{s'\in S}[R(s,a,s') + \gamma V^\pi(s')]\mathcal{P}^a_{ss'} \\ &= R(s,a) + \sum_{s'} \mathcal{P}^a_{ss'} V^\pi(s') Q^\pi(s,a) \\ &= R(s,a) + \gamma \sum_{s'\in\mathcal{S}} P^a_{ss'} \sum_{a'\in\mathcal{A}} \pi(a'|s')Q^\pi(s',a'). \end{aligned}$$

On déduit (3.3) de (3.1) et (3.2). Ces équations sont primordiales à la résolutions de problème de *reinforcement learning* car elles permettent :

1. D'évaluer la politique mise en oeuvre par l'agent
2. D'améliorer cette politique.
3. De résoudre notre problème d'optimisation

La stratégie naïve π qui s'impose en phase d'exploitation est celle qui maximise la fonction de valeur est maximale : $\pi_t(s) = \underset{a}{\operatorname{argmax}}\, Q_t^\pi(s,a)$. En cas d'exploration, l'agent choisit une action aléatoire. Il est cependant possible de définir d'autres stratégies.

3.2 Recherche de la décision optimale par un réseau de neurones

Une fois l'équation de Bellman établie, il faut également s'assurer de la capacité à calculer la fonction de valeur au fur et à mesure de l'apprentissage. Dans le cas d'un problème à dimensionnalité finie et limitée, il est possible de calculer directement la fonction de valeur Q associant une valeur réelle à toute combinaison action/état du système. A chaque fois que l'algorithme teste un couple (état,action) donné, il met à jour la fonction de valeur avec la nouvelle récompense reçue pour cette dernière. Cependant, beaucoup de problèmes que s'attachent à résoudre le *reinforcement learning* se définissent avec un caractère continu de l'espace des états ou des actions, ce qui ne permet plus raisonnablement de calculer une fonction de valeur pour chaque couple (état,action). Cela demanderait une mémoire beaucoup trop importante et un jeu de test également très important.

L'introduction d'un réseau de neurones va alors permettre d'approximer la fonction de valeur $Q^\pi(s,a)$ au fur et à mesure de l'apprentissage. Grâce au théorème d'approximation universelle (voir section 3.2.2), le réseau de neurone multi-couche s'avère être un candidat pertinent pour réaliser cette approximation.

En entrée du réseau, les valeurs de transition du système sont fournies : pour chaque état du système, l'action $a \in A_t$ réalisée, le niveau de *reward* reçu et le nouvel état du système après l'action.

Puis le réseau calcule en sortie la fonction de coût qu'elle va minimiser :

$$L(\theta) = \mathbb{E}\Big(L\big(Q^\pi(s,a;\theta), \hat{Q}^\pi(s,a;\theta)\big) \mid (s,a,r,s')\Big).$$

3.2.1 Principes généraux d'un réseau de neurones simple : le perceptron

Inspirés de la dynamique biologique des neurones du cerveau humain qui reçoivent des stimulis électriques leur permettant de transmettre une information, les réseaux de neurones se sont développés en *machine learning* pour résoudre des problèmes d'apprentissage complexes.

Le perceptron permet de formaliser le principe général sous-jacent des réseaux de neurones.

Un problème de classification peut se définir à l'aide des composantes suivantes :

— Un N-échantillon de données $((x_i, y_i))_i \in [\![1 : N]\!]$, dont les x_i représentent les données d'entrée et les y_i les valeurs de sorties que l'on souhaite estimer.
— Une fonction de prédiction paramétrée $g(x;\theta)$, $\theta \in \mathcal{R}^n$ qui prend en entrée les x_i et qui cherche à prédire la valeur des y_i.
— Une fonction de coût L dont le rôle est de quantifier l'erreur de prédiction et donc la qualité d'estimation.

Le neurone artificiel peut être caractérisé par la fonction suivante :

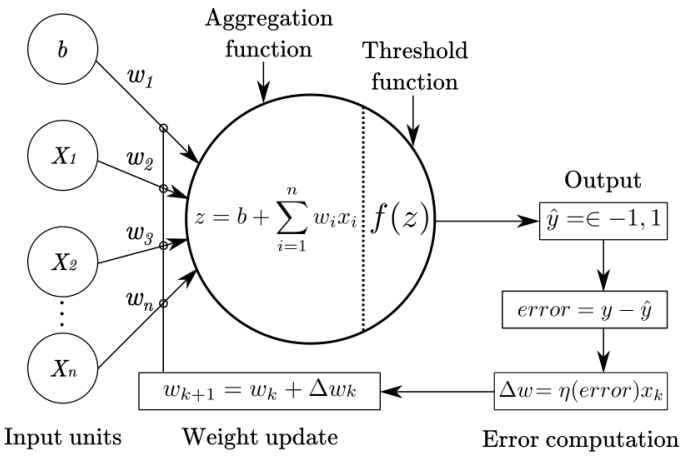


FIGURE 3.4 – Fonctionnement d'un perceptron

$$P : \mathcal{R}^n \longrightarrow \mathcal{R}$$
$$x = (x_1, ..., x_n) \longrightarrow f(\Sigma_i x_i w_i + b)$$

Les poids $w_i \in \mathbb{R}$ sont les paramètres caractérisant le perceptron et affectés à chaque signal d'entrée. Au niveau du perceptron la fonction d'agrégation est une fonction linéaire de chaque neurone de la couche d'entrée affectés des poids w_i.

Le traitement des signaux munis de leur poids altéré d'un biais, $\Sigma_i x_i w_i + b$, est modélisé par une fonction f appelée fonction d'activation.

Puis l'objectif est d'obtenir l'estimation la plus précise des poids $w_i \in \mathbb{R}$ qui minimisent l'erreur moyenne de prédiction et donc de trouver $\hat{\theta}$ en résolvant :

$$\hat{\theta} \in \arg\min_{\theta} \mathbb{E}[L(f(x;\theta), y)]$$

.

Cette estimation est réalisée via un processus d'apprentissage itératif illustré par la figure 3.4 et détaillé en 3.2.3 permettant de converger vers une solution du problème.

3.2.2 Le réseau de neurones multi-couches

L'intérêt des réseaux de neurones réside dans le fait de pouvoir concaténer un nombre important de neurones simples pour former un réseau capable de modéliser des problèmes plus complexes (théorème d'approximation). Dans le cadre de cette étude, le perceptron multi-couches représenté sur la figure 3.5 a servi de base au réseau de neurones modélisé.

La figure 3.5 représente un réseau à 3 couches. 1 couche d'entrée, une couche cachée et une couche de sortie avec respectivement M, N et K neurones. Chaque neurone de la couche cachée $y_n^1, n \in [\![1 : N]\!]$ est le résultat d'un perceptron dont les inputs sont les neurones de la couche d'entrée $y_m^0, m \in [\![1 : M]\!]$ et munis des poids $W^0 = (w_{nm}^0), n \in [\![1 : N]\!], m \in [\![1 : M]\!]$. Idem pour les neurones de la couche de sortie : chaque neurone de la couche de sortie est le résultat d'un perceptron

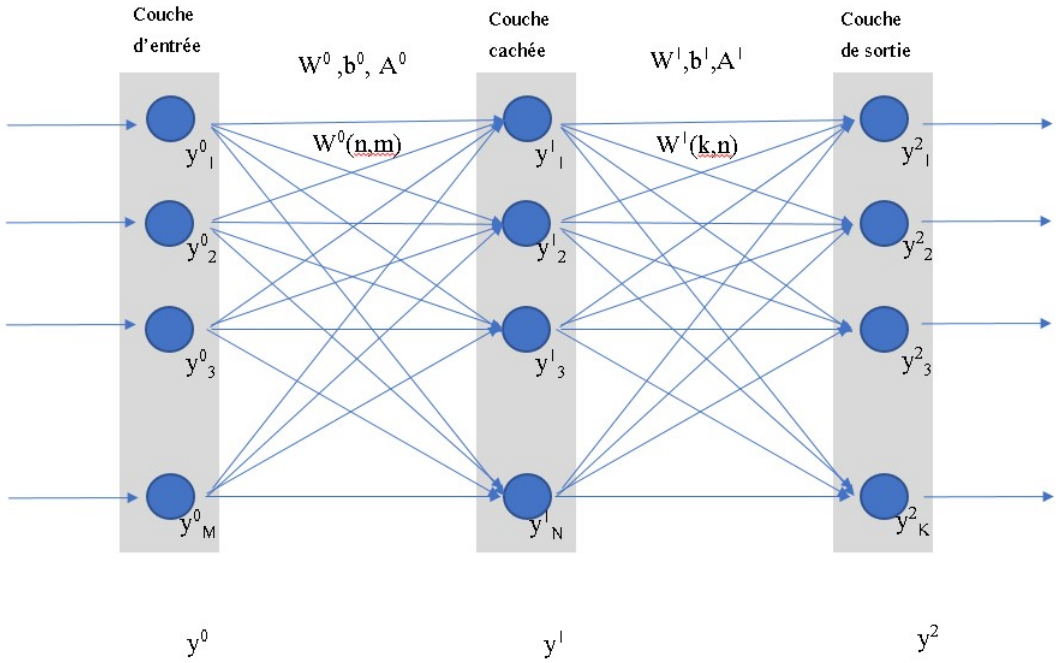


FIGURE 3.5 – Perceptron multicouches

dont les inputs sont ceux en sortie de la couche cachée $y_n^1, n \in [\![1:N]\!]$ et munis des poids $W^1 = (w_{kn}^1), k \in [\![1:K]\!], n \in [\![1:N]\!]$. Autrement dit on a :

$$\begin{cases} y_n^1 = N_{entrée}(y^0) = \mathcal{A}_{entrée}^0(\Sigma_m y_m^0 w_{nm}^0 + b_n^0) \\ y_k^2 = N_{cachée}(y^1) = \mathcal{A}_{cachée}^1(\Sigma_k y_k^1 w_{kn}^1 + b_k^1) \end{cases} \tag{3.4}$$

Cette notation a l'avantage de montrer le lien qui existe entre le perceptron et le perceptron multicouche, on lui préfèrera son écriture matricielle afin de simplifier les calculs dans les parties suivantes.

Le théorème d'approximation universelle, ou théorème de Cybenko, nous assure que toute fonction continue peut être approximée avec une précision arbitraire par un réseau de neurones à trois couches.

Soit ϕ une fonction continue, non-constante, bornée et croissante. Notons par I_m l'hypercube unitaire de dimension m $[0,1]^m$. L'espace des fonctions continues sur I_m est noté $C(I_m)$. $\forall f \in C(Im)$ et $\epsilon > 0, \exists N \in \mathbb{N}, (v_i, b_i) \in \mathbb{R}^2$ et $w_i \in \mathbb{R}^m$, où $i \in \{1, \cdots, N\}$ tel que :

$$F(x) = \sum_{i=1}^{N} v_i \phi(w_i^T x + bi) \tag{3.5}$$

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

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

En d'autres termes, les fonctions de la forme F(x) sont denses dans C(Im). Cela reste vrai lorsque nous remplaçons Im par n'importe quel sous espace compact de $\mathbb{R}^m$.

L'erreur uniforme converge vers 0 lorsque la taille de la couche cachée tend vers l'infini.

Chaque couche de neurones peut contenir autant de neurones qu'on le souhaite ; de même, il est possible de paramétrer un type de fonction d'activation différente en sortie de chaque couche.

Dans cette étude les performances de l'apprentissage de l'agent sur plusieurs structures de réseaux de neurones ont pu être testées.

3.2.3 Rétro-propagation

Au départ d'une simulation, les poids du réseau de neurones sont définis aléatoirement. Une fois le réseau de neurones stimulé par un nouvel état du système, ce dernier va mettre à jour tous les poids de tous les neurones pour minimiser la fonction de coût en sortie du réseau.

La méthode la plus répandue pour cette opération est la descente de gradient. L'idée est de partir d'un point x_0 suffisamment proche du minimiseur x^*. On cherche à construire une suite (x_n) telle que $x_{n+1} = x_n + \eta h_n$, avec $h \in \mathbb{R}^n$ et $\eta \in \mathbb{R}$ et $F(x_{n+1}) < F(x_n)$. on dira de (h_n) qu'il s'agit d'une suite de direction de descente. La suite $F(x_n)$ est donc décroissante et bornée inférieurement et convergera vers une valeur qu'on espère être $F(x^*)$.

Pour permettre cette itération décroissante, on s'appuie sur le gradient. Il correspond à la plus grande direction de descente lors de chaque itération. En effet, le développement de Taylor à l'ordre 2 d'une fonction de classe C^2 appliqué à $x + \mu h$ donne la formule suivante :

$$f(x + \mu h) = f(x) + \mu < \nabla f(x), h > + \mu^2 h^\intercal H_f(x) h + \circ(h^2),$$

où $\nabla f(x) \in \mathbb{R}^n$ est le gradient de F et, $H_f \in \mathcal{M}(\mathbb{R})$ est la hessienne de f.

D'après l'inégalité de Cauchy-Schwartz on a :

$$| < \nabla f, h > | \leq \|\nabla f\| \|h\|,$$
$$et | < \nabla f, h > | = \|\nabla f\| \|h\| \Longleftrightarrow h = \pm \nabla f(x).$$

On peut donc en déduire l'encadrement suivant :

$$-\|\nabla f\| \|h\| \leq < \nabla f, h > \leq \|\nabla f\| \|h\|.$$

Puis en gardant la borne inférieure, on a $\forall h$:

$$\begin{aligned} f(x+\mu h) &\geq f(x) - \mu\|\nabla f(x)\|\|h\| + \mu^2 h^\intercal H_f(x)h + \circ(h^2) \\ &\geq f(x) - \mu\|\nabla f(x)\|^2 + \mu^2 h^\intercal H_f(x)h + \circ(h^2) \\ &\geq f(x+\mu(-\nabla f(x))). \end{aligned}$$

Ainsi en posant $h_n = -\nabla f(x_n)$, on a $x_{n+1} = x_n - \mu\nabla F(x_n)$ et $\forall h \in \mathbb{R}^n$ $f(x_{n+1}) \leq f(x_n + \mu h)$ ce qui assure donc à chaque étape entre n et n+1 de se rapprocher du minimum local.

Apprentissage des poids du réseau de neurones

Le principe de l'algorithme de rétro-propagation est de recalculer les poids du réseau en partant des neurones les plus proches de la sortie et en appliquant la méthode du gradient : le nouveau poids correspond au poids précédent auquel on soustrait le gradient affecté d'un paramètre $\mu \geq 0$ dit taux d'apprentissage. Les poids sont remis à jour jusqu'à ce qu'on atteigne une erreur minimale (qui peut correspondre à un minimum local car cette méthode ne permet pas de calculer un minimum absolu).

