

Mémoire présenté le : 07/09/2021

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : SHALUPENKA Yuliya
Titre : Construction de la table de maintien
d'expérience pour la garantie rente éducation au
sein du portefeuille de Prévoyance Collective
Confidentialité : NON OUI (Durée : 1 an 2 ans)

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

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

*Entreprise : AXA France
Signature :*

C.PIGEON, R.EI SOKHON, G.BIESSY

Nom :

Signature :

Membres présents du jury de l'ISFA

Directeur de mémoire en entreprise :

Nom : Brahim JAMALEDDINE

P.RIBEREAU

Signature :


Invité :

Nom :

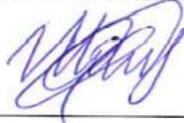
Signature :

**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 du candidat



Résumé

Mots clés : rente éducation, modèle de durée, table d'expérience, exposition au risque, taux de hasard, fonction de survie, espérance de durée résiduelle, estimateur de Hoem, estimateur de Kaplan-Meier, lissage, Whittaker-Henderson, hasard proportionnel, modèle de Cox, provisions, SMR

Ce mémoire a pour objectif d'étudier la loi de durée de présence des individus dans le champ d'application de la garantie rente éducation et d'estimer son impact sur le calcul des engagements de la garantie. L'intérêt de cette étude est de proposer une approche alternative pour la construction de la loi d'expérience et réfléchir à de potentiels usages. L'étude porte sur un portefeuille de rentes temporaires dans le cadre de garantie proposée dans les contrats de prévoyance collective.

L'appréciation du risque de prolongement de la durée des études est, en effet, indispensable pour l'assureur. Le provisionnement des garanties en cas de vie ou de décès est encadré par les articles A335-1 du Code des Assurances et A931-10-10 du code de la sécurité sociale. Ces derniers permettent aux assureurs de mettre en place, sous respect de certaines conditions, des tables d'expérience. Étant donné que le portefeuille des bénéficiaires de la rente éducation est souvent composé d'enfants provenant d'un milieu social favorisé, l'utilisation des données disponibles au niveau national semble ne pas être prudente, surtout dans le contexte de la hausse des inégalités sociales.

Avant de commencer l'étude de la base de données issue du portefeuille de l'assureur, les données disponibles au niveau national ont permis d'avoir une première idée sur l'évolution du système éducatif en France. Une loi de maintien de rente fictive a été construite à partir des données nationales dans le but de créer une référence pour la comparer avec la table d'expérience retenue.

Plusieurs méthodes ont été exploitées pour la construction de la loi de maintien d'expérience : les estimateurs non-paramétriques de Hoem et de Kaplan Meier, le modèle semi-paramétrique de Cox et le modèle segmenté. Les résultats obtenus par les différentes méthodes ont été analysés et soumis à une validation par de différents tests pour être comparés à la fin dans le but de ne retenir qu'une seule loi d'expérience.

Enfin, la loi retenue a été utilisée pour le calcul des engagements de la garantie rente éducation. Ces montants ont été comparés avec ceux calculés à l'aide de la loi d'expérience utilisée au sein de l'équipe depuis 2014 et actualisée avec la base de données construite dans le présent mémoire. La performance de la nouvelle loi a également été évaluée par la comparaison des paramètres estimés obtenus avec les données réelles.

Abstract

Keywords: education annuity, duration model, experience table, risk exposure, hazard rate, survival function, residual time expectancy, Hoem, Kaplan-Meier estimator, smoothing, Whittaker-Henderson, proportional hazard, Cox model, reserves, Standardizes Mortality Ratio

This essay aims to study the duration model within the scope of the education annuity and to estimate its impact on the calculation of reserves. The interest of this study is to suggest an alternative approach for the elaboration of the experience law and to reflect on its potential uses. This work takes place in the framework of a temporary annuity portfolio and within the personal protection policy contracts.

Being able to assess the risk related to continued education is key to the insurance company. The provisioning of life insurance and death benefit coverages is defined in the articles A335-1 of the Insurance code and A931-10-10 of the Code of Social Security. These articles allow insurers to set up experience tables if certain conditions are met. Given that most of the beneficiaries of the education annuity come from a privileged background, it would appear unwise to use data at the national level, and even more so in the context of rising educational inequalities linked to social background.

Before we started studying the database derived from the insurer's portfolio, we have used the data that is available at the national level to get a first glimpse of the evolution of the French education system. An abstract life-table model has been built and used as a reference for comparison with the retained final table.

The following models have been used for the elaboration of the abstract life-table model: The Hoem estimator, the Kaplan-Maier estimator, the Cox proportional hazards model and the model with segmentation by sex of beneficiary. The results obtained following the application of the models have been analyzed and controlled and validated by a set of tests. They were then compared so that we could retain only one experience law.

Lastly, the retained law has been used to calculate the reserves of the education annuity. These amounts were compared with those calculated using the law of experience used within the team since 2014 and updates with the database built into this essay. The performance of the new experience table is also assessed through a comparison of the estimated parameters obtained with the real data.

Remerciements

Je voudrais tout d'abord remercier Monsieur Brahim JAMALEDDINE, qui m'a soutenue et m'a orientée dans mon projet. Il a su se rendre disponible et partager ses connaissances et son expertise malgré un emploi de temps très chargé.

Je souhaite remercier Monsieur Frédéric PLANCHET pour ses conseils précieux ainsi que pour ses interventions en tant qu'enseignant dans le cadre de la formation au sein de l'Institut de Science Financière et d'Assurance.

Je tiens aussi à remercier l'équipe pédagogique de l'Institut, particulièrement Madame Anne EYRAUD-LOISEL pour sa formation de haute qualité et son soutien permanent auprès des étudiants.

Mes remerciements s'adressent aussi à mes collègues rencontrés à des moments différents au sein d'AXA France, plus particulièrement Romain, Abdellatif, Pascal, Camille, Renaud, Vladimir, Mamadou et Younes pour leur aide et leur bonne humeur au quotidien.

Je souhaite remercier ma famille, mes parents en particulier, et mes amis qui ont su me soutenir, me changer les idées quand il le fallait et m'encourager durant ces années de formation. Enfin, un merci particulier à mon conjoint Mathieu pour son soutien quotidien et pour m'avoir choisie comme épouse pendant la préparation de ce mémoire.