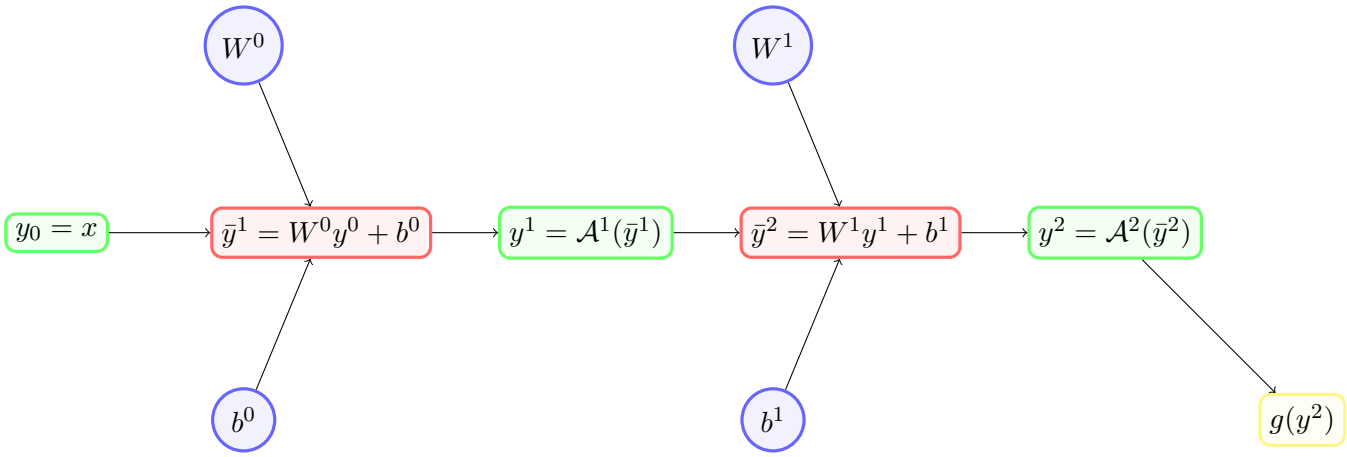


Diagram 3.1 – Graphe computationnel pour un réseau de neurones à 3 couches

Ci-dessus nous avons le graphe computationnel d'un réseau à 3 couches (dont une cachée) :
- en vert : sont représentés les inputs de chaque couche avant application de la fonction d'activation,
- en rouge : sont représentés les outputs de chaque couche,
- en bleu : sont représentés les paramètres de chaque couche,
- en jaune : est représentée l'erreur d'estimation du réseau.

Ce graphe nous permet de décrire le séquençage machine des calculs réalisés par un réseau de neurones. Les calculs entre les couches de la figure 3.5 sont réecris comme des calculs matriciels. En effet, la couche d'entrée prend un vecteur de taille M et en retourne un de taille N par une operation affine. Ce qui équivaut à appliquer une matrice à coefficients réels de taille $N \times M$ sommé d'un vecteur de taille N. Soit H le nombre de couches, pour chaque couche $h \in [\![1, H]\!]$, on a :
- $W^h = (w^h_{ij})_{(i,j)\in[\![1:dim_{output}]\!]\times[\![1:dim_{input}]\!]}$
- $b^h = (b^h_i)_{i\in[\![1,dim_{output}]\!]}$
- $\mathcal{A}^h : \mathbb{R}^{dim_{output}} \to \mathbb{R}^{dim_{output}}$, $x = (x_i)_i \mapsto \mathcal{A}^h(x) = (a^h(x_i))_i$

Cette représentation matricielle des calculs est particulièrement utile afin d'écrire les équations de propagation rétrograde qui permettent d'actualiser les poids b^h et W^h de notre réseau de neurones

synonyme d'apprentissage. En effet, d'après l'algorithme de la descente de gradient on a :

$$w_{ij}^h = w_{ij}^h - \mu \frac{\delta \mathcal{L}}{\delta w_{ij}^h} \tag{3.6}$$

$$b_i^h = b_i^h - \mu \frac{\delta \mathcal{L}}{\delta b_i^h} \tag{3.7}$$

Ainsi pour actualiser les poids du réseau de neurones il faut déterminer les quantités $\frac{\delta \mathcal{L}}{\delta b^h}$ et $\frac{\delta \mathcal{L}}{\delta W^h}$ décrites par les équations de propagation rétrogrades suivantes. Par souci de clarté nous nous restreignons au cas $H = 2$ et à l'utilisation d'une écriture symbolique des dérivées. Le lecteur désireux d'obtenir plus de détails concernant le calcul exact des dérivées intermédiaires pourra se référer à l'article de Nielsen (2018) :

$$\frac{\delta \mathcal{L}}{\delta W^1} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta W^1} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta W^1} \tag{3.8}$$

$$\frac{\delta \mathcal{L}}{\delta b^1} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta b^1} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta b^1} \tag{3.9}$$

$$\frac{\delta \mathcal{L}}{\delta W^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta W^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta W^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta y^1} \cdot \frac{\delta y^1}{\delta W^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta y^1} \cdot \frac{\delta y^1}{\delta \bar{y}^1} \cdot \frac{\delta \bar{y}^1}{\delta W^0} \tag{3.10}$$

$$\frac{\delta \mathcal{L}}{\delta b^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta W^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta b^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta y^1} \cdot \frac{\delta y^1}{\delta b^0} = \frac{\delta \mathcal{L}}{\delta y^2} \cdot \frac{\delta y^2}{\delta \bar{y}^2} \cdot \frac{\delta \bar{y}^2}{\delta y^1} \cdot \frac{\delta y^1}{\delta \bar{y}^1} \cdot \frac{\delta \bar{y}^1}{\delta b^0} \tag{3.11}$$

Le graphe computationnel des équations rétrogrades est représenté ci-dessous :

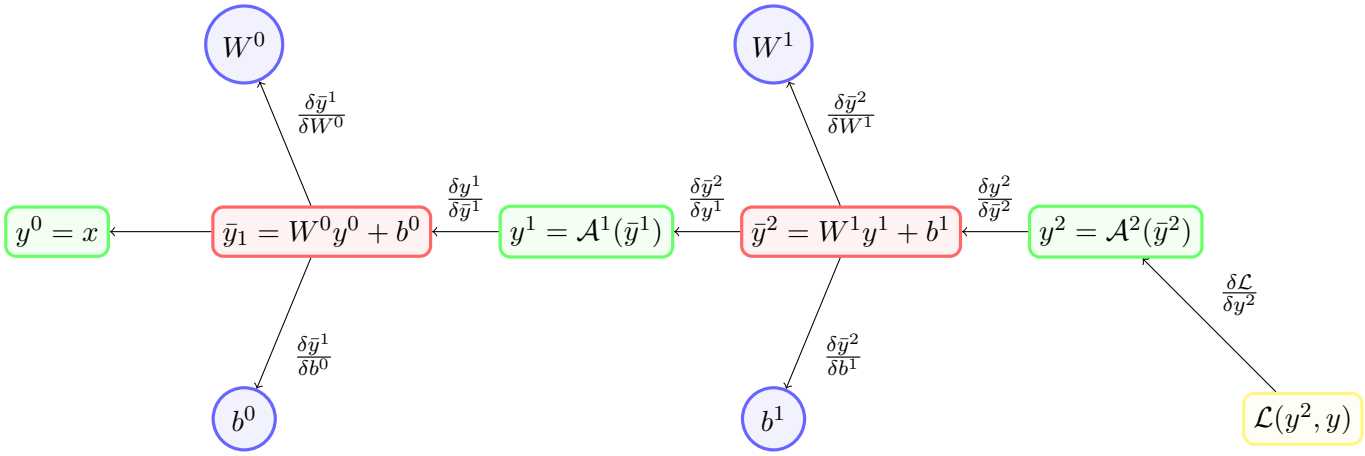


Diagram 3.2 – Graphe computationnel des équations rétrogrades

Une fois les dérivées intermédiaires calculées, on obtient le gradient qui sert à l'actualisation des poids du réseau de neurones. On applique donc l'algorithme de descente de gradient jusqu'à sa convergence en un minimum local. En pratique, les calculs permettant cette convergence peuvent s'avérer coûteux, c'est pourquoi les algorithmes de descente de gradients comportent un second critère d'arrêt fondé sur un nombre maximal d'itérations à effectuer. Le choix de cet hyperparamètre est à définir par l'utilisateur au même titre que le pas de descente μ, le nombre de couches ainsi que le nombre de neurones par couche.

3.3 Présentation de l'ALM sous forme d'un problème de *reinforcement learning*

De Tetris à la gestion actif-passif

Etudiant les interactions entre les actifs et les passifs, l'ALM est au cœur de toute activité d'assurance. Son objectif est de définir une stratégie d'investissement permettant de maximiser les marges futures de l'assureur sous contraintes de risque. Cette stratégie fournit une allocation stratégique ou « SAA » (*Strategic Asset Allocation*) qui est ensuite allouée par les équipes chargées des investissements. Les études ALM sont effectuées régulièrement, environ tous les deux ans pour chaque portefeuille chez AXA France, afin de mettre à jour ces allocations en fonction des changements de conditions de marché, de l'évolution des actifs et du passif. Dans cette approche, nous nous plaçons dans la peau d'un directeur des investissements d'une compagnie d'assurance gérant un portefeuille de retraite en run-off. Celui-ci s'appuie sur les études de son équipe ALM pour prendre une décision de gestion de son portefeuille d'actifs auquel sont adossés ces contrats de retraite. Ainsi, il peut être vu comme un agent d'un jeu qui vise à gérer les actifs de la compagnie pour atteindre des objectifs fixés par son actionnaire. Les parallèles avec Tetris évoqué précédemment sont multiples :

- **Aléa** : comme pour les nouveaux blocs dans Tetris, un aléa impacte l'évolution du jeu. Différents scénarios économiques peuvent impacter les valorisations et les rendements des actifs alors que des rachats dynamiques peuvent se déclencher au passif.
- **Espace d'action discret** : les contraintes opérationnelles ne donnent à l'assureur qu'un nombre limité d'actions possible. Il est en effet impossible de changer radicalement le portefeuille à cause de frais de transaction et surtout de la liquidité des marchés. Ainsi on ne peut qu'effectuer des pas dans différentes directions en ajustant l'allocation et le gap de duration.
- **Observation de l'intégralité du système** : l'agent dispose de toutes les informations sur les actifs et les passifs sous gestion.
- ***Reward* explicite** : bien qu'il s'agisse d'une simplification de la réalité, la gestion de portefeuilles d'assurances peut être modélisée comme une maximisation d'une métrique de rendement pour les actionnaires normalisé par le risque pris.
- ***Game over*** : un parallèle peut être fait avec la fin d'une partie de Tetris et la faillite de la compagnie. Si la gestion n'est pas suffisamment performante, des pertes ou une trop grande prise de risque peuvent entraîner l'insolvabilité de l'assureur qui se retrouve en incapacité à poursuivre la gestion du portefeuille.

L'étude qui suit s'intéresse donc à la résolution de ce problème de gestion dans un cadre de *reinforcement learning*.

Deep ALM

L'approche de ce mémoire est inspirée par les travaux de Thomas Krabichler et Joseh Teichmann dans leur article « *Deep Replication of a runoff portfolio* », Krabichler & Teichmann (2020), qui s'intéresse à l'application du *deep learning* dans le domaine financier. Dans cet article publié en septembre 2020, ils s'intéressent à l'optimisation du bilan d'une banque de détail en utilisant un agent se basant sur un réseau de neurones pour prendre des décisions - une approche qu'ils nomment *Deep ALM*. Ils montrent que cette approche est performante à la fois dans la maximisation du capital de l'actionnaire mais également d'un point de vue calculatoire.

On peut représenter le principe générale de ce modèle ALM par le graphe 3.7, en phase d'en-

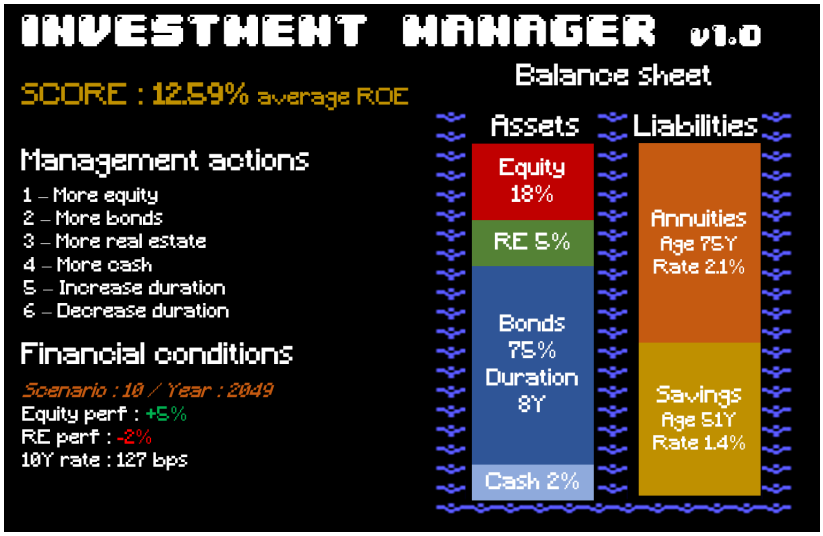


FIGURE 3.6 – De Tetris au jeu de la gestion ALM

traînement, puis le graphe 3.8 en phase d'inférence.

Le modèle ALM central permet de réaliser tous les calculs financiers relatifs au portefeuille étudié, pour chaque scenario et chaque pas de temps. Ce dernier est initialisé avec des valeurs issues de calculs externes. Pendant l'apprentissage, le réseau de neurones est alimenté régulièrement avec en entrée l'état du système et le *reward* résultant. Ce dernier approxime la fonction de valeur en actualisant au fur et à mesure des scenarios les poids de chaque neurone et de chaque couche.

En choisissant l'action qui maximise la fonction de valeur sachant l'état du système donné, une nouvelle allocation et un nouveau gap de duration sont choisis comme nouveaux paramètres du modèle ALM.

Le paramètre d'exploration du système génère occasionnellement une décision aléatoire. Comme déjà expliqué en section 3.1.2, ce paramètre est indispensable pendant la phase d'apprentissage du modèle pour permettre au réseau d'apprendre sur toutes les actions possibles.

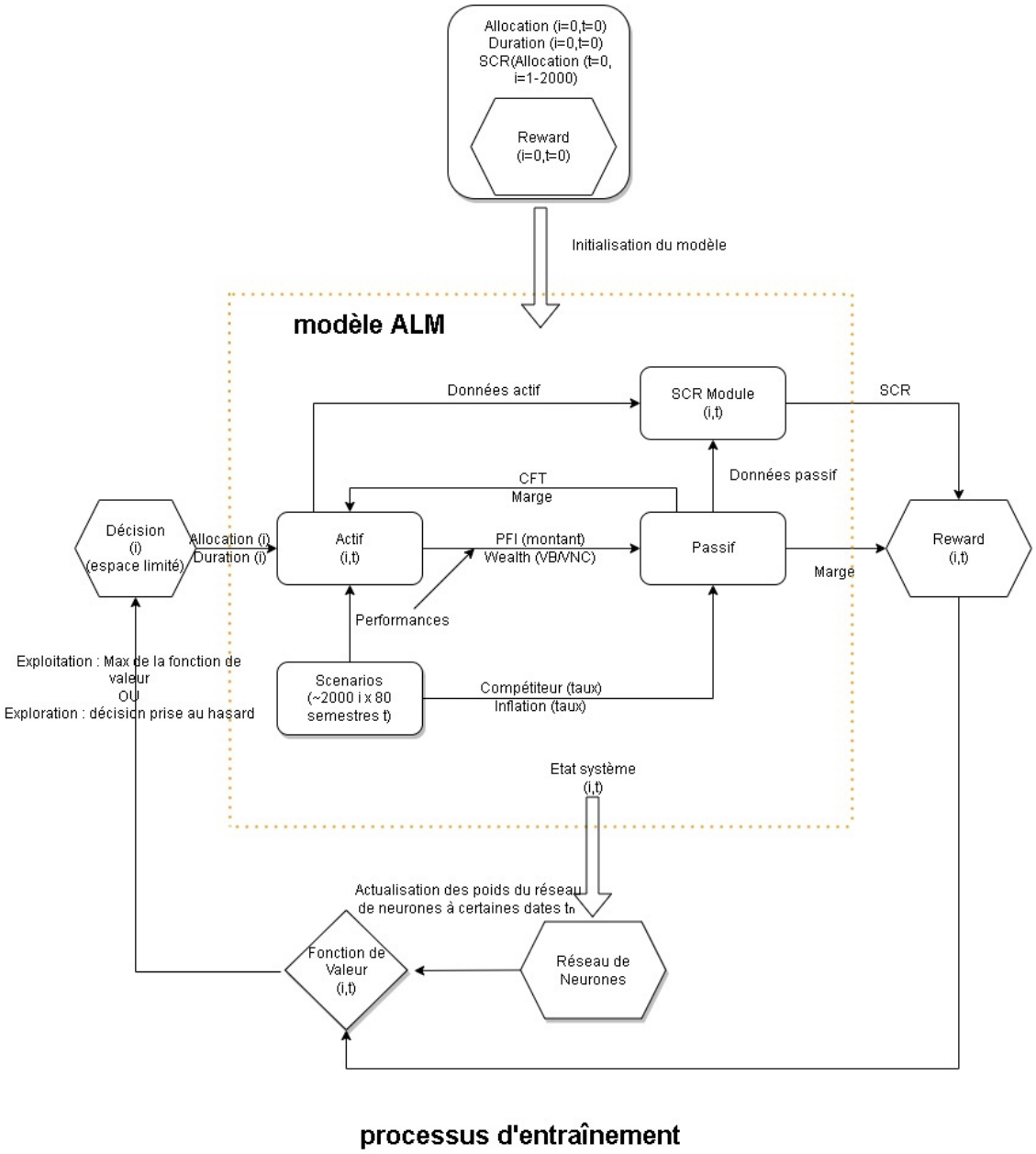


FIGURE 3.7 – Deep ALM - Schéma fonctionnel en phase d'entraînement

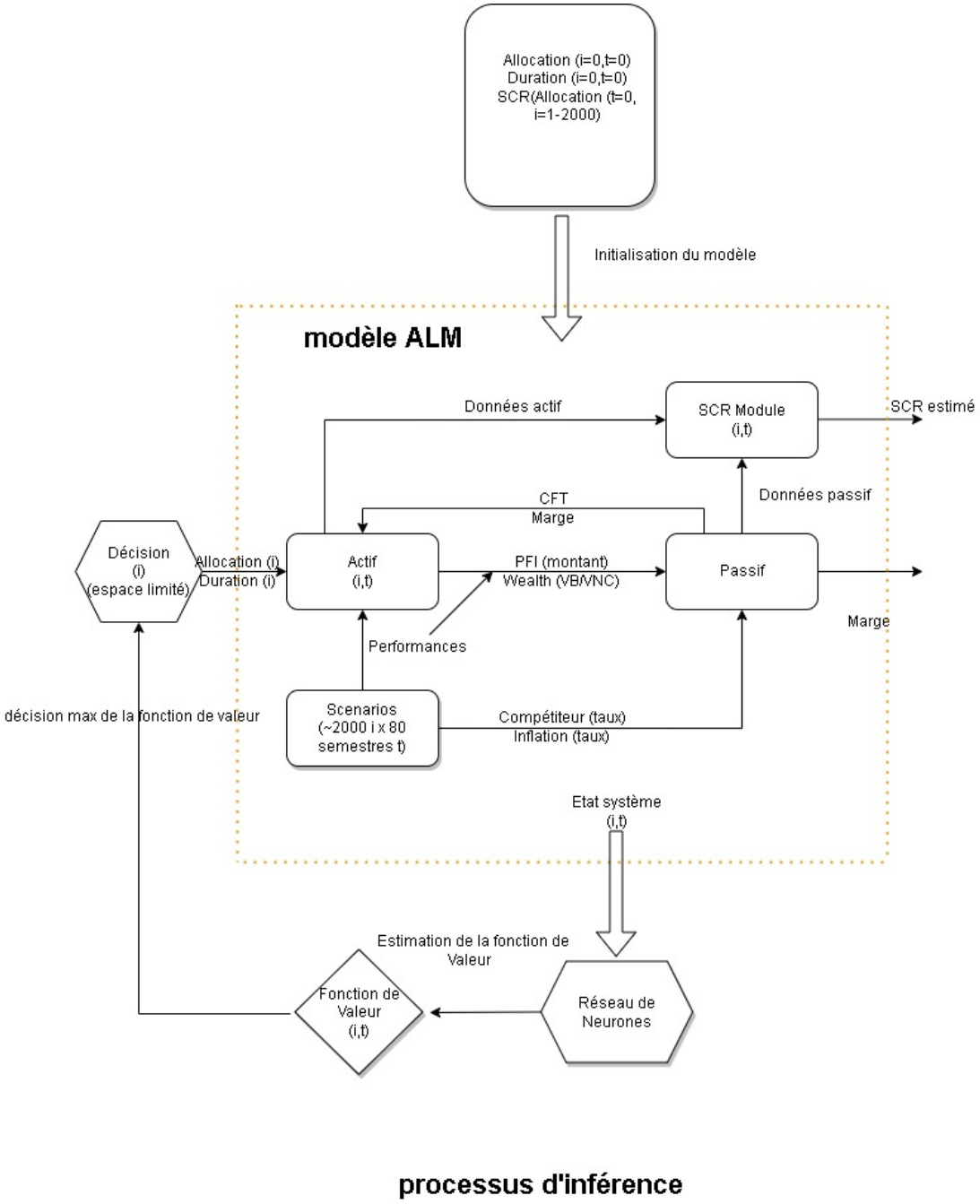


FIGURE 3.8 – Deep ALM - Schéma fonctionnel en phase d'inférence

Chapitre 4

Implémentation du modèle d'apprentissage

4.1 Caractéristiques du portefeuille étudié

Comme décrit dans les chapitres précédents, nous avons choisi de modéliser un portefeuille fictif de retraite collective simplifié pour mener à bien cette étude. Le portefeuille de retraite collective d'AXA a servi de référence pour calibrer les caractéristiques du portefeuille afin de refléter une réalité possible dans la modélisation et les résultats obtenus. L'idée générale est de prendre en compte la diversité des contrats dans le portefeuille pour refléter les risques techniques et financiers associés, tout en limitant le nombre de *model points* à projeter pour garder un temps de *run* du modèle raisonnable.

4.1.1 Descriptif du passif

La provision mathématique du portefeuille total est de 13,1 milliards d'euros, avec une proportion de contrats en phase de constitution plus importante. Les caractéristiques techniques principales pour chaque phase sont présentées dans la table 4.1.

	Provision Mathématique à t=0 (en M€)	taux garanti	âge moyen	marge sur encours moyenne	% partage des produits financiers
Constitution	8 816	0,87%	47,5	0,50%	97,50%
Restitution	4 312	1,50%	72	0,50%	97,50%
TOTAL	13 128	1,08%	55,5	0,50%	97,50%

Table 4.1 – Volumes et caractéristiques moyennes principales du passif

Tous les contrats sont assimilables à des « Articles 83 » : ils bénéficient d'une phase de constitution qui peut donner lieu à des transferts (individuels ou de l'entreprise complète), puis d'une phase de restitution avec une rente. Pour simplifier cette dernière est non réversible. Les contrats arrivant à date de liquidation optent soit pour un capital, soit pour la rente. En effet, la loi autorise

les détenteurs de petits capitaux constitutifs à recevoir la totalité du capital au moment de la liquidation.

Plusieurs niveaux de taux garantis (TMG ou taux technique) sont pris en compte afin de capturer les effets non linéaires associés à la diversité d'un portefeuille qui s'est construit sur plusieurs générations de taux. En effet, les niveaux de marge financière après participation aux bénéfices dépendent de la distribution des taux techniques, du fait de devoir servir ce minimum contractuel et sont sensibles aux scenarios de marché simulés. (Voir table 4.2)

TMG/ taux technique	Constit	Restit
0,00%	25%	14%
0,50%	25%	14%
1,00%	25%	14%
1,50%	0%	14%
2,00%	25%	14%
2,50%	0%	14%
3,00%	0%	14%

TABLE 4.2 – Distribution des taux garantis en pourcentage des volumes

En ce qui concerne la marge sur encours et le niveau de partage des produits financiers, on peut considérer que ces niveaux sont un peu plus homogènes dans le portefeuille, avec toutefois deux niveaux pour chaque paramètre. Cela correspond à la fois à une réalité du portefeuille réel, tout en limitant le nombre de *model points* modélisés et en simulant des niveaux de marge différents pour les contrats. Les tables 4.3 et 4.4 présentent les paramètres choisis.

Marge sur encours	Constit	Restit
0,30%	50%	50%
0,70%	50%	50%

TABLE 4.3 – Distribution des marges sur encours en pourcentage des volumes

% partage des produits financiers	Constit	Restit
95%	50%	50%
100%	50%	50%

TABLE 4.4 – Distribution des taux de partage des produits financiers en pourcentage des volumes

Dans cette modélisation, chaque *model point* représente une tête moyenne, dont l'âge est échelonné entre 29 et 77 ans. Cette distribution des âges dans le portefeuille permet de refléter un portefeuille mature qui contient toutes les générations d'âge, avec des passages en rente ou des décès de rentiers réguliers.

Compte tenu de toutes ces caractéristiques, il y a en tout 608 *model points* pour la phase de constitution et 308 pour la phase de restitution, ce qui fait déjà 916 *model points* de passif à simuler.

Autres caractéristiques du passif

Une loi de rachat partiel représentant le taux de transfert annuel de 2% fixes est appliquée. En complément, un niveau de sortie en capital est appliqué à hauteur de 35%. Comme présenté dans le chapitre 3, une loi de rachat dynamique est modélisée pour compléter les rachats partiels structurels dans des situations de marché adverses.

Un taux de chargement de 2% est prélevé sur les rentes.

Concernant les frais généraux futurs, le choix de modélisation ici est de considérer un niveau de coût proportionnel aux volumes mais différenciés entre la PM sous gestion et les prestations servies.

Une approche par coûts unitaires pourrait être développée pour prendre en compte des coûts fixes par contrat, et cela nécessiterait une modélisation tenant compte du nombre de contrats à chaque pas de temps, mais cela n'est pas toujours possible en assurance Collective de suivre le nombre d'assurés sous-jacent au contrat.

Par ailleurs, il serait possible d'ajuster l'évolution des coûts à l'inflation, mais l'approche retenue proportionnelle aux PM est déjà prudente car la PM est régulièrement revalorisée par la participation aux bénéfices, qui est le plus souvent plus élevée que l'inflation dans le contexte financier actuel.

On utilise donc les taux suivants pour la modélisation :

- Taux-exp-pm : 0,25%

- Taux-exp-clms : 1%

4.1.2 Descriptif de l'actif

Ce portefeuille simplifié bénéficie d'une allocation initiale centrée sur les obligations gouvernementales, les actions et l'immobilier :

Voir table 4.5

	Valeur Comptable à t=0 (en M€)	Valeur Boursière à t=0 (en M€)	Plus-value Latente à t=0	Taux de PVL	Allocation (hors cash)
GOVIES	10 412	16 077	5 665	54%	83%
ACTIONS	1 167	1 637	470	40%	9%
IMMOBILIER	940	1 281	341	36%	8%
CASH	609	609	-	0%	
TOTAL	13 128	19 605	6 477	49%	100%

TABLE 4.5 – Volumes et caractéristiques moyennes principales de l'actif

Ce portefeuille détient un volume de plus-values latentes important sur toutes les classes d'actifs, y compris sur les actifs dont la richesse est pilotable (Actions, Immobilier). C'est le reflet d'une certaine ancienneté par cohérence avec le passif dont les caractéristiques représentent plusieurs

générations de taux techniques ou d'âges souscrits sur plusieurs années.

4.1.3 Solvabilité initiale

A l'aide du modèle interne d'AXA France, on peut déterminer les métriques de solvabilité pour ce portefeuille, avec la méthodologie décrite dans le chapitre 2.

Voir tables 4.6 et 4.7

(en M€)	Brut IS	Net IS
VIF	630	428
SCR	858	742
RM	268	182
EOF hors Fonds propres		246
Besoin minimal en Fonds Propres		496

Table 4.6 – Indicateurs de solvabilité du portefeuille et besoins en fonds propres

Décomposition du SCR Brut (en M€)	Brut IS
SCR technique	655
Longevite	*570*
Rachat	*71*
Expense	*136*
SCR marché	321
Taux	*200*
Action	*120*
Immobilier	*55*
BSCR - Adj	798
Risque Opérationnel	60
SCR BRUT	858

Table 4.7 – Décomposition du SCR

Malgré un niveau de marge future élevé sur le portefeuille (428 M€), le besoin en capital est important, avec un SCR total de 742M€. Ce dernier est principalement lié au SCR de longévité élevé (570M€), et un SCR marché qui vaut moitié moins que le SCR technique. Cette composition n'est pas surprenante du fait d'un portefeuille avec un fort taux de sortie en rente (65% du portefeuille de constitution bascule en rente) et des taux garantis pouvant aller jusqu'à 3%. La *Risk Margin* a également un poids significatif (24% du SCR), du fait du risque de longévité qui génère un besoin en capital sur le long terme pour l'actionnaire. Cette dernière a été calculée avec une approche par *risk driver* pour la projection de chaque SCR technique et opérationnel.