Table des matières

Résumé.....	2
Abstract	3
Remerciements	4
Table des matières	5
Introduction.....	7
Chapitre 1. Contexte de l'étude	9
1.1 Garantie Rente Education	9
1.2 Tables d'expérience et rôle d'actuaire	10
1.3 Système d'éducation en France	11
1.3.1 Statistiques descriptives.....	12
1.3.2 Construction de la loi de sortie de référence.....	19
Chapitre 2. Construction de la base de données	21
2.1 Extraction des données	21
2.2 Retraitement des données.....	22
2.3 Statistiques descriptives des données	26
Chapitre 3. Construction de la loi de maintien d'expérience.....	36
3.1 Principales étapes du calcul de la loi de maintien.....	36
3.1.1 Définitions et notations.....	36
3.1.2 Limites de la méthode	38
3.1.3 Détection des observations censurées.....	38
3.1.4 Calcul d'exposition au risque.....	40
3.2 Estimation des taux bruts de sortie par les méthodes non paramétriques.....	42
3.2.1 Estimateur de Hoem.....	42
3.2.2 Estimateur de Kaplan Meier	44
3.3 Lissage des taux bruts de sortie	47
3.3.1 Lissage de l'estimateur de Hoem	49
3.3.2 Lissage de l'estimateur de Kaplan Meier.....	54
3.3.3 Comparaison des résultats	59
3.4 Méthode semi-paramétrique : analyse d'hétérogénéité.....	61
3.4.1 Modèle de Cox : modèle à risques proportionnels	62
3.4.2 Hypothèses du modèle.....	64
3.4.3 Vraisemblance du modèle de Cox.....	65
3.4.4 Choix des variables et mise en place du modèle.....	66

3.5 Loi de maintien segmentée	77
Chapitre 4. Validation de la loi de maintien d'expérience et son impact sur le provisionnement	81
4.1 Validation des résultats	81
4.1.1 Analyse graphique	81
4.1.2 Calcul de SMR	82
4.2 Impact du changement de la loi de sortie sur les provisions	85
4.2.1 Calcul des provisions	85
4.2.2 Comparaison des résultats	88
4.2.3 Limites de l'approche et axes d'amélioration	89
Conclusion	93
Bibliographie.....	95
Table de figures	98

Introduction

L'enseignement est une étape majeure de la vie de chaque citoyen. En effet, les études occupent une place importante dans les projets de vie de la majorité des Français : les études deviennent de plus en plus accessibles, le besoin de se former est de plus en plus marqué. La volonté des individus à trouver leur vocation et à devenir expert dans ce domaine semble les préoccuper de plus en plus.

Les chiffres ne font que confirmer l'intérêt que portent les Français pour les études : le nombre d'élèves et d'étudiants recensé en 2019 en France Métropolitaine et DROM a augmenté de +5,5% depuis 2010 et de +0,3% depuis 2018 selon les chiffres de l'INSEE. De plus, la durée des études ne cesse d'augmenter depuis 2009, et cela s'explique principalement par le fait que de plus en plus d'individus intègrent des formations d'enseignement supérieur.

L'évolution permanente du système d'enseignement et la croissance du nombre d'enfants scolarisés et diplômés s'accompagnent cependant d'un renforcement des inégalités sociales qui a un impact sur la scolarisation des enfants : selon le rapport du Conseil national d'évaluation du système scolaire, les inégalités scolaires liées aux origines sociales se sont approfondies sur le territoire français depuis l'année 2000.

Le présent mémoire concerne la garantie rente éducation qui permet aux enfants de bénéficier d'un soutien financier pour poursuivre les études en cas de décès d'un de ses parents. La hausse de l'effectif présent dans le système éducatif, la croissance de l'espérance de durée des études et l'accentuation des inégalités scolaires incitent les sociétés d'assurance à contrôler régulièrement l'état des stocks de provisions constituées pour faire face aux engagements pris envers les bénéficiaires de la rente éducation.

Le suivi des risques auxquels l'assureur doit faire face est motivé davantage par le contexte réglementaire avec la mise en place de la norme IFRS 17 qui doit entrer intégralement en vigueur à partir du 1^{er} janvier 2023. La norme impose, entre autres, une comptabilisation précise des contrats d'assurance ce qui nécessite de bien valoriser et connaître les risques encourus.

Le présent mémoire a été réalisé au sein de la direction actuariat assurance de personnes et plus exactement dans le pôle Actuariat Prévoyance Collective et Santé de la société AXA France et a pour objectif d'améliorer l'évaluation des risques liés à la garantie Rente éducation.

A ce jour, selon la Convention collective nationale du personnel des prestataires de services dans le domaine du secteur tertiaire du 13 août 1999, le provisionnement de la garantie rente éducation doit s'effectuer selon l'approche « tête par tête » et à l'aide des tables de mortalité réglementaires par sexe et par génération (TGF 2005/TGH 2005) et les taux techniques en vigueur (Planchet F., 2007, Tables de mortalité d'expérience pour les portefeuilles de rentiers). Il n'y a pas donc d'encadrement précis à propos de la loi de maintien

en études à utiliser, ce qui explique l'intérêt de la construction d'une table d'expérience, basée sur les données issues du portefeuille de la compagnie d'assurance. Cette table permettra de mieux refléter le comportement des rentiers et de capter les différences par rapport à la population française. En effet, il est judicieux de supposer que la souscription de la garantie rente éducation concerne plutôt les populations favorisées dont les membres sont davantage susceptibles de faire de longues études. C'est une raison de plus pour ne pas se baser sur les données nationales, mais de se référer plutôt aux données propres à l'assureur.

La première partie du présent mémoire définit de manière détaillée le cadre de l'étude, notamment la définition de la garantie rente éducation et l'explication de l'utilité de la construction de la table d'expérience. Elle inclut également un point sur la situation du système éducatif en France, ainsi qu'une description de la construction de la loi fictive basée sur les données nationales. La deuxième partie du mémoire présente la construction de la base de données pour la mise en place de la loi d'expérience, ainsi qu'une description de la population des rentiers d'AXA. La partie suivante, qui représente le cœur du présent mémoire, décrit le travail de la construction de la table d'expérience, en abordant les aspects théoriques et en les appliquant aux données réelles. Enfin, la dernière partie de l'étude aborde l'appréciation de la qualité de la loi d'expérience construite, ainsi que les réflexions sur ses usages potentiels, ses limites et les axes d'améliorations envisageables.