Le besoin minimal en fonds propres pour ce portefeuille est de 496M€, ce qui fait que la VIF ne peut à elle seule couvrir le besoin en capital. Il y a donc un besoin minimal complémentaire de capital par l'actionnaire de 250M€.

4.2 Description des scenarios économiques

Pour l'entraînement de notre modèle nous utilisons 2000 scénarios économiques générés par les équipes du risk-management du groupe AXA, et utilisés par les équipes ALM dans leurs modèles. Il s'agit de scénarios *real world* qui intègrent donc une prime de risque. Etant donné que nous cherchons à modéliser la marge réelle de l'assureur, la prime de risque est nécessaire pour refléter l'intérêt de l'investissement en actifs risqués. Cela n'a pas d'impact sur les calculs de SCR, qui doivent rester en environnement risque neutre, car nous utilisons des drivers à partir du SCR fourni par le modèle interne d'AXA France.

Dans le cadre de notre modèle simplifié, nous projetons un indice pour un indice de performance et un taux de dividendes pour les actions et l'immobilier ainsi qu'une courbe des taux OAT et un indice d'inflation. Ce dernier est utilisé pour inflater les frais généraux et participe à la satisfaction client pouvant déclencher des rachats dynamiques.

	Inflation	Equity TR	Equity div	Real Estate TR	Real Estate div
Mean	1,94%	5,15%	2,39%	4,71%	2,35%
Std Variation	3,18%	18,00%	1,76%	9,52%	0,33%
Minimum	-9,06%	-51,92%	0,00%	-31,94%	1,23%
Q1 - 25%	-0,10%	-7,57%	1,28%	-1,92%	2,12%
Q2 - 50%	1,70%	3,67%	2,01%	4,25%	2,33%
Q3 - 75%	3,77%	16,20%	2,91%	10,82%	2,56%
Maximum	20,20%	122,04%	35,58%	64,10%	4,35%

TABLE 4.8 – Statistiques descriptives des scénarios utilisés

	Inflation	Equity TR	Equity div	Real Estate TR	Real Estate div
Inflation	100%	13%	-19%	21%	-4%
Equity TR		100%	-8%	40%	-4%
Equity div			100%	-8%	21%
Real Estate TR				100%	-10%
Real Estate div					100%

TABLE 4.9 – Corrélations entre les scenarios

4.3 Description du modèle

4.3.1 Déroulé d'une année de simulation

A chaque pas de temps, le modèle exécute les étapes suivantes :

— **Etape 1 – Déroulé de l'actif** :
  — Calcul des valorisations, gestion des maturités et calcul des revenus générés,
  — Adossement - Alignement de la valeur comptable des actifs avec le montant de PM via un ajustement de cash,
  — Ajustement du levier : si la position de cash est supérieure à une limite fixée (15% dans nos simulations), on investit le surplus au prorata de l'allocation d'actifs actuelle. Si la part de cash est inférieure à un minimum fixé (-15% dans nos simulations), on vend des actifs

au prorata de l'allocation actuelle. On aurait pu envisager un modèle sans cette règle de gestion. Toutefois, étant donné que le portefeuille est structurellement en décollecte, on se retrouve naturellement dans des situations de cash négatif. Si ces situations arrivent trop brutalement, il faudrait plusieurs pas de simulation pour revenir à une allocation désirée par l'algorithme ce qui limiterait son influence.

— **Etape 2 – Décision d'allocation** : calcul des variables utilisées par l'agent et détermination de la décision prise par l'agent. Calcul d'une nouvelle allocation cible correspondant à l'allocation actuelle incrémentée de l'action choisie puis implémentation de cette nouvelle allocation via des achats/ventes.
— **Etape 3 – Déroulé du passif** : Calcul des produits financiers qui dépendent à la fois des conditions économiques, de l'historique du portefeuille et de la décision prise par l'algorithme via les plus-values potentiellement générées. Application de la participation aux bénéfices et mise à jour du passif (rachats, décès, passages en rente, ...).
— **Etape 4 – Mise à jour de l'agent** :
  — Calcul de la marge de l'assureur et mise à jour du SCR afin de déterminer le reward de l'agent,
  — Entrainement du modèle le cas échéant.

4.3.2 Gestion de l'actif

Les actions et l'immobilier sont chacun gérés via une seule ligne pour laquelle nous suivons la valeur comptable ainsi que la valeur de marché. Les produits financiers générés par ces classes d'actifs sont constitués des dividendes payés et des plus ou moins values réalisées nettes des frais de transaction. On suppose par simplification qu'il n'y a pas de frais de gestion des actifs ce qui revient de façon équivalent à supposer que la surperformance des gérants d'actifs est égale aux frais qu'ils prélévent.

En ce qui concerne les obligations, nous les modélisons en ligne à ligne à partir de l'actif initial. Pour les investissements, nous investissons en zero-coupons au taux de marché actuel. La maturité choisie est déterminée à partir de la cible de duration fixée par l'agent avec une limite à 50 ans. Les produits financiers générés par les obligations sont constitués des coupons pays par les obligations du stock intial, l'amortissement de la surcote/décote du stock initial et des zero-coupons achetés et des plus ou moins values réalisées nettes de frais de transaction. On note ici que contrairement à ce qui se passerait dans le cadre assurentiel français, nous ne modélisons pas de réserve de capitalisation dans laquelle doivent aller les plus ou moins values obligataires. Ce choix a été fait pour ne pas complexifier le modèle.

Lors de transactions à l'actif, nous appliquons des frais de transactions sur la valeur de marché en fonction de la classe d'actifs : 0.2% pour les actions, 1.0% pour l'immobilier et 0.8% pour les obligations.

4.3.3 Choix de la fonction de *reward*

Le *reward* est lié au bénéfice généré par le portefeuille pour l'actionnaire de la compagnie. Ainsi la marge nette nous semble être une métrique adaptée. Toutefois, nous avons identifié deux contraintes supplémentaires pour le choix du *reward* dans notre simulation :

— Avoir un taux : pour maintenir une stratégie homogène lors de l'écoulement du portefeuille indépendamment du montant de PM. Dans le cas d'un reward fixé en montant, l'agent serait

poussé à externaliser de la richesse en début de projection pour réaliser le plus de marge possible en valeur absolue. Cela serait toutefois contraire aux intérêts de la compagnie étant donné qu'elle pourrait être confrontée à des sanctions réglementaires si sa gestion n'est pas soutenable à long terme. On pourrait contourner cette contrainte en modélisant des pénalités en cas de chute du ratio de solvabilité trop fort mais nous ne l'avons pas implémenté pour simplifier l'analyse.

— Ajustement en fonction des conditions de marché : un taux de marge n'a pas forcément la même valeur dans différentes conditions de marché. Ainsi la métrique choisie doit tenir compte des différents scénarios.

Nous aboutissons donc à un choix similaire à la métrique optimisée par l'ALM d'AXA France, à savoir une métrique de *Return on Equity* (RoE) au-delà du taux sans risque :

$$\rho_{t,i} = M_{t,i}/K_{t,i} - \kappa - \tau_{t,i}$$

où
- $\rho_{t,i}$ est le *reward* à la date t du scénario i
- $M_{t,i}$ est la marge nette de l'assureur à la date t du scénario i
- $K_{t,i}$ est le capital immobilisé par l'assureur à la date t du scénario i. Il s'agit du produit entre le ratio SII cible et le SCR.
- κ est le coût du capital exigé par l'actionnaire.
- $\tau_{t,i}$ est le taux sans risque à la date t du scénario i. On utilise le taux OAT 10 ans en tant que taux sans risque.

4.3.4 Description de l'état du système

Contrairement au jeu de Tetris, il n'est pas possible de fournir à l'agent une vision complète du système car la dimensionalité du problème serait trop importante. En effet, il faut rappeler qu'il y a 608 model points qui sont définis par une dizaine de paramètres qui évoluent dans le temps. Il y aurait un ordre de grandeur de quelques dizaines milliers de variables à fournir à l'agent pour avoir l'information complète. Avec un réseau de neurones comptant 8 neurones dans la première couche, cela ferait une centaine de milliers de paramètres à calibrer. Cela n'est clairement pas envisageable étant donné la puissance de calcul disponible lors de l'étude et surtout de la quantité de données (2000 scénarios à simuler).

Ainsi un ensemble de variables ont été sélectionnées en se basant sur les éléments observés par les équipes ALM, d'allocation et du risk management d'AXA France. De plus, la sélection ne porte que sur des variables normalisées par la taille du portefeuille qui décroit dans le temps. L'objectif est d'éviter à l'agent de faire des ajustements d'échelle et également l'inciter à avoir une stratégie de gestion cohérente tout au long de la vie du portefeuille.

Les variables sélectionnées sont donc les suivantes :

— **Allocation d'actifs du portefeuille** : pourcentage de la valeur boursière de l'actif pour chacune des classes d'actifs (actions, immobilier, obligations, cash).

— **Duration** : duration de l'actif et duration du passif en années – il est intéressant d'avoir en plus du gap de duration le niveau absolu de duration afin d'appréhender le risque de taux.

— **SCR total** : montant de SCR normalisé par la PM.

— **SCR marché** : montant de SCR normalisé par la PM pour chacune des composantes du SCR marché (SCR taux, SCR actions, SCR Immobilier) – donne des éléments sur la contribution des différents actifs au capital immobilisé ainsi qu'une notion de l'exposition aux différents

facteurs de risque. Nous avons choisi de ne pas détailler les autres composantes du SCR car elles sont essentiellement subies par l'agent via l'écoulement du passif.

— **Taux de marge** : montant de marge actionnaire normalisée par la PM du portefeuille.
— **Taux de richesse de l'actif** : ratio entre la valeur boursière et la valeur comptable de l'actif – donne une idée des réserves futures disponibles et du surplus de marge qui peut être généré lors d'une vente.
— **Taux de produits financiers** : taux de produits financiers générés lors de l'exercice précédent.

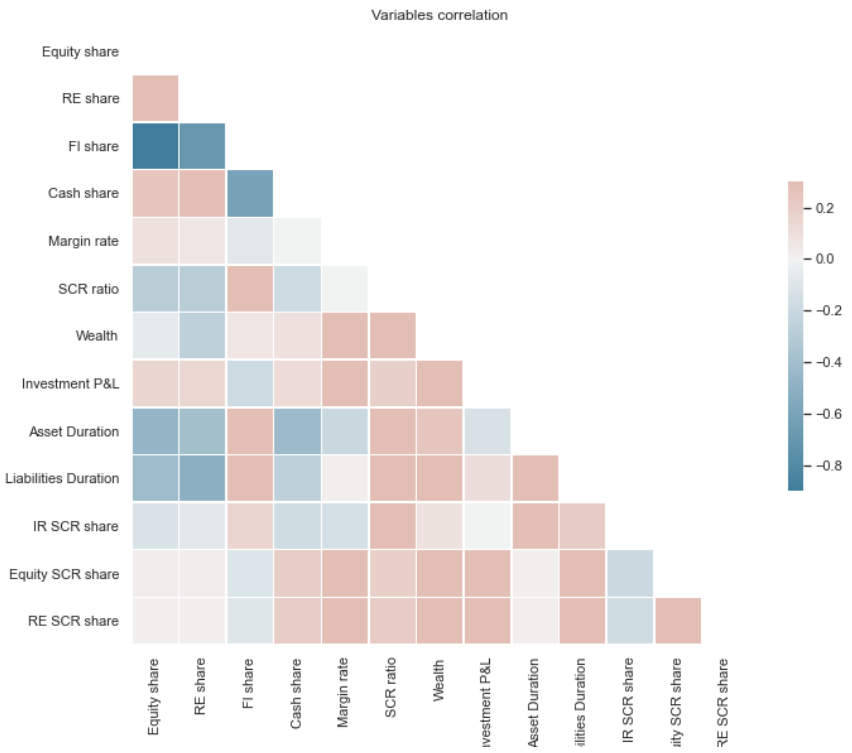


FIGURE 4.1 – Matrice de corrélation des variables d'état de l'environnement

4.3.5 Description de l'agent

Espace des actions

Dans le cadre de cette étude, c'est le cadre traditionnel du *reinforcement learning* où l'agent prend une décision dans un espace d'actions multi-discret qui a été mis en application. Bien qu'il existe des extensions à des espaces continus, il s'agit d'éviter une trop forte complexité tout en répliquant des actions qui restent possibles dans le contexte d'une compagnie d'assurance réelle. Kanervisto, Scheller, et Hautamäki ont montré dans leur étude Kanervisto et al. (2020) qu'un espace d'actions dit multi-discret (soit une discrétisation selon plusieurs axes différents) peut certes conduire à des performances plus faibles qu'en espace continu, mais il assure pourtant une capacité d'apprentissage de l'agent plus efficace en étant notamment moins gourmand en exploration et plus rapide pour converger. Dans la même logique, ils préconisent une réduction du nombre d'actions possibles, même si un nombre plus élevé d'actions pourrait donner de meilleures performances sur

les choix à réaliser en situations extrêmes.

Les pas de réallocation retenus sont d'une amplitude de 2% ou 8%. Cela permet de faire des mouvements faibles ou amples tout en restant sur des montants qui peuvent être envisagés dans la réalité. Pour chaque classe d'actifs, il est possible de faire un pas positif ou négatif. Pour les obligations, les pas peuvent être faits en augmentant la duration de l'actif ou en la diminuant. A cela s'ajoutent l'action nulle (aucun mouvement), l'achat et la vente d'actifs au prorata de l'allocation actuelle. Cela fait un total de 25 actions possibles pour l'agent.

Choix de l'architecture du réseau de neurones

Il n'existe pas de méthodologie permettant de choisir une architecture de réseau de neurone optimale a priori pour ce problème. Il faut donc se baser sur différentes heuristiques issues de l'expérience des praticiens du *deep learning* et appliquer en fonction des caractéristiques des données du modèle implémenté.

Contrairement au jeu de Tetris qui peut être joué une infinité de fois, nous ne disposons pour cette étude que d'un nombre limité de scénarios financiers et nous n'avons pas la capacité à en générer des supplémentaires issus du même modèle. Ainsi la quantité de données d'apprentissage étant limitée il a fallu restreindre le nombre de paramètres du réseau de neurones. Pour cela nous limitons déjà le nombre couches cachées à deux en nous appuyant sur le théorème d'approximation vu au chapitre précédent. Cela a en plus l'avantage de limiter le risque d'*overfitting*. La limitation en termes de quantité de données nous pousse aussi à nous orienter vers de fonctions d'activation de type *Relu* entre les couches cachées qui permettent de garder suffisamment de gradient pour avoir un apprentissage rapide. En ce qui concerne la couche de sortie, c'est plutôt une fonction d'activation de type *Softmax* qui est retenues ; c'est une généralisation de la fonction sigmoïde utilisée pour les problèmes de classification non-binaires comme le nôtre.

Pour le choix de la taille des couches cachées il est généralement conseillé dans la littérature d'utiliser des tailles qui sont des puissances de deux, plus efficientes lors des calculs, et avoir un nombre de neurones d'un ordre de grandeur similaire à la taille des données en entrée. Le vecteur d'entrée du modèle est de dimension 13, les tests réalisés ont donc porté sur des couches de taille 8, 16, 32 et 64. Afin de limiter le risque de surapprentissage et forcer l'entrainement de l'intégralité des neurones du réseau, nous avons mis en place un mécanisme de *dropout* qui désactive à chaque entrainement 20% des neurones de façon aléatoire au sein du réseau. L'inconvénient de ce mécanisme est de ralentir la vitesse d'apprentissage mais cela se fait en échange d'une plus grande généralité du modèle.

Stratégie d'entrainement

L'agent est constitué de deux réseaux de neurones, d'une part un modèle cible qui est le modèle utilisé pour prendre des décisions à chaque pas de temps et d'autre part un modèle temporaire qui est entrainé à chaque pas de temps. Le modèle temporaire est entrainé à partir des différents enregistrements état, action, *reward* historiques. A chaque fin de scénario, le modèle cible prend les poids du modèle temporaire. Cela a comme intérêt de garder une logique de décision stable tout le long d'un scénario. A chaque pas, lors de l'entrainement du modèle temporaire, un échantillon de 32 enregistrements est tiré parmi les données des états, décisions et *rewards*. Le modèle est ensuite entrainé via le principe de rétropropagation du gradient. Comme mentionné dans la partie précédente, une part de décisions aléatoires est introduite en phase d'entrainement, paramétrée par

un taux d'exploration afin de forcer l'algorithme à explorer l'espace des actions.

4.3.6 Métriques de performance

Le *reward* est évidemment la métrique de performance à analyser en priorité étant donné qu'il s'agit à la fois de la métrique *business* utilisée et de la métrique maximisée par le réseau de neurones entraîné. Il est possible de comparer la distribution de celui-ci en fonction des différents agents prenant des décisions. Etant donné qu'il s'agit d'une variable aléatoire, la séléction de l'agent passe par celui qui offre le meilleur compromis entre un *reward* important et stabilité de celui-ci. Ainsi il faut se concentrer sur le positionnement des agents dans le plan défini par la moyenne du *reward* et de son écart-type. Les scénarios défavorables extrêmes sont également un point d'attention car la capacité à éviter des faillites doit également être valorisée.

Le portefeuille, contenant notamment des taux garantis, n'est pas rentable dans de nombreux scénarios, il n'est donc pas possible de conclure sur la base d'un *reward* négatif qu'un agent a une mauvaise performance. Cela pousse donc à choisir de définir des stratégie intuitives qui serviront de *benchmarks* pour l'analyse de performance :

- **Aucune action** : aucun achat et aucune vente ne sont faits mis à part les rebalancements automatiques liés aux contraintes fixées sur le cash,
- **Action aléatoire** : une des décisions est choisie aléatoirement par l'agent à chaque étape,
- **Allocation d'actifs fixe** : on maintient l'allocation d'actifs identique à l'allocation initiale en faisant les achats et les ventes nécessaires sans limite de volume,
- **Allocation d'actifs cible** : à chaque étape l'agent effectue l'action qui le rapproche le plus vers une allocation cible qui est définie ici comme étant l'allocation initiale.

4.3.7 Exemple de déroulé actif/passif

Les graphiques 4.2 à 4.5 présentent l'évolution de l'état du système sur un scénario simulé en utilisant la stratégie d'allocation d'actifs cible. On peut y voir les différentes actions prises, au sein de l'espace défini ci-dessous, qui permettent à l'algoritheme de s'approcher à chaque pas de projection de l'allocation cible.

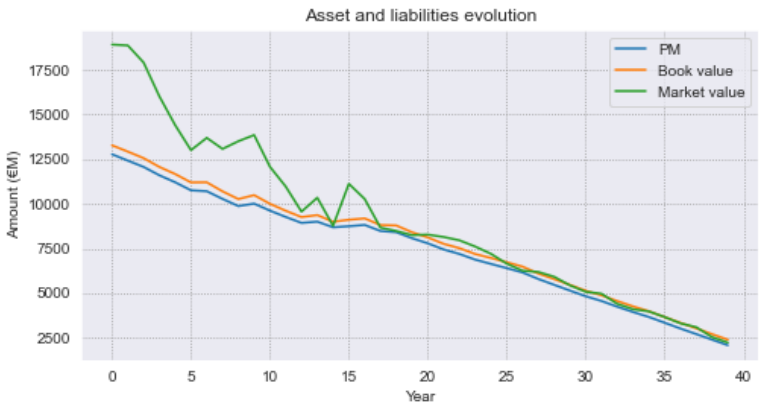


FIGURE 4.2 – Déroulé d'une simulation - évolution de la PM

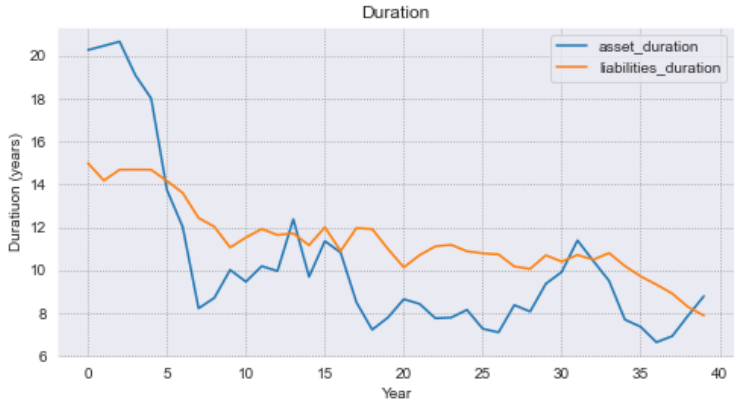


FIGURE 4.3 – Déroulé d'une simulation - évolution de la duration

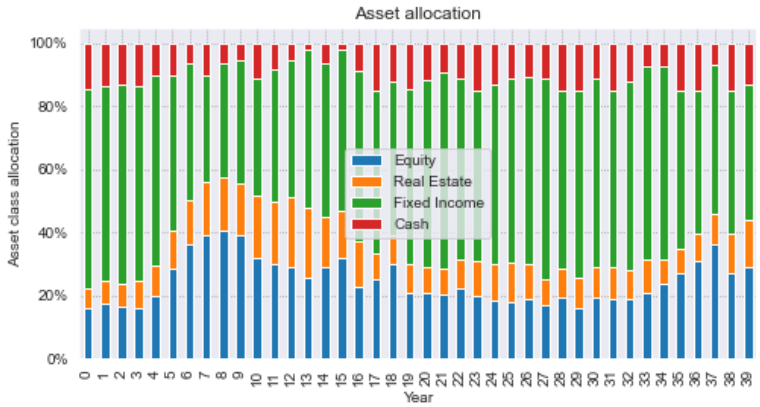


FIGURE 4.4 – Déroulé d'une simulation - évolution de l'allocation d'actifs

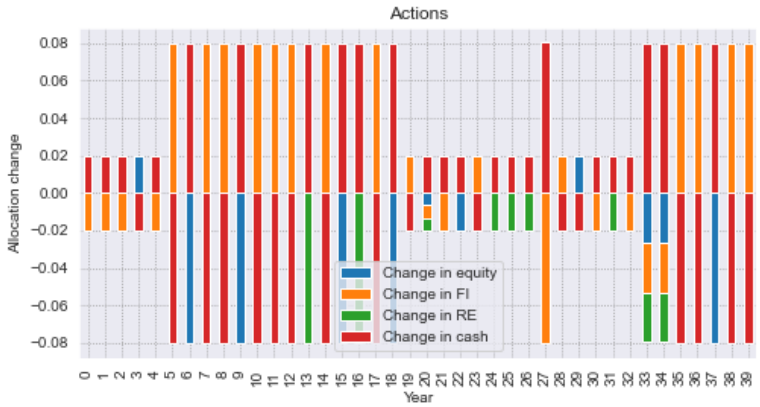


FIGURE 4.5 – Déroulé d'une simulation - actions prises par la stratégie de référence

Chapitre 5

Présentation et analyse des résultats

5.1 Séléction d'un *benchmark*

Dans un premier temps, la cible est de déterminer la meilleure stratégie de référence contre laquelle il sera ensuite possible de comparer la performance des agents. La table 5.1 présente pour chaque année de simulation * le *reward* moyen et son écart-type par stratégie. D'après ces résultats, la non-action (en bleu) apparaît comme la stratégie la moins performance alors que les stratégies d'allocation fixe et cible surperforment toutes les deux les décisions aléatoires. Les tables 5.1 et 5.2 donnent les statistiques des *reward*. On y remarque qu'en moyenne l'agent n'arrive pas à obtenir le rendement sur capital attendu de 14% au delà du taux sans risque sur le portefeuille. La stratégie d'allocation fixe est la plus performante dans l'absolu et en relatif à la stratégie aléatoire. Toutefois, les mouvements effectués par cette stratégie peuvent être très importants (jusqu'à 85% de l'actif réalloué). De tels mouvements ne seraient pas possibles dans la réalité à cause notamment de contraintes de liquidité. De plus cette stratégie se caractérise par un espace d'actions continu qui n'est pas comparable à l'espace des actions disponibles pour l'agent. Nous décidons donc de garder la stratégie d'allocation cible comme référence pour analyser les résultats de l'agent.

Stratégie	Moyenne	Ecart-type	Minimum	Q1 - 25%	Médiane	Q3 - 75%	Maximum
Fixed	-6,10%	18,49%	-590,35%	-6,23%	-4,12%	-1,87%	49,13%
Target	-7,01%	21,19%	-848,59%	-7,55%	-5,32%	-3,04%	56,35%
Noaction	-7,64%	6,71%	-174,83%	-9,45%	-6,23%	-4,69%	8,33%
Random	-9,09%	19,75%	-791,82%	-10,06%	-6,52%	-4,41%	69,59%

Table 5.1 – *Rewards* des stratégies de référence

Stratégie	Moyenne	Ecart-type	Ratio de Sharpe
Fixed	2,99%	24,59%	12,16%
Target	2,09%	26,85%	7,78%

Table 5.2 – Stratégies de référence comparées à des décisions aléatoires

∗. Les points sombres correspondent aux premières années de projection

FIGURE 5.1 – Evolution des moyennes et écarts-types des *rewards* des stratégies de référence

5.2 Sélection de l'achitecture du réseau de neurones

Dans un premier temps, il faut sélectionner une architecture de réseau de neurones performante. Ainsi le modèle a été lancé sur l'intégralité des 2000 scénarios avec les couches cachées suivantes : 8x16, 16x32, 32x32, 32x64, 64, 64x64, 64x128, 128, 128x64. Les agents ont été entrainés avec un taux d'exploration de 20% et en choisissant à chaque pas de temps l'action ayant le score le plus important. Le tableau 5.3 donne des statistiques sur les surperformances par rapport à la stratégie aléatoire qui permettent de faire le choix.

A ce stade, nous éliminons les réseaux qui ont une performance inférieure à la stratégie d'allocation cible pour se concentrer sur les trois architectures qui surperforment afin de n'en garder qu'une pour la suite de l'analyse. Le réseau 32x64 a une performance moyenne et un écart type en ligne avec la stratégie d'allocation cible. Toutefois sa médiane est sensiblement inférieure ce qui pousse à penser que la surperformance est concentrée sur quelques cas favorables. En effet, cette architecture ne surperforme le *benchmark* que dans 47% des observations. Pour les deux autres options sélectionnées ce taux est supérieur à 60%. Ainsi, nous éliminons ce réseau et poursuivons l'étude avec les réseaux 32x32 et 64x128.

Sur le graphique 5.2, nous avons représenté l'évolution de la performance des deux réseaux retenus tout au long de l'apprentissage sur les 2000 scénarios. On constate que sur les derniers 500 scénarios les deux scénarios ils ont une performance similaire alors que sur les 1000 derniers le réseau 64x128 a à la fois une meilleure performance moyenne et un écart-type plus faible de celle-ci. De plus, on observe une peformance en progression et un écart-type qui se réduit avec l'entrainement

Stratégie	Moyenne	Ecart-type	Ratio de Sharpe	Q1 - 25%	Médiane	Q3 - 75%
Fixed	2,99%	24,59%	12,16%	-0,05%	2,27%	6,58%
32x32	2,90%	30,09%	9,65%	-0,20%	2,82%	7,33%
64x128	2,73%	33,13%	8,23%	-0,04%	2,66%	6,84%
32x64	2,10%	26,61%	7,90%	-0,96%	0,94%	4,38%
Target	2,09%	26,85%	7,78%	-0,40%	1,51%	4,32%
128	2,53%	33,71%	7,52%	-0,17%	3,20%	7,55%
16x32	1,95%	31,13%	6,28%	-1,26%	0,90%	5,19%
64x64	1,51%	30,08%	5,01%	-1,31%	0,97%	4,91%
8x16	1,46%	29,42%	4,96%	-1,47%	0,57%	3,85%
64	1,32%	37,77%	3,49%	-0,46%	2,87%	7,53%
128x64	1,15%	34,03%	3,37%	-1,25%	1,72%	6,59%

TABLE 5.3 – *Rewards* relatifs à la stratégie aléatoire pour les réseaux de neurones testés

pour ce réseau alors que les statistiques restent stables pour le 32x32.

Etant donnés ces résultats, nous avons choisi de poursuivre notre étude en ajustant les paramètres pour un agent muni d'un réseau de neurones de taille 64x128.