Chapitre 1. Contexte de l'étude

1.1 Garantie Rente Education

L'une des offres commerciales proposées par AXA France à ses clients est Adaptalia Prévoyance. Elle permet aux dirigeants d'entreprises de choisir une formule adaptée et personnalisée mais aussi aux salariés de souscrire des garanties complémentaires qui protègent leur famille et les ayants droits des assurés en cas de décès de l'assuré. Les garanties de type assurance décès incluent plusieurs sous-garanties (Baillard E., 2011, Pilotage prévoyance individuelle) :

- Décès toutes causes qui assure le versement d'un capital aux bénéficiaires désignés par l'assuré
- Rente de conjoint qui garantit le versement d'une rente au conjoint de l'assuré dans le cas de son décès
- Rente éducation qui prévoit le versement d'une rente à chaque enfant à charge de l'assuré dans le cas de son décès. Elle permet de pallier le manque de revenus du parent et couvre ainsi une partie du coût lié à l'éducation jusqu'à la fin des études des enfants bénéficiaires
- Décès accidentel qui multiplie par deux le capital décès toutes causes si le décès de l'assuré a été provoqué par un accident
- Décès accidentel mission qui garantit le versement du capital supplémentaire si le décès de l'assuré est survenu au cours d'une mission dans le cadre de son travail

La présente étude porte sur une garantie de type assurance décès, plus précisément sur la sous-garantie de rente éducation. Elle permet de compenser la perte de revenus du foyer dans le cas du décès de l'un ou des deux conjoints. La garantie sous-entend que l'assuré et le bénéficiaire ne peuvent pas être la même personne et donc exige l'existence de trois acteurs de contrat d'assurance au minimum.

La rente éducation peut être vue comme un contrat d'assurance dont l'objectif est de répondre aux besoins de financements du ou des enfants de l'assuré dans le cadre de leurs études. La garantie se déclenche généralement au décès de l'assuré. La rente peut cependant être versée de manière anticipée lorsqu'une invalidité totale et irréversible est déclarée chez l'assuré. Le contrat peut être souscrit par un seul parent ou par les deux simultanément.

En pratique, la rente éducation est versée trimestriellement jusqu'au 16^e anniversaire du bénéficiaire sans conditions de poursuite d'études. Si le bénéficiaire reste dans le système d'enseignement, la rente est versée jusqu'à ses 26 ans. Elle peut aussi être étendue jusqu'à ses 28 ans. En effet, jusqu'à l'année 2018 le statut d'étudiant reconnu par le système de sécurité sociale était limité à 26 ans. A partir de l'année scolaire 2019-2020, le régime étudiant de la sécurité sociale disparaît complètement, ce qui veut dire que la notion d'âge-limite du statut étudiant disparaît (<https://www.service-public.fr/particuliers/vosdroits/F675>). Les

compagnies d'assurance doivent alors adapter leurs offres par rapport à ce changement réglementaire. Il est notamment prévu d'étendre la limite d'âge de la garantie rente éducation. À l'heure actuelle, certains nouveaux contrats souscrits par AXA France fixent dorénavant l'âge limite du bénéficiaire à 28 ans, alors qu'il était de 26 ans auparavant. En revanche, pour les contrats qui existent déjà, la fin théorique de la rente survient au 26^e anniversaire du bénéficiaire. Il est donc naturel que l'âge de clôture de la majorité des rentes en cours du portefeuille soit actuellement fixé à 26 ans. La tendance changera au cours des prochaines années.

Le montant de la rente évolue avec l'âge de l'enfant. La hausse du montant prend effet le 1^{er} jour du trimestre qui suit l'anniversaire du rentier. Les conditions de versement de la rente prévoient une prise d'effet le jour suivant celui du décès de l'assuré et le versement à terme échu à la fin de chaque trimestre.

Au sein d'AXA France, la garantie Rente éducation vient généralement compléter la garantie décès.

Le tableau de la *figure 1* présente la formule minimale de la garantie rente éducation proposée aux cadres des entreprises par AXA Santé et Prévoyance Collectives dans l'offre Adaptalia.

Tranches d'âge		Assuré			
		Jusqu'à 40 ans	De 41 à 45 ans	De 46 à 50 ans	51 ans et plus
Enfant	0 – 10 ans	15 %	10 %	5%	-
	11 – 18 ans	20 %	15 %	10 %	-
	19 – 26 ans (si études)	25 %	20 %	15 %	-

Figure 1. Offre Adaptalia d'AXA France : parts utilisées pour le calcul de la rente par tranche d'âge d'assuré et de bénéficiaire

1.2 Tables d'expérience et rôle d'actuaire

Actuellement le provisionnement des domaines d'assurance vie/décès est encadré par une réglementation qui impose de nombreuses contraintes sur la table de survenance ou de maintien et sur le taux technique qui doivent être utilisés pour le calcul. Cependant les articles A335-1 du Code des Assurances et A931-10-10 du code de la sécurité sociale permettent aux assureurs de mettre en place, sous respect d'un certain nombre de conditions, des tables d'expérience. Elles doivent être construites à partir des données propres au portefeuille de l'assureur. L'intérêt des tables d'expérience est donc de mieux estimer les engagements envers les assurés. (Planchet F., Septembre 2020, Modèles de durée. Tables de mortalité)

En effet, il est intéressant et bénéfique pour un organisme d'assurance de prendre en compte toutes les différences qui existent entre sa population d'assurés et celle de référence. Même si la notion du comportement d'assuré n'est pas évidente à appréhender, l'assureur a tout intérêt à tenter de détecter les anomalies comportementales afin de les intégrer dans

son travail de provisionnement. Ceci s'avère utile et même nécessaire dans les situations où le risque étudié est sur ou sous-estimé dans la population de référence par rapport au risque réel du portefeuille d'assurés.

La construction de la table d'expérience consiste à compter l'occurrence d'un évènement dans le portefeuille de l'assureur. Il s'agit de l'évènement qui déclenche une garantie ou bien celui qui engendre sa cessation. Ce mémoire étudie l'évènement appelé « sortie de la rente éducation » qui peut survenir pour plusieurs causes possibles, comme la fin des études, le décès du bénéficiaire ou l'absence de documents justificatifs.

Le calcul du montant des engagements, qui est actuellement mis en place au sein de la branche Santé et Prévoyance Collectives d'AXA France est basé sur la table construite en 2014. Il s'agit de la loi qui combine la table de mortalité TPG93 jusqu'à l'âge de 16 ans avec la table de maintien d'expérience pour un intervalle allant de 16 à 28 ans. La modélisation de la table d'expérience de 2014 a été effectuée à l'aide de l'estimateur de Hoem. Eu égard au temps écoulé depuis la mise en place de la table, ce mémoire a pour objectif de proposer des approches alternatives pour sa construction.