Optimisation de la stratégie de décision

Enfin, nous avons voulu tester une stratégie de décision alternative. Au lieu de sélectionner l'action avec le score le plus élevé, l'action est sélectionnée à partir d'un tirage aléatoire dont les probabilités sont pondérées par les scores de chacune des actions. Cette stratégie a été développée dans le cadre d'un entraînement sur 800 scénarios. La performance en moyenne est similaire pour l'algorithme de décision stochastique. Toutefois en comparant les distributions de *reward* on constate que tous les quantiles de 0.2 à 0.8 sont meilleurs pour l'algorithme de décision *argmax*. On choisit donc de sélectionner ce dernier.

5.3 Analyse des résultats

5.3.1 Mesure de performance de l'agent entrainé

Maintenant qu'un agent a été sélectionné et entraîné, il faut mesurer sa surperformance par rapport à la stratégie de référence. La performance présentée ci-dessus a permis de faire une analyse comparative mais ne donne pas de performance absolue étant donné qu'elle intègre les décisions de l'agent lors de la phase d'entraînement et une part de décisions aléatoires liées au taux d'exploration.

Pour obtenir une performance comparable à la stratégie de référence, les scenarios sont divisés en deux : d'une part 90% des scénarios soit 1800 sont utilisés pour entraîner l'agent et les 10% restants (200 scénarios) sont utilisés pour mesurer sa performance sans entraînement et sans exploration. Cela nécessite, à l'issue des scénarios d'entraînement de figer les poids du réseau de neurones avant de continuer sur la phase test.

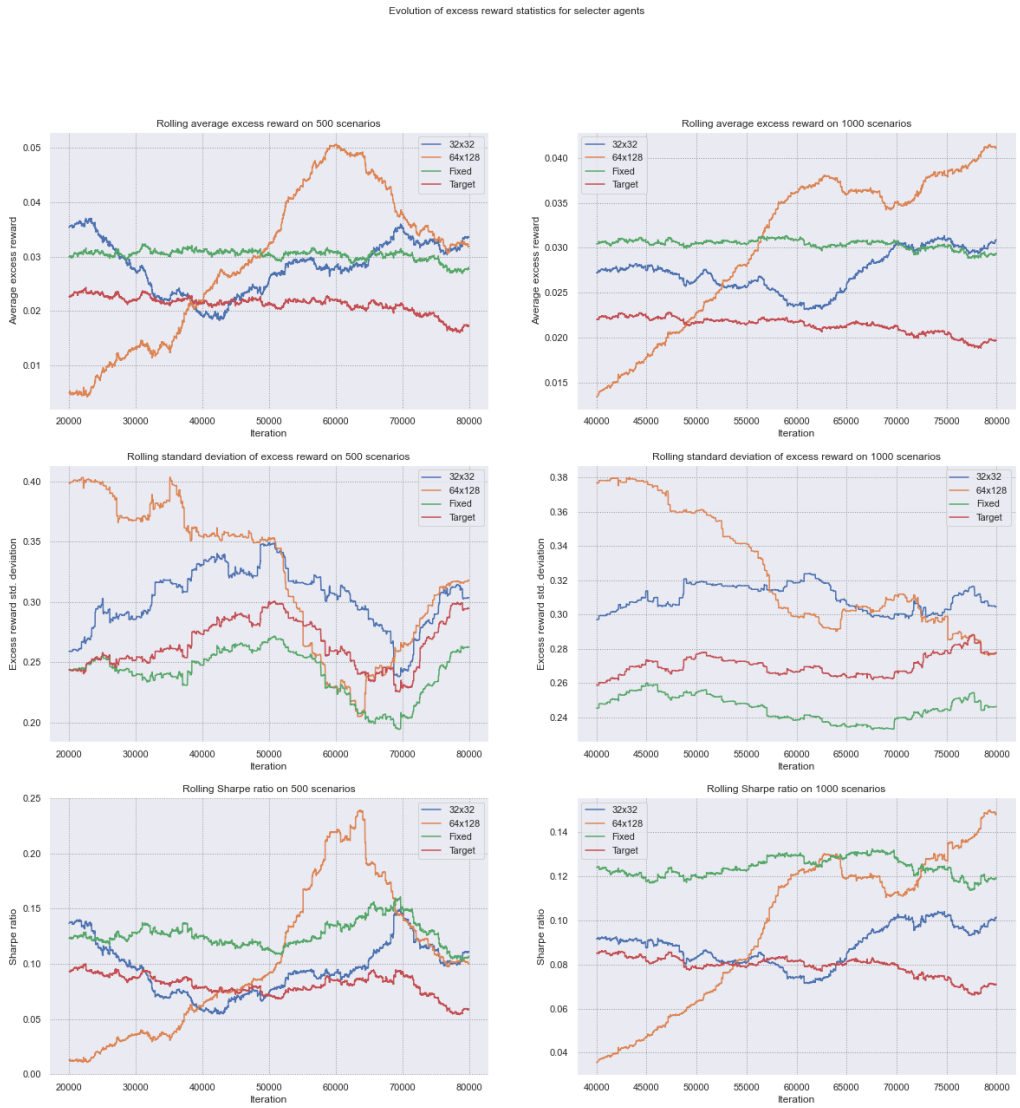


FIGURE 5.2 – Evolution des statistiques des *rewards* relatifs

Strategie	Moyenne	Ecart-type	Ratio de Sharpe	Q1 - 25%	Mediane	Q3 - 75%
Target	1,59%	33,15%	4,79%	-0,53%	1,32%	4,04%
Fixed	2,76%	29,54%	9,33%	-0,22%	2,25%	6,97%
64x128 - train + explo	3,89%	30,71%	12,68%	0,23%	2,87%	7,29%
64x128 - test	**4,12%**	**30,08%**	**13,69%**	**-0,48%**	**2,87%**	**8,16%**

TABLE 5.4 – Evolution des statistiques des *rewards* relatifs

La figure 5.4 présente les performances de l'agent en test, lors de la phase d'entraînement (200 derniers scénarios de la simulation d'entrainement) et la stratégie de référence. Les résultats de test sont meilleurs sur tous les indicateurs par rapport à la stratégie de référence. Ils sont égalment meilleurs que lors de l'entrainement, ce qui est logique étant donné que les décisions y sont diluées par 20% de décisions aléatoires.

5.3.2 Description des décisions l'agent sélectionné

Le graphique 5.3 donne l'histogramme des actions prises par l'agent sur les 200 scénarios de test. La grande majorité des actions effectuées par l'agent sont des ventes ce qui est cohérent avec les caractéristiques du passif qui est fortement en décollecte. Ces ventes sont principalement sur les actions et l'immobilier qui sont des actifs avec une forte performance sur les scénarios étudiés avec respectivement 5.30% et 4.7% de performance annuelle. Sur les actions, 25% des itérations voient des performances supérieures 16.2%. On comprend donc que l'agent en profite pour réaliser des plus-values permettant à la fois d'augmenter les produits financiers (et par conséquence la marge et la satisfaction client) et de diminuer le SCR actions coûteux en capital.

5.3.3 Calibration de paramètres supplémentaires

Maintenant que l'architecture du réseau de neurones est déterminée, il est intéressant d'optimiser certains paramètres afin d'améliorer les résultats. Ce sont ici deux paramètres qui sont étudiés, dans un premier temps le taux d'actualisation de l'agent et dans un second temps le pas utilisé pour définir l'espace des actions.

Taux d'actualisation de l'agent

Le taux d'actualisation utilisé par l'agent est la valorisation par celui-ci de *rewards* futurs. Les actions prises par l'agent ont un impact futur notamment via l'effet mémoire de la satisfaction client. Par défaut le taux d'actualisation à été fixé à 1% pour rester proche d'un taux d'actualisation économique. Toutefois les deux ne sont pas forcément liés et il est intéressant de jouer sur ce paramètre afin d'améliorer les décisions de l'agent. On peut faire le parallèle entre le taux d'actualisation (γ) et un horizon de temps (ΔT) sur lequel l'action a un impact.

$$\gamma = e^{-1/\Delta T}$$

Le taux d'actualisation à 1% implique un horizon de temps de 99 ans ce qui, étant donné le portefeuille, semble trop long. Sur un nombre de scénarios réduit (500) les performances de l'agent avec des taux d'actualisation correspondant à des horizons de temsp de 5 ans, 10 ans, 20 ans et 40 ans soit respectivement 18%, 10%, 5% et 2% sont comparées entre elles.

Bien que le nombre de scénarios soit limité, l'agent avec un taux d'actualisation de 10% soit 10 ans d'horizon d'impact des actions a les meilleures performances sur toutes les métriques regardées.

Pas des actions disponibles

On peut observer sur le graphe 5.3 que les actions prises par l'agent essentiellement avec un pas de 8% et que les actions avec un pas réduit ne sont que très peu utilisées. Pour rappel, le pas

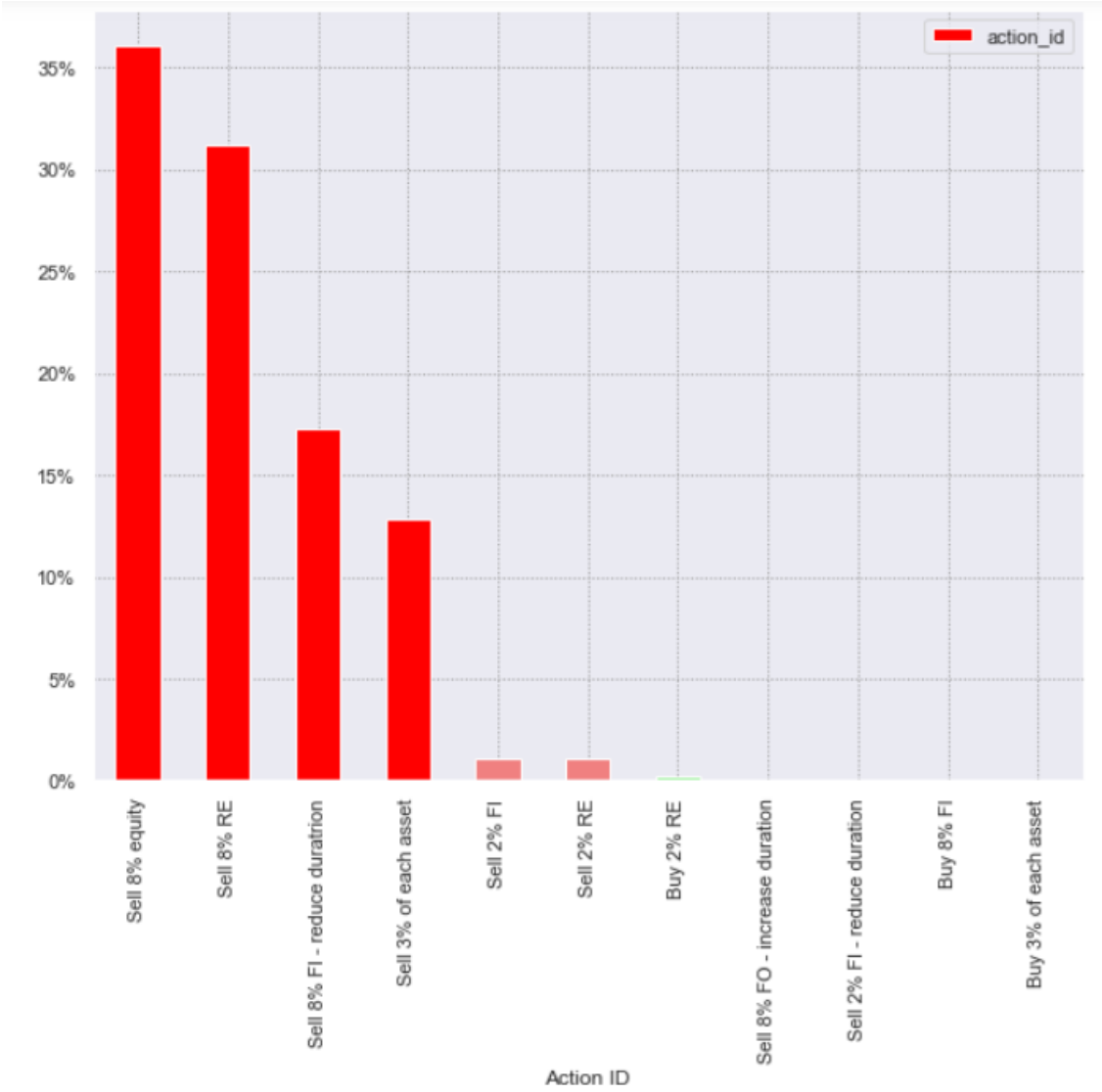


FIGURE 5.3 – Histogramme des actions prises par l'agent

Strategie	Moyenne	Ecart-type	Ratio de Sharpe	Q1 - 25%	Médiane	Q3 - 75%
Target	2,26%	24,39%	9,26%	-0,42%	1,56%	4,38%
10%	2,42%	27,63%	8,77%	-1,32%	0,86%	5,92%
18%	2,30%	28,64%	8,05%	-1,36%	0,85%	5,86%
5%	1,82%	31,56%	5,75%	-1,58%	0,66%	5,27%
0%	1,64%	31,51%	5,20%	-1,57%	0,97%	5,99%
2%	1,75%	34,56%	5,05%	-1,43%	0,72%	5,59%
-5%	0,86%	33,80%	2,55%	-1,85%	0,53%	5,84%
1%	0,49%	39,87%	1,22%	-1,18%	1,78%	6,88%

TABLE 5.5 – Performances de l'agent en fonction du taux d'actualisation sur 500 simulations

d'action a été placé à 8% pour une action ample et 2% pour une action réduite. Ce pas de décision a potentiellement un impact important sur la performance de l'agent qui n'a pas toujours la capacité à s'adapter à la situation avec des variations importantes de l'allocation d'actif à cause d'échéances ou de fortes réductions du passif pouvant être liées à des rachats dynamiques notamment. Ainsi, il est intéressant d'optimiser le pas des actions qui sont disponibles pour l'agent. La stratégie ayant

Stratégie	Moyenne	Ecart-type	Ratio de Sharpe	Q1 - 25%	Médiane	Q3 - 75%
Target	2,26%	24,39%	9,26%	-0,42%	1,56%	4,38%
Pas de 4%	1,60%	21,53%	7,44%	-1,05%	0,52%	2,36%
Pas de 10%	0,74%	38,78%	1,90%	-2,54%	0,82%	11,01%
Pas de 8%	0,49%	39,87%	1,22%	-1,18%	1,78%	6,88%
Pas de 15%	0,34%	36,39%	0,94%	-1,92%	0,44%	7,18%

TABLE 5.6 – Performances de l'agent en fonction du pas de réallocation sur 500 simulations

des pas de 4% s'avère plus performante sur les 500 scénarios étudiés aussi bien en espérance qu'en variance.