1.3 Système d'éducation en France

Avant de commencer la construction de la table d'expérience, il est important d'analyser les données disponibles au niveau national. Chaque année, le ministère de l'éducation nationale publie, en collaboration avec le ministère de l'Enseignement supérieur, un ouvrage qui recueille de nombreuses données sur la formation et l'enseignement en France. Ces données sont étudiées dans l'objectif d'expliquer pourquoi leur utilisation ne serait pas bien adaptée aux besoins du provisionnement de la garantie Rente éducation d'AXA France.

L'ouvrage « Repères et références statistiques sur les enseignements, la formation et la recherche 2019 » publié par le ministère de l'éducation nationale, en collaboration avec le ministère de l'Enseignement supérieur, contient de nombreuses informations sur le système éducatif français.

L'étude de cet ouvrage et des données mises à disposition doit permettre de répondre à plusieurs questions importantes pour la suite du mémoire. Il est notamment intéressant de connaître la part des étudiants selon les filières d'enseignement, de savoir s'ils privilégient ou non les études longues ainsi que leur âge moyen lorsqu'ils quittent le système éducatif. Il semble également important de connaître l'origine sociale des élèves et son impact sur la durée de leurs études. En effet, le portefeuille des rentiers d'AXA France regroupe principalement des enfants d'origines sociales favorisées. L'utilisation des données nationales pour les travaux de provisionnement de la garantie crée dans ce cas-là un risque de sous-estimation de la durée de maintien en études.

1.3.1 Statistiques descriptives

Le système d'éducation français est composé de trois degrés qui sont partagés en cycles pour les deux premiers degrés. Le premier degré comprend les enseignements préélémentaires et élémentaires qui sont dispensés dans les écoles maternelles, élémentaires et primaires. L'enseignement du second degré est dispensé au collège puis au lycée. Quant à l'enseignement post-secondaire et supérieur, il est assuré par les lycées, les grandes écoles, les universités et les instituts spécialisés. Il faut également prendre en compte l'apprentissage qui représente une voie d'accès direct à l'enseignement professionnel.

L'instruction est obligatoire en France à partir de 3 ans et jusqu'à l'âge de 16 ans. Ainsi le cursus scolaire, qui débute à l'âge de 3 ans et qui se déroule sans redoublement, ni saut de classe, ni interruption, est appelé « âge théorique ».

Selon les données de la *figure 2*, lors de la rentrée scolaire 2019, le système éducatif public et privé a recensé 15,8 millions d'élèves, d'étudiants et d'apprentis en France métropolitaine et dans les DROM, soit une hausse de 50 000 personnes par rapport à la rentrée scolaire précédente. Cette hausse est principalement portée par l'enseignement supérieur avec une augmentation de 1,6% de l'effectif. Cette évolution montre la volonté et la motivation des élèves de poursuivre leurs études après le baccalauréat et augmente donc la durée des études. Pour un assureur, cela se traduit par une hausse du montant de provisions nécessaires pour respecter les engagements de la garantie Rente éducation. Une hausse des effectifs est aussi constatée au sein des établissements d'enseignement du second degré et des formations en apprentissage.

Evolution des effectifs d'élèves, d'étudiants et d'apprentis, en milliers	1980	1990	2000	2010	2015	2016	2017	2018	2019	Evolution 2019/2018 (%)
Premier degré Education nationale	7 396	6 953	6 552	6 664	6 805	6 806	6 783	6 750	6 704	- 0,7
Second degré Education nationale	5 309	5 726	5 614	5 353	5 536	5 579	5 629	5 645	5 677	0,6
Second degré Agriculture	117	116	151	150	145	143	141	137	137	- 0,5
Scolarisation dans établissement de la Santé	96	88	81	72	71	72	69	70	67	- 5,1
Apprentissage	244	227	376	434	411	417	435	453	479	5,8
Enseignement supérieur	1 181	1 717	2 160	2 299	2 510	2 555	2 622	2 682	2 725	1,6
TOTAL	14 347	14 828	14 935	14 972	15 478	15 573	15 680	15 737	15 788	0,3

Figure 2. Evolution du nombre d'élèves, d'étudiants et d'apprentis en France métropolitaine et DROM depuis 1980 (en milliers)

(Source : <https://www.education.gouv.fr/reperes-et-references-statistiques-sur-les-enseignements-la-formation-et-la-recherche-2019-3806>)

En revanche, une baisse des effectifs de 0,7% des formations du premier degré est constatée au niveau national. Ce phénomène est observé pour la troisième année consécutive et semble être contre-intuitive vis-à-vis du postulat selon lequel l'instruction serait obligatoire à partir de l'âge de 3 ans et jusqu'à l'âge de 16 ans. Il peut cependant être lié à l'évolution de la population de l'âge concerné. En effet, selon les données de l'INSEE apparaissant sur le graphique de la *figure 3*, le nombre de naissances en France est en baisse systématique depuis l'année 2010. Les enfants nés en 2010 ont 3 ans en 2013 et en 2014, il semble alors naturel de constater une baisse du nombre d'enfants inscrits dans le premier degré depuis 2014. Il ne

serait donc pas prudent de la part de l'assureur de tenir compte de cette baisse de l'effectif dans le cadre du provisionnement de la garantie rente éducation.