Performances de l'agent optimisé

Suite aux deux optimisations locales ci-dessus, nous avons choisi de lancer sur l'ensemble des 2000 scénarios l'agent avec un pas de réallocation de 4% et un taux d'actualisation de 10% ains que des simulations uniquement avec un pas de 4% et un taux d'actualisation à 10%.

Le graphique 5.4 présente les résultats des analyses avec des paramètres modifiés. Sur cette simulation plus longue on constate que le pas de 4% est finalement moins performant que le pas standard de 8% bien qu'en début d'apprentissage ce paramétrage était meilleur. Cela montre les limites de l'optimisation de paramètres sur un nombre de scénarios limités étant donnée la longueur de la convergence et des aléas des scénarios financiers.

On constate toutefois que l'actualisation à 10% a une performance similaire à l'agent initial qui semble plus stable notamment entre les itérations 50000 et 75000 où la performance de l'agent initial monte fortement avant de chuter alors que celle de l'agent modifié continue à augmenter avant de se stabiliser.

Il est assez intéressant de noter, comme le montre le graphe 5.5 que l'augmentation du taux d'actualisation amène l'algorithme à prendre des décisions plus variées. Ce graphe représente l'écart de fréquence des différentes actions, colorées en vert s'il s'agit d'achats ou en rouge pour des ventes. On y constate que les principales actions, qui étaient sur-représentées pour l'agent initial sont bien moindres pour l'agent modifié qui les remplace par une série d'actions variées. L'action 2, qui correspond à une vente d'obligations à hauteur de 8% accompagnée d'une baisse de la duration, était l'action choisie dans 30% des cas par l'agent initial. Elle passe à 17% pour l'agent modifié. On peut faire l'hypothèse que l'approche plus long-terme de l'agent modifié l'amène à élaborer des stratégies plus complexes.

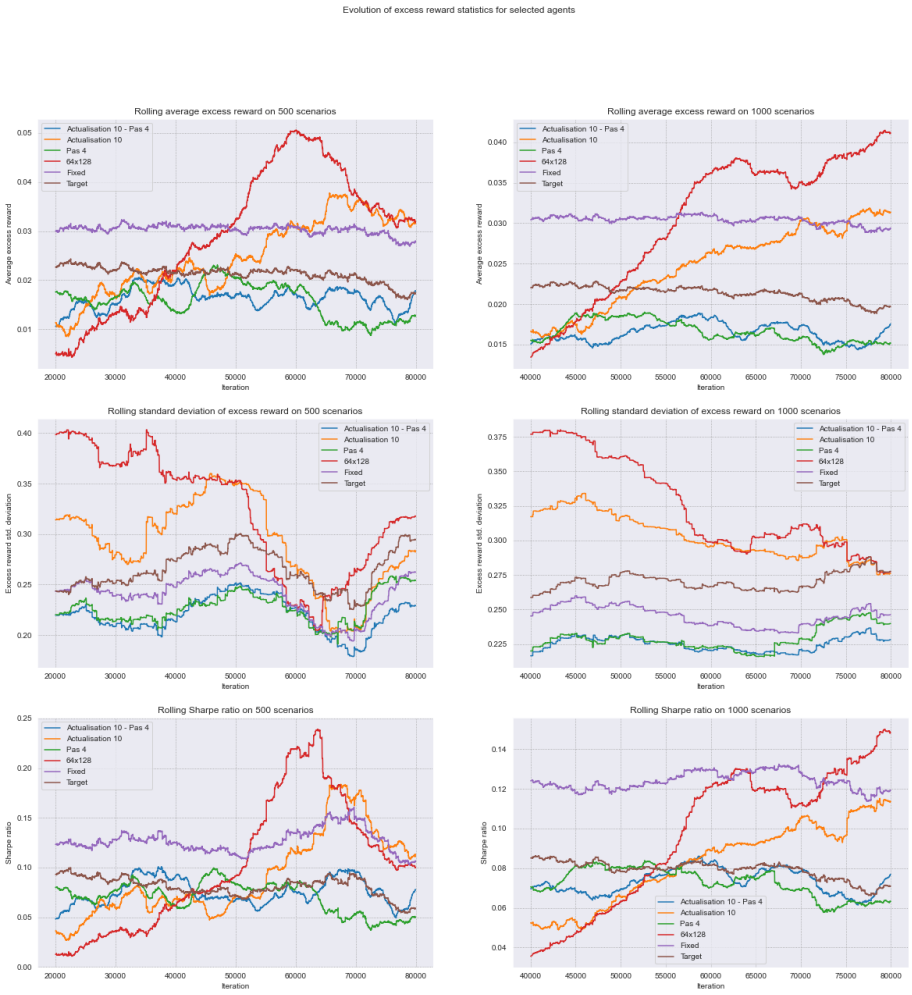


FIGURE 5.4 – Evolution des statistiques des *rewards* relatifs avec optimisation des paramètres

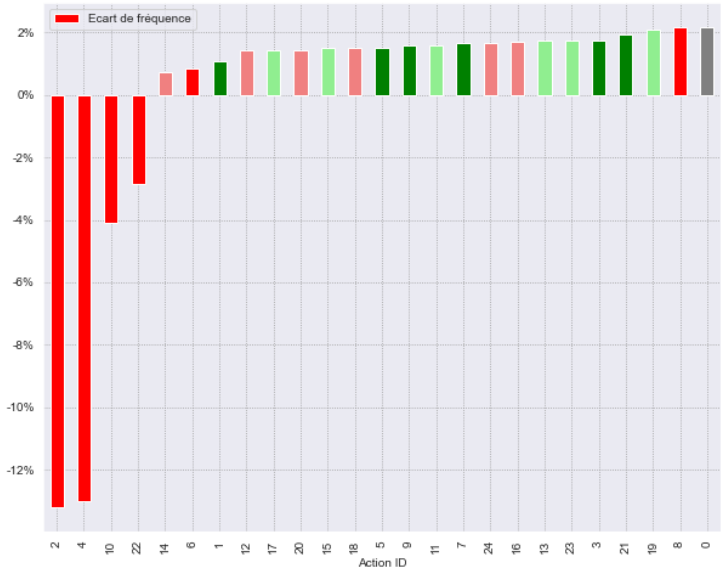


FIGURE 5.5 – Histogramme des actions prises par l'agent

5.4 Points d'amélioration

5.4.1 Amélioration de l'environnement

Dans le cadre de cette étude, il a été volontairement choisi de concevoir un modèle plus simple que le modèle interne utilisé au sein d'AXA France et qui ne représente pas l'intégralité des contraintes auxquelles est soumis l'assureur. L'une d'entre elles est la réserve de capitalisation dans laquelle doivent être placées les plus ou moins-values obligataires réalisées par l'assureur. Il s'agit d'une contrainte importante dans la gestion du portefeuille qui n'est pas modélisée dans notre modèle et qui aurait potentiellement un impact significatif sur le comportement de l'agent. Cela étant dit, il est possible de transposer les décisions de vente d'obligations par l'agent dans un contexte où elles impactent les produits financiers. Pour cela on peut mettre en place des SICAV obligataires qui peuvent externaliser les plus ou moins-values réalisées à l'intérieur via des distributions. De façon plus générale il est intéressant de doter l'agent de la capacité d'investir dans des OPCVM afin qu'il puisse piloter au mieux le résultat financier via des distributions ou des réalisations de plus-values à travers ces investissements.

D'autre part, le nombre de classes d'actifs modélisées reste assez limité. Dans un contexte plus réaliste, il est possible de mettre en place des stratégies pour profiter de primes d'illiquidité sur certains actifs comme le *private equity*, d'effet de diversifications importants au niveau du SCR ou encore de couvertures en utilisant des produits dérivés. Il est notamment assez handicapant pour l'algorithme de gérer la duration de l'actif via des achats et ventes d'obligations qui induisent des frais importants. Un développement de couvertures de taux comme des IRS ou des *swaptions* pourrait s'avérer utile. Par ailleurs, aucune exposition au risque de crédit n'est modélisée alors que les obligations d'entreprise représentent une part significative de l'actif d'un assureur vie.

5.4.2 Amélioration du modèle de *reinforcement learning*

Performance d'éxecution

L'analyse produite se base sur un nombre de scénarios limités faute de disponibilité de scénarios supplémentaires et de la complexité de mise en place d'un générateur de scénarios économiques performant pour y pallier. Par ailleurs, le temps d'exécution de l'analyse est également une variable limitante pour effectuer plus de simulations.

On observe en pratique que l'exécution d'un scénario prend environ 20 secondes de temps de calcul en local sur une machine de travail standard. Ainsi, pour un jeu de 2000 scénarios, une simulation s'exécute en environ 11 heures. Ce temps est important pour plusieurs raisons. D'une part, on fait appel 80 000 fois à la procédure de mise à jour de l'environnement qui doit entre autres revaloriser tous les actifs et les passifs en appliquant des exponentiations qui sont coûteuses en temps de calcul. D'autre part nous faisons appel au réseau de neurones pour prendre des décisions et effectuer des rétropropagations d'entraînement ou de calcul matriciel et des exponentiations (via la fonction sigmoïde) sont effectuées. Ces durées de simulation rendent impossible d'effectuer des recherches de paramètres optimaux de façon automatisée via des méthodes comme le *Grid Search*. Cela limite également la taille du vecteur d'état du système. On peut démontrer que la complexité d'application d'un réseau de neurones est en $O(n^3)$ où n est la dimension du vecteur d'état, comme explicité dans l'article Fredenslund (2021) d'un étudiant de l'Université de Copenhague. La rétropropagation est quant à elle en $O(n^4)$. Ainsi multiplier par 2 le nombre de variables d'état nous aurait amené à multiplier par près de 16 le temps de calibration du modèle.

Pour remédier à ce problème le plan d'action consisterait à faire un audit de la performance calculatoire de l'environnement programmé afin d'optimiser les fonctions utilisées et les aspects algorithmiques. Ensuite cet environnement optimisé pourrait être complié pour pouvoir être exécuté plus rapidement. Enfin, le processus d'apprentissage pourrait être parallélisé sur des GPU qui sont particulièrement adaptés à l'exécution de produits matriciels utilisés par les réseaux de neurones.

Variables d'état

Nous utilisons dans notre algorithme un nombre limité de variables pour représenter l'état du système. Avec une puissance de calcul très importante et un générateur de scénario, on pourrait envisager d'alimenter l'agent avec l'intégralité de la connaissance de l'environnement. Sans aller jusque là, plusieurs variables d'état pourraient avoir un impact bénéfique si elles étaient ajoutées au modèle. Par exemple :

— Taux d'inflation – il s'agit d'une variable en lecture directe dans les scénarios qui a une influence sur la satisfaction client.

— Taux de réinvestissement – le taux OAT 10 que nous utilisons en tant que taux sans risque peut aussi être utilisé pour obtenir une indication sur les revenus futurs générés par un investissement en obligations.

— TMG moyen du passif – il donne une indication sur les produits financiers minimum au-delà duquel l'assureur commence à générer de la marge.

Agent

Nous pouvons bien évidemment complexifier l'architecture du réseau de neurones utilisé par l'agent. Mais avant cela, il y a plusieurs paramètres du modèle que nous n'avons pas optimisés :

— Rétro propagation – il existe différents algorithmes d'optimisation effectuant l'entrainement par rétro propagation du réseau de neurones. Nous n'avons travaillé qu'avec l'algorithme *Adam* avec un taux d'apprentissage ou *learning rate* fixé à 0.001. Ce taux peut être optimisé pour avoir potentiellement de meilleurs résultats.

— Taux de *dropout* - les algorithmes qui optimisent le dropout pour limiter le risque d'*overfitting* restent à explorer. Ils pourraient permettre d'avoir un meilleur compromis entre précision et généralisation.

L'approche que nous avons implémenté est appelée *model-free* car notre agent choisit des actions à partir de sont expérience uniquement. Il existe des approches appelées *model-based* où l'agent dispose en plus d'un modèle qui est à sa disposition pour anticiper *a priori* les impacts de sa décision. Par exemple on pourrait fournir à l'agent un modèle qui à partir des rendements moyens et d'une matrice de covariance des actifs disponibles donne une estimation l'espérance et la volatilité de sa marge de l'année suivante en fonction de l'action choisie. L'agent peut alors se constituer des estimation de ces paramètres avec son expérience et appuyer le choix de son action par le résultat du modèle à sa disposition.

Reward

Nous avons choisi en *reward* le ROE de l'actionnaire qui est une métrique activement optimisée par les assureurs. Toutefois cette optimisation se fait généralement sous contraintes. Il y a d'une part la contrainte réglementaire de maintenir un niveau de solvabilité minimal et d'autre part une contrainte économique forte : éviter la faillite de l'assureur. Dans le contexte de *reinforcement learning* ces contraintes peuvent être matérialisées par des pénalités appliquées au *reward*. Ainsi, on pourrait déduire une pénalité forfaitaire au *reward* à chaque année où le ratio de solvabilité effectif est inférieur à un seuil minimal. En ce qui concerne la faillite, il est plus difficile de choisir un critère simple car on peut imaginer que l'actionnaire refinance le capital de l'assureur. On peut toutefois imaginer un suivi du *drawdown* du *reward* qui pourrait amener à une pénalité lorsqu'il dépasse un seuil fixé. Dit autrement, on peut considérer qu'après plusieurs années consécutives de pertes l'assureur se retrouve en faillite. Cela entrainerait d'ailleurs un arrêt du scénario équivalent au *game over* du jeu de Tetris.

Stratégie d'apprentissage et sélection de modèle

Etant donné temps requis pour générer une simulation avec apprentissage sur 2000 scénarios (environ 8h avec nos machines), nous n'avons pas formellement optimisé la stratégie d'apprentissage en modulant le taux d'exploration. Les résultats du paragraphe précédent ont été générés en utilisant un taux de 20% qui dans était le meilleur paramètre parmi ceux que nous avons testé. Avec un temps d'exécution plus court, nous aurions effectué une recherche par dichotomie pour trouver l'optimum.

Ayant choisi d'utiliser tous les scénarios à notre disposition pour séléctionner le modèle, nous avons mesuré les performances en phase de test sur un échntillon de ce jeu de scénarios. Il existe

donc un risque d'erreur dans la sélection du modèle, il est possible qu'une architecture différente aurait été plus adaptée en analysant des résultats de test sur des scenarios de validation. Par exemple on aurait pu envisager de prendre un jeu de scénarios générés à une autre période avec un paramétrage différent du générateur de scénarios. Nous aurions pu aussi démultiplier le nombre de scenarios d'apprentissage en disposant d'un générateur de scenarios pour améliorer encore les résultats. D'ailleurs, il serait même possible de brancher directement un générateur de scenarios au modèle afin de faire un nombre arbitraire de simulations.