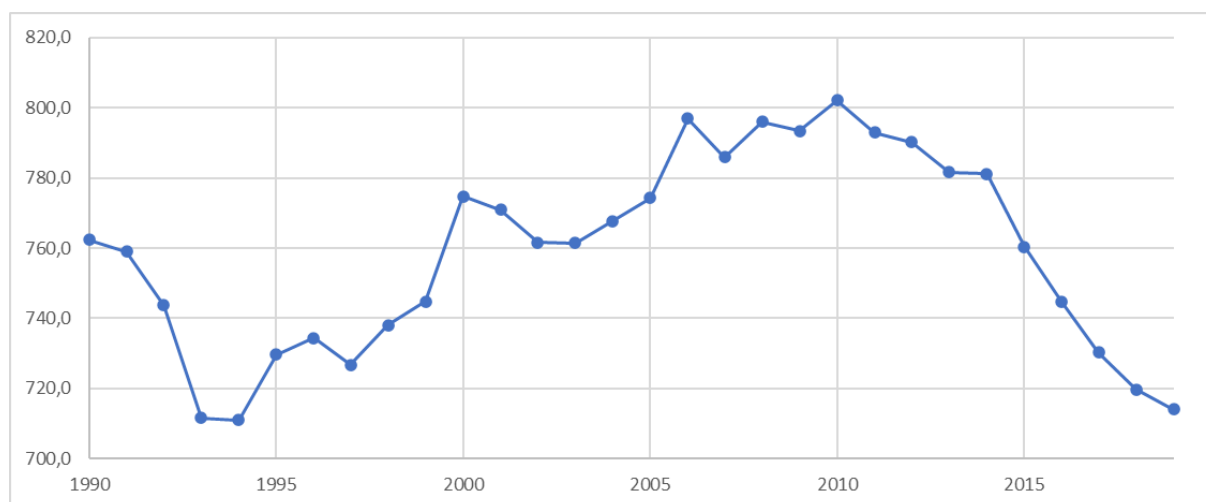


Figure 3. Nombre de naissances vivantes en France métropolitaine 1990-2019 (en milliers)

(Source : <https://www.insee.fr/fr/statistiques/4277635?sommaire=4318291>)

Selon les chiffres de la *figure 4* l'espérance globale de scolarisation est en hausse à la rentrée scolaire 2017. Cette évolution est principalement portée par le secteur de la formation supérieure, tandis que l'espérance de scolarisation dans le secondaire stagne depuis l'année scolaire 2014-2015. Selon le taux de scolarisation constaté pendant l'année scolaire 2017-2018, un jeune de 15 ans peut espérer suivre encore 6,5 années d'études en formation initiale dont 3,1 années dans l'enseignement supérieur. L'espérance de scolarisation dans un établissement supérieur est plus longue pour les filles (3,4 années) que pour les garçons (2,8 années). En outre, l'espérance de scolarisation dans un établissement secondaire est légèrement plus longue pour les garçons que pour les filles. Ce constat permet de formuler une hypothèse à propos du taux de redoublement qui semble être plus important pour les garçons que pour les filles.

Par ailleurs, l'espérance de scolarisation à 15 ans pour l'ensemble des formations reste quasiment inchangée depuis 1995-1996 avec uniquement la modification de la répartition entre les cycles d'enseignement. Cette observation permet de supposer que le modèle de scolarisation reste sans changements significatifs dans sa globalité depuis 1995. Il est donc possible de se baser sur un historique de données assez large pour étudier la durée globale des études.

Evolution de l'espérance de scolarisation à 15 ans, en année	1985-1986	1995-1996	2000-2001	2010-2011	2014-2015	2015-2016	2016-2017(p)	2017-2018(p)
Ensemble des formations	4,7	6,5	6,3	6,2	6,3	6,3	6,4	6,5
Garçons	4,6	6,3	6,2	6,0	6,1	6,1	6,2	6,3
Filles	4,8	6,6	6,5	6,4	6,5	6,6	6,6	6,7
Dont dans le Secondaire	3,4	4,0	3,9	3,6	3,4	3,4	3,4	3,4
Garçons	3,4	4,1	3,9	3,6	3,4	3,4	3,4	3,4
Filles	3,4	4,0	3,8	3,5	3,4	3,3	3,4	3,3
Dont dans le Supérieur	1,3	2,4	2,5	2,6	2,8	2,9	3,0	3,1
Garçons	1,2	2,1	2,2	2,3	2,6	2,7	2,7	2,8
Filles	1,4	2,7	2,7	2,8	3,1	3,2	3,3	3,4

Figure 4. Evolution de l'espérance de scolarisation à 15 ans depuis 1985-1986 en France métropolitaine et DROM (en nombre d'années)

(Source : <https://www.education.gouv.fr/reperes-et-references-statistiques-sur-les-enseignements-la-formation-et-la-recherche-2019-3806>)

Selon le graphique de la *figure 5*, en 2019, 34% des personnes de la tranche d'âge allant de 25 à 34 ans ont suivi une formation supérieure de longue durée à l'issue de laquelle ils ont obtenu un diplôme.

La part des personnes ayant suivi une formation supérieure longue est la plus importante chez les enfants des cadres et des professions intellectuelles supérieures (soit 67%). Par ailleurs cette proportion n'est que de 16% chez les enfants d'ouvriers. Il apparaît alors que l'origine sociale des personnes a un impact important sur la durée de leurs études.

Sur l'ensemble de la population, près de la moitié des personnes âgées de 25 à 34 ans ont obtenu un diplôme d'études supérieures en 2016 (soit 48% d'effectif total, dont 34% en formation supérieure longue). Ces chiffres ont un impact sur le provisionnement de la garantie rente éducation car cette dernière est effective à partir des 16 ans de bénéficiaires sous condition que ce dernier poursuive ses études.

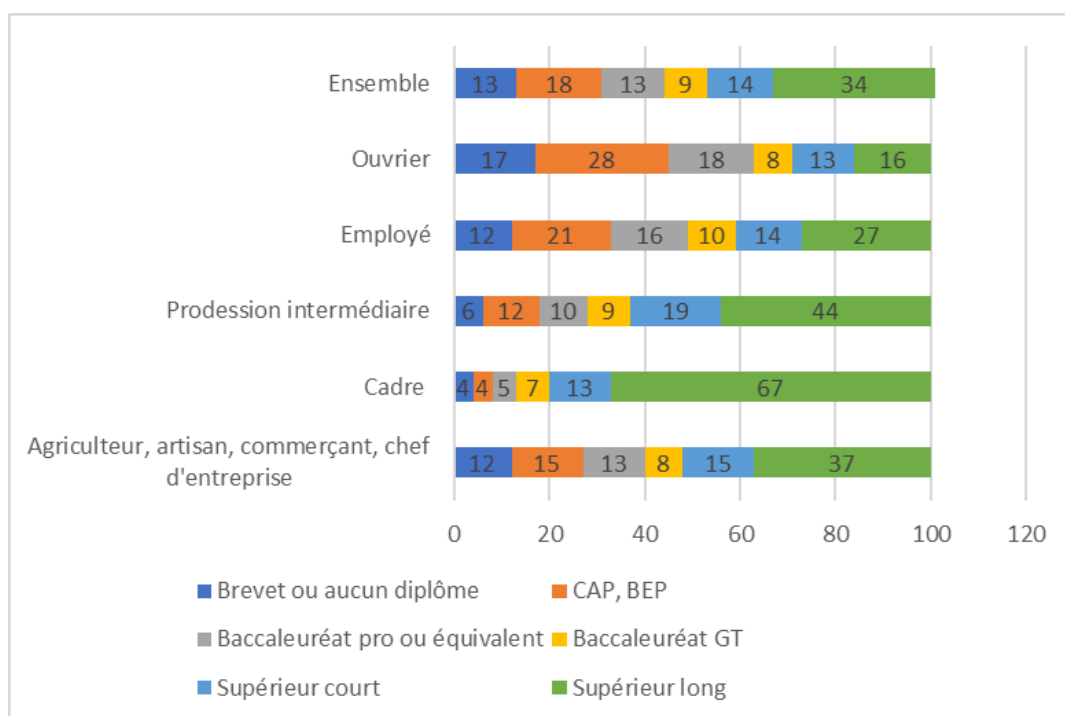


Figure 5. Niveau de diplôme le plus élevé des personnes de 25-34 ans selon le milieu social en France métropolitaine et DROM en 2019 (en %)

(Source : <https://www.education.gouv.fr/l-etat-de-l-ecole-2020-307185>)

Le tableau de la *figure 6* met en évidence la répartition des effectifs présents dans chaque cycle de l'enseignement en fonction de la catégorie socio-professionnelle de leurs parents. En 2019-2020, les enfants de cadres et de parents exerçant une profession intellectuelle supérieure représentent un tiers des étudiants des universités françaises. De plus, les enfants des cadres représentent plus de la moitié des inscrits en classes préparatoires aux grandes écoles, dans les écoles d'ingénieurs et les écoles normales supérieures ainsi que la moitié des effectifs des écoles de commerce. Les tendances observées s'expliquent par le coût élevé de ces formations.

Les enfants d'agriculteurs, artisans, commerçants et chefs d'entreprise sont les moins présents dans les universités (soit 9,1%). En revanche ils représentent 1/5^{ème} des effectifs dans les écoles de commerce (soit 19%). La part d'enfants d'ouvriers est relativement faible dans tous les types de formation à l'exception des formations de techniciens supérieurs (soit 23%).

Ces données permettent de poser une hypothèse selon laquelle l'origine sociale des parents aurait un impact non négligeable sur la présence des enfants dans les formations supérieures et donc sur la durée de scolarisation. Etant donné que l'objet de ce mémoire est l'étude du portefeuille des rentiers d'Axa France, on parlera plutôt des assurés de catégories sociales favorisées qui ont tendance à poursuivre des études longues. Ceci montre que l'utilisation des données nationales pour le portefeuille d'AXA France peut créer une sous-estimation de la durée espérée des études et donc un provisionnement insuffisant. Il est

important, dans ce contexte, de faire appel à la construction de la loi d'expérience qui est adaptée à la population d'assurés du portefeuille.

Origine sociale des études de nationalité française en 2019-2020, en %	Agriculteurs, artisans, commerçants et chefs d'entreprise	Cadres et professions intellectuelles supérieures	Professions intermédiaires	Employés	Ouvriers	Retraités et inactifs	TOTAL
Universités	9,1%	33,2%	15,1%	17,3%	10,9%	14,5%	100,0%
Formations d'ingénieurs hors université	12,1%	54,4%	11,9%	9,1%	5,4%	7,2%	100,0%
STS	12,2%	15,3%	14,6%	19,7%	23,0%	15,2%	100,0%
CPGE et préparations intégrées	10,8%	51,9%	12,6%	11,0%	7,1%	6,6%	100,0%
Ecoles de commerce, gestion et comptabilité	19,2%	49,9%	9,6%	9,4%	4,5%	7,3%	100,0%
Etablissement privés d'enseignement universitaire	15,3%	45,9%	11,7%	11,3%	6,8%	9,0%	100,0%
Ecoles normales supérieures (ENS)	7,1%	64,2%	9,9%	6,8%	2,3%	9,7%	100,0%
Ecoles artistiques, d'architecture et de journalisme	7,1%	64,2%	9,9%	6,8%	2,3%	9,7%	100,0%
Ecoles paramédicales et sociales	15,2%	23,7%	9,4%	31,1%	18,5%	2,0%	100,0%
Autres écoles et formations	12,6%	46,4%	11,1%	11,5%	6,5%	12,0%	100,0%
Ensemble des étudiants français	10,9%	34,4%	14,0%	16,8%	11,5%	12,5%	100,0%

Figure 6. Origine sociale des étudiants de nationalité française en France métropolitaine et DOM en 2019-2020 (en %)

(Source : <https://www.education.gouv.fr/media/70905/download>)

Le graphique de la *figure 7* présente l'évolution du taux de scolarisation par âge pour un intervalle allant de 12 à 29 ans sur les années 2010, 2015 et 2018. Le choix de l'intervalle est motivé par les taux de scolarisation assez proches et élevés pour les âges inférieurs à 12 ans. La notion du taux de scolarisation décrit la proportion d'élèves, d'étudiants et d'apprentis inscrits en formation initiale pour l'âge donné parmi l'ensemble des personnes de cet âge.

Les courbes des années 2010 et 2015 sont assez proches jusqu'à l'âge de 19 ans. Pour la tranche d'âge de 19 à 26 ans, la courbe de l'année 2015 se situe au-dessus de celle du 2010, ce qui indique que le taux de scolarisation de cet intervalle d'âge est supérieur en 2015 par rapport à 2010. En outre, la courbe de l'année 2018 se retrouve au-dessus des autres pour tous les âges, ce qui montre que la scolarisation est en évolution. Par ailleurs, contrairement aux années 2010 et 2015, le taux de scolarisation jusqu'à l'âge de 13 ans est de 100% en 2018, ce qui permet de supposer que l'obligation de scolarisation a un effet sur la population française. Le graphique montre ainsi que le taux de scolarisation est en croissance depuis l'année 2010, ce qui incite les assureurs à contrôler régulièrement leur approche en matière de valorisation de risque de maintien en études.