5.5 Champs d'application

Les résultats encourageants que nous obtenons ouvrent des perspectives d'application de notre approche dans plusieurs domaines. L'application la plus directe de notre étude est dans la réalisation d'études ALM pour une compagnie d'assurance. Il serait possible d'intégrer dans un cadre de *reinforcement learning* les modèles qui sont utilisés actuellement par les compagnies afin d'entrainer un agent prenant des décisions d'allocation. Contrairement aux études traditionnelles qui fournissent des allocations optimales statiques, l'analyse des décisions de l'agent pourrait fournir une stratégie dynamique avec un plan de convergence de l'allocation actuelle à une allocation cible en minimisant les frais de transaction, en optimisant les taux de produits financiers servis et en respectant les contraintes de liquidité des actifs. De plus, étant donné que nous ne mettons pas de contraintes de gestion *a priori* notre modèle peut explorer de nouvelles stratégies d'investissement non intuitives.

L'absence de stratégie prédéterminée permet à l'agent de parcourir l'ensemble des actions possibles et d'identifier des arbitrages de modèle qui lui permettent de générer du *reward* en exploitant des failles de l'environnement. Il existe de nombreux exemples où des agents arrivent à créer des situations avantageuses en exploitant des propriétés de l'environnement mises en place de façon non intentionnelles. Par exemple, un article d'OpenAI, Clark & Amodei (2016), présente le cas d'un jeu de conduite de bateau où à la place de finir la course rapidement, le bateau tourne en rond afin d'améliorer son score en percutant des objets. Nous pouvons utiliser ce constat à notre avantage pour auditer les différents modèles utilisés en leur connectant un agent de *reinforcement learning* . Celui-ci va, au terme d'un nombre d'itérations suffisant, identifier les opportunités d'arbitrage au sein des scénarios économiques ou encore des erreurs de modélisation du passif qui pourraient être avantageuses pour lui. Grâce à une analyse du comportement de l'agent, il est possible de corriger le fonctionnement du modèle.

Il serait évidemment intéressant de faire le lien entre notre étude et le modèle interne d'AXA France. En remplaçant l'algorithme ALM de celui-ci par un agent entraîné on observerait une amélioration des indicateurs de valeur générée pour l'actionnaire et potentiellement de solvabilité. Cela aurait un impact financier significatif étant donnée la valeur générée par l'augmentation de la solvabilité de celui-ci. Cependant, deux problèmes se présentent à nous : premièrement il s'agit d'un défi important d'un point de vue informatique. Le modèle interne est bien plus complexe que l'environnement que nous avons développé et il est techniquement compliqué d'y introduire cet agent et de l'entraîner. Le second défi est l'approbation par le régulateur des indicateurs générés par un tel modèle interne. Il est difficile d'interpréter le fonctionnement d'un réseau de neurones sous la forme d'un ensemble de règles qui peut être présenté sous forme d'arbre de décision pour expliquer les choix de gestion ALM. Bien qu'il existe des outils permettant d'expliquer les résultats d'une prédiction faite par un réseau de neurones, une approbation d'un modèle interne avec un agent de *reinforcement learning* requiert un changement en profondeur de la validation de modèle. On pourrait par exemple envisager une validation par échantillonnage : le régulateur fournirait à l'assureur un jeu de scénarios spécifiques et analyserait les sorties de modèle pour ces scénarios en

jugeant de la pertinence des actions effectuées par l'algorithme d'allocation d'actif.

Au-delà de la gestion actif/passif de portefeuilles d'assurance vie investie sur des supports euro, on peut imaginer des applications dans le domaine de la gestion d'actifs et des produits d'épargne en unités de compte. En effet, de nombreux assureurs, dont AXA France, commercialisent des offres de gestion pilotée ou sous mandat où les allocation UC des clients sont rebalancées régulièrement avec des allocations dépendant du contexte des marchés financiers et des vues des gérants. Il est possible d'entraîner un agent de *reinforcement learning* sur des données historiques et des scénarios pour sélectionner les allocations optimales et maximiser le ratio d'information de la stratégie, à savoir la surperformance de la poche sous mandat par rapport à un *benchmark* normalisée par l'erreur de suivi. Il s'agit d'un problème très similaire au nôtre dans un cadre de gestion d'actifs.

5.6 Gouvernance d'un modèle de *reinforcement learning*

En juin 2020, l'ACPR a publié un rapport sur la gouvernance des algorithmes d'intelligence artificielle dans le secteur financier (Dupont et al. (2020)). Dans ce document elle définit un cadre d'évaluation et de gouvernance pour les modèles de *Machine Learning* qui s'appliquerait à une éventuelle mise en place du modèle présenté dans ce mémoire, dans le cadre d'une prise de décision d'allocation au sein du modèle interne.

L'évaluation du modèle doit être faite selon les quatre angles suivants (5.6) :

FIGURE 5.6 – Axes d'évaluation d'un modèle de *Machine Learning*

— **Le traitement de données** : dans notre cas, l'algorithme utilise uniquement des données du modèle interne qui sont déjà inscrits dans des processus de validation garantissant leur qualité.

— **L'explicabilité** : interprétabilité du modèle pour les utilisateurs internes. Elle répond d'une part à la question de savoir comment fonctionne l'algorithme (ex : vendre des actions lorsque les taux d'intérêt sont au-dessus d'un certain niveau) et d'autre part pourquoi l'action est faite (ex : pour réinvestir en obligations à un taux plus haut et réduire le SCR). L'ACPR introduit différents niveaux d'explications allant de 1 à 4 :

  — Le premier niveau est l'observation : il s'agit de donner les résultats pour un jeu de

données d'entrée.

— Le second niveau est la justification : il s'agit de donner des justifications de ces résultats par la génération de justifications par l'algorithme. Il peut s'agir, par exemple, d'indicateurs synthétiques ou KPI constitués à partir des variables d'entrée qui sont constitués par la première couche de neurones.

— Le troisième niveau d'explicabilité est l'approximation. Il s'agit de donner les éléments nécessaires pour approximer la décision du modèle par une méthode simple.

— Le quatrième niveau d'explicabilité est la réplication. Il s'agit d'analyser le code source et de donner les paramètres et données suffisantes au régulateur pour répliquer le modèle.

Le tableau 5.6 donne un exemple d'explications qui sont exigés pour une implémentation de modèle de *Machine Learning* dans un modèle interne. Ainsi il faut fournir des explications de niveau 4 lors du changement de modèle à une équipe de validation qui serait constituée de représentatnts du *Risk Management* et des équipes des investissements. En plus de cela des explications de niveau 2 seraient exigées tout au long de la vie du modèle par les organes d'administration et de surveillance.

— **La stabilité** : capacité de l'algorithme à généraliser son apprentissage à de nouvelles données et à maintenir ses performances dans le temps. Nous avons pu tester la stabilité du modèle en effectuant un test sur des données nouvelles. Toutefois il est nécessaire de faire des tests complémentaires et de continuer à observer la performance du modèle lors de sa mise en œuvre pour assurer une revue du paramétrage voire un réapprentissage lors d'un décrochage de performance en production.

— **La performance** : l'utilisation d'un modèle complexe doit être justifiée par une performance supérieure à une approche traditionnelle qui, en général, est plus explicable et plus stable. Nous avons vu dans la section précédente que notre algorithme surperforme légèrement une stratégie basée sur l'expérience qui est communément utilisée en ALM. Les résultats ne sont pas encore suffisants pour justifier la mise en production d'un tel modèle mais il s'agit d'un champ d'études intéressant qui pourrait amener à une surperformance bien plus significative.

En termes de gouvernance, la réglementation exige une déclaration de changement de modèle lorsque celle-ci est jugée matérielle par l'assureur. L'utilisation d'un modèle de *RL* pose la question de la matérialité de la modification de celui-ci. On pourrait envisager une métrique basée sur les valeurs des paramètres mais celle-ci serait insuffisante car de faibles changements de coefficients peuvent avoir des impacts significatifs. Il est plus adapté de déclencher de telles déclarations sur la base de résultats de *backtesting* lorsqu'une variation importante de performance est observée. En ce sens, l'ACPR considère qu'un modèle d'IA se rapproche d'un modèle interne classique car on peut faire un parallèle entre les processus de calibration de paramètre de modèle fait par des experts avec l'apprentissage de paramètres.

L'ACPR appelle également à adapter les méthodes d'audit interne et externe, notamment dans le cadre d'une de leurs missions, aux algorithmes de *Machine Learning*. Plusieurs solutions innovantes sont proposées pour permettre à un auditeur d'effectuer son contrôle :

— Solution 1 – « *benchmarking* » : l'auditeur fournit un jeu de données de test qui est ingégré par

Cas d'usage			Critères d'explicabilité			Niveau d'explication requis
Domaine	Processus métier	Fonctionnalité de l'IA	Audience de l'explication	Contexte	Risque associé	
Modèles internes	Conception du modèle	Calcul des ratios de solvabilité	Équipe de validation	Validation des modèles, et de la politique de changement de modèle	- Risque de modèle (de solvabilité) - Risque de conformité	4
			Organes d'administration, de gestion ou de surveillance	Approbation	- Risque de modèle (de solvabilité) - Risque de conformité	2

TABLE 5.7 – Niveaux d'explicabilité requis pour la mise en place d'un modèle de *Machine Learning*

l'algorithme. Les résultats sont ensuite transmis par l'assureur. L'idée est d'avoir des cas de test pour lesquels l'auditeur a des résultats réels ou éventuellement déterminés par un modèle de référence. Il peut ainsi comparer les jeux de résultats et mesurer la performance du modèle testé. Cette méthodologie s'approche du concept de Kaggle, une plateforme organisant des compétitions de *data science* où les participants soumettent les prédictions de leurs modèles qui sont ensuite confrontés à la réalité connue par l'organisateur.

— Solution 2 – mise en concurrence de modèles : ici c'est l'auditeur qui fournit directement à l'assureur un modèle qu'il a développé pour répondre au problème donné. Ce modèle est implémenté par l'assureur qui doit démontrer que le sien est plus performant. Dans notre cas de figure il s'agirait de montrer sur un jeu de scénarios assez large que la stratégie de l'agent développée génère un *reward* plus important. Dans d'autres cas de figure, comme par exemple le traitement de sinistres, on peut penser à une approche par *A/B testing* où une partie des cas est gérée par le modèle testé et l'autre par le modèle *challenger*.

Conclusion

Dans ce mémoire, nous avons pu étudier en détail un portefeuille de retraite collective en analysant le contexte assurantiel et la modélisation de celui-ci dans le modèle interne d'AXA France. Cela nous permis de dresser des parallèles entre le problème de la gestion de l'allocation d'actifs et des environnements de jeu tels que Tetris. Par conséquent, nous avons entrepris la résolution de celui-ci dans un cadre du *reinforcement learning* en mettant en place un modèle simplifié, une métrique cible à maximiser et un agent muni prenant des décisions d'allocation en s'appuyant sur un réseau de neurones entraîné sur la base de ses actions et *rewards* passées. Pour la gestion du portefeuille en *run-off*, l'étude montre qu'un agent choisissant l'action ayant le score maximal prédit par réseau de neurones dense de dimension 64x128 est plus performant que le suivi d'une allocation stratégique cible. Cette surperformance est caractérisée par un rendement sur capital supérieur pour l'actionnaire en espérance avec un risque identique. Il ne s'agit que d'un début d'exploration des applications du *deep learning* à l'ALM étant donné que de nombreux paramètres restent à optimiser et que l'environnement ainsi que la stratégie d'entraînement peuvent être largement améliorés.

Le développement croissant de modèles d'apprentissage artificiels en assurance soulève la question de leur interprétabilité et de leur contrôle notamment par le régulateur dans un contexte de calcul d'indicateurs officiels. En effet, la plupart de ces modèles sont, une fois entrainés, des boîtes noires qui ne peuvent pas être résumées à des règles de décision simples. Le cas échéant, il faut réfléchir à de nouvelles approches de contrôle et de validation. Pour les modèles d'apprentissage supervisé, il existe des métriques permettant de donner l'importance des différentes variables et le sens de leur impact, comme les *SHAP values*. Pour les modèles de *reinforcement learning* la question est plus complexe. Des approches cherchent à étudier des « circuits » de neurones connectés entre eux par des poids élevés afin d'en déduire des *features* créés à partir des données d'entrée qui seraient interprétables.

Dans notre cas de figure, le régulateur pourrait s'intéresser à un ensemble de scenarios économiques spécifiquement choisis pour induire un comportement évident pour un expert. Ces scénarios seraient alors joués dans l'environnement de l'assureur par son algorithme d'allocation. Les résultats peuvent ensuite être analysés en s'assurant que l'algorithme se comporte façon cohérente. Par ailleurs, des critères de stabilités pourraient également être étudiés en envoyant un jeu de scénarios proches et en vérifiant que le comportement en sortie est suffisamment stable.

Bibliographie

Arnaud, F. (2021), 'Les retraités et les retraites'.
URL: *https ://drees.solidarites-sante.gouv.fr/publications-documents-de-reference/panoramas-de-la-drees/les-retraites-et-les-retraites-edition-0*

Clark, J. & Amodei, D. (2016), 'Faulty reward functions in the wild'.
URL: *https ://openai.com/blog/faulty-reward-functions/*

Dupont, L., Fliche, O. & Yang, S. (2020), Gouvernance des algorithmes d'intelligence artificielle dans le secteur financier, Technical report, ACPR. **URL :** https ://acpr.banque-france.fr/sites/default/files/medias/documents/20200612_gouvernance$_e$*valuation_ia.pdf*.

Fredenslund, K. (2021), 'Computational complexity of neural networks'.
URL: *https ://kasperfred.com/series/introduction-to-neural-networks/computational-complexity-of-neural-networks*

Kanervisto, A., Scheller, C. & Hautamäki, V. (2020), 'Action space shaping in deep reinforcement learning', *CoRR* **abs/2004.00980**.
URL: *https ://arxiv.org/abs/2004.00980*

Krabichler, T. & Teichmann, J. (2020), 'Deep replication of a runoff portfolio'.
URL: *https ://arxiv.org/abs/2009.05034*

Nielsen, M. A. (2018), 'Neural networks and deep learning'.
URL: *http ://neuralnetworksanddeeplearning.com/*

Sutton, R. S. & Barto, A. G. (2018), *Reinforcement Learning : An Introduction*, second edn, The MIT Press.