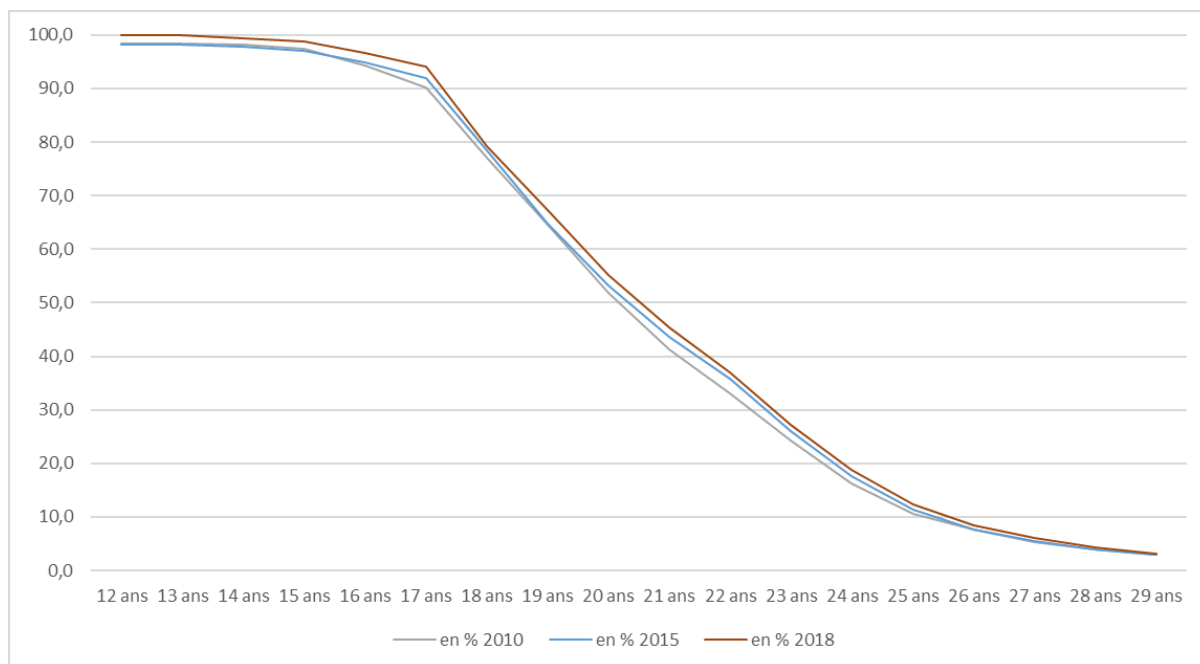


Figure 7. Evolution de taux de scolarisation par âge (12 - 29 ans) pour les années 2010, 2015, 2018 en France métropolitaine et DROM

(Source : <https://www.education.gouv.fr/reperes-et-references-statistiques-2020-1316>)

Le graphique de la *figure 8* permet d’observer l’évolution de l’espérance de scolarisation à l’âge de 2 ans dans la population française pour la période 1985 – 2015. Cette espérance est en croissance régulière entre les années 1985 et 1995. En revanche, elle baisse ensuite jusqu’à l’année 2008 pour se mettre de nouveau à augmenter après. Finalement la durée de scolarisation espérée à l’âge de 2 ans semble être relativement stable depuis l’année 1989 : la variation ne dépasse pas 1 an. Ce constat permet de supposer que les conditions de scolarisation n’ont pas évolué de manière significative pendant cette période. Il est donc acceptable de choisir un intervalle de temps assez large (de 30 ans) pour la constitution de la base de données pour la construction de la loi d’expérience. En 2015, la durée de scolarisation espérée pour un enfant de 2 ans est de 18,3 années dont 8,2 dans le cursus primaire ; 6,9 dans le secondaire ; 2,7 dans le supérieur et 0,5 en apprentissage.

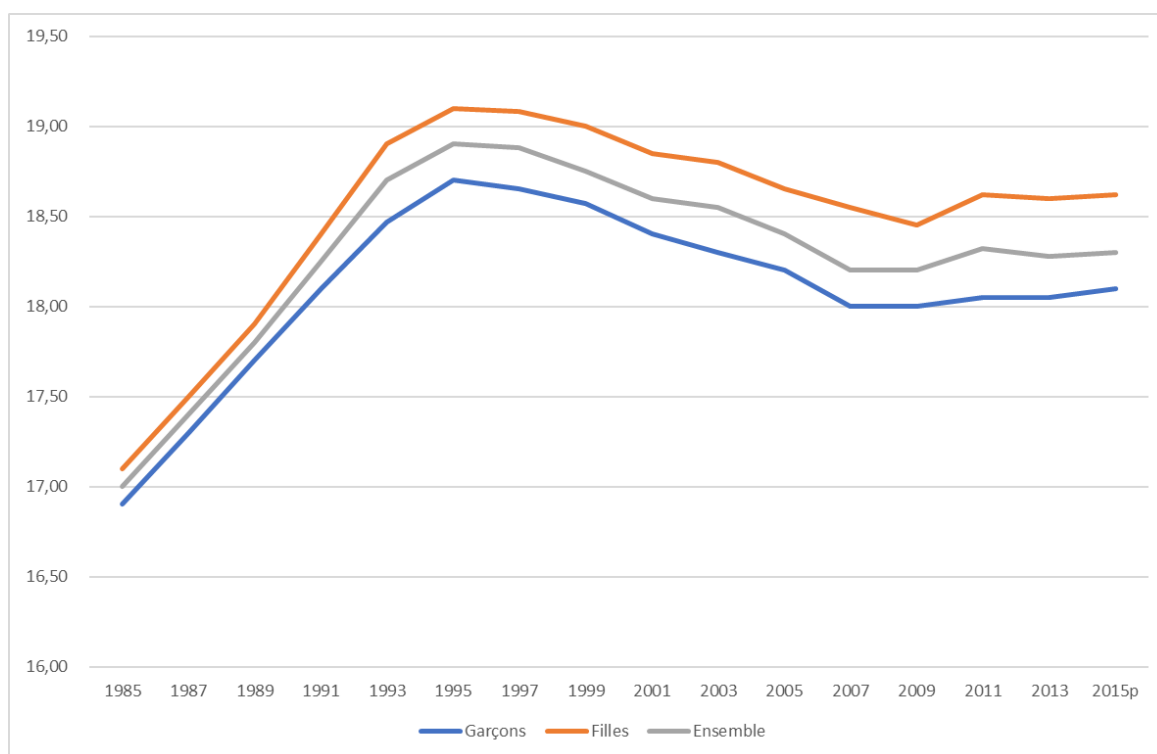


Figure 8. Evolution de l'espérance de scolarisation à l'âge de 2 ans selon le sexe en France métropolitaine et DROM depuis 1985

(Source : https://www.insee.fr/fr/statistiques/fichier/3526077/Formemp18f1_F1.1_ecole.pdf)

Le tableau de la *figure 9* nous informe quant au diplôme le plus élevé obtenu par les élèves issus d'une formation initiale. La part de diplômés supérieurs augmente de 2 points pour l'ensemble de la population en 2017 comparé à la moyenne annuelle 2013-2015. Cette hausse est portée par l'augmentation du nombre de titulaires d'un diplôme de niveau Master aussi bien chez les hommes (4 points de plus) que chez les femmes (1 point de plus). La part de diplômés de licence reste stable. De plus, la proportion des non diplômés chez les hommes et chez les femmes ainsi que la part de sortants du système éducatif avec seulement un diplôme du 2nd degré est en baisse. Ces chiffres permettent de conclure que la tendance de la hausse de la durée des études est observée dans la population française, ce qui a un impact direct sur le provisionnement de la garantie rente éducation.

Il est cependant important de préciser que l'obtention du diplôme le plus élevé n'est pas forcément un indicateur parfait pour faire une conclusion sur la durée moyenne des études. En effet, une personne peut très bien s'engager dans une formation assez longue après le baccalauréat sans forcément réussir à la valider et à obtenir le diplôme qui y est associé. De plus une même personne peut suivre plusieurs formations et obtenir plusieurs diplômes de même niveau, ce qui joue sur la durée de scolarisation mais n'est pas visible dans la *figure 9*.

Répartition des sortants de formation initiale en fonction de leur diplôme le plus	Moyenne annuelle 2013,2014 et 2015			2017		
	Ensemble	Homme	Femme	Ensemble	Homme	Femme
Diplôme supérieur	44,0%	40,0%	48,0%	46,0%	40,0%	51,0%
Dont Master ou doctorat	13,0%	9,0%	17,0%	15,0%	13,0%	18,0%
Dont Licence	10,0%	9,0%	10,0%	10,0%	9,0%	11,0%
Diplôme du 2nd degré	42,0%	44,0%	41,0%	41,0%	44,0%	38,0%
Diplôme national du brevet seul	7,0%	7,0%	6,0%	7,0%	8,0%	6,0%
Aucun diplôme	7,0%	8,0%	5,0%	6,0%	7,0%	4,0%

Figure 9. Répartition des sortants de formation initiale en fonction de leur diplôme le plus élevé en 2013,2014,2015 et 2017 en France métropolitaine et DROM

(Source : https://www.insee.fr/fr/statistiques/fichier/3526077/Formemp18f2_F1.2_ecole.pdf)

1.3.2 Construction de la loi de sortie de référence

Avant de procéder à la construction de la loi d'expérience basée sur les données issues du portefeuille de l'assureur, il est intéressant de générer une table de maintien de référence à partir des indicateurs nationaux. La comparaison de cette table avec la loi d'expérience d'AXA permettra de dire dans quelle mesure le comportement des bénéficiaires de la rente éducation de l'entreprise ressemble à celui de la population française. Il sera donc possible de conclure si les données nationales peuvent être utilisées pour les travaux de provisionnement de la société.

Selon les conditions de contrat de Rente éducation proposé par AXA, le versement de la rente éducation est effectif si le bénéficiaire est âgé de moins de 16 ans ou si son âge est inclus dans l'intervalle de 16 à 26 ans et il est reconnu en tant qu'inscrit au système d'éducation. La rente peut être clôturée dans deux situations :

- Le rentier cesse ses études et sort du système éducatif
- Le rentier décède

Ces deux évènements sont supposés être indépendants entre eux. La loi de maintien de référence est alors le croisement des indicateurs de scolarisation et de mortalité de la population française par âge. Soit RE – évènement appelé « versement de la rente éducation », D – évènement correspondant au décès du rentier, F – évènement correspondant à l'arrêt des études pour le rentier (ou à sa déscolarisation) pour un motif autre que le décès. Si les évènements D et F sont indépendants, alors la probabilité du versement de la rente se calcule comme suit :

$$\Pr(RE(n)) = \Pr(\bar{D}(n)) * \Pr(\bar{F}(n))$$

$$\Pr(RE(n)) = (1 - \Pr(D(n))) * (1 - \Pr(F(n)))$$

$$\Pr(E(n)) = (1 - q_d(n)) * (1 - q_f(n))$$

$$\Pr(E(n)) = 1 - q_f(n) - q_d(n) + q_f(n) * q_d(n)$$

$$\Pr(E(n)) = 1 - q_f(n) - q_d(n)$$

Pour construire la loi de référence fictive, il est nécessaire d'utiliser deux indicateurs : le taux de mortalité et le taux de déscolarisation. L'INSEE met à disposition des lecteurs les taux de scolarisation par âge. Le taux de déscolarisation ou de sortie du système éducatif pour un âge n peut ainsi être obtenu à l'aide de la formule suivante : $q_f(n) = 1 - q_s(n)$.

Il est important de noter que, selon les données de l'INSEE, le taux de scolarisation pour les âges de 0 et 1 an est égal à 0% et qu'il est de 12% pour l'âge de 2 ans. Si la formule ci-dessus était utilisée pour ces âges, les taux de déscolarisation seraient de 100% pour 0 et 1 an et de 88% pour 2 ans. En réalité ces taux s'expliquent non pas par des sorties du système éducatif mais plutôt par le fait que les enfants n'y sont pas encore entrés. Ces taux ne doivent pas être pris en compte tels qu'ils sont pour la construction de la loi de maintien fictive car ils donneraient l'impression que tous les rentiers sortent de la rente à un âge situé entre 0 et 1 an. En tenant compte de cette particularité et en se rappelant que la scolarisation n'est obligatoire qu'à partir de l'âge de 3 ans, il est décidé d'utiliser une formule spécifique pour la probabilité de maintien pour les âges 0, 1 et 2 ans :

$$\Pr(E(n)) = 1 - q_d(n)$$

Le graphique de la *figure 10* illustre la courbe de maintien de la rente éducation construite à l'aide des données de l'INSEE caractérisant l'intégralité de la population française. Selon la loi fictive de référence, il ne reste pratiquement plus d'individus entrant dans le champ d'application de la rente éducation vers l'âge de 24 ans (la probabilité de maintien en rente à 24 ans est de 0,22%). Etant donné que la garantie rente éducation proposée par AXA s'étale jusqu'à l'âge de 26 ans (ou 28 ans pour certains contrats), le fait de s'appuyer sur les données nationales créerait donc un risque de sous-provisionnement.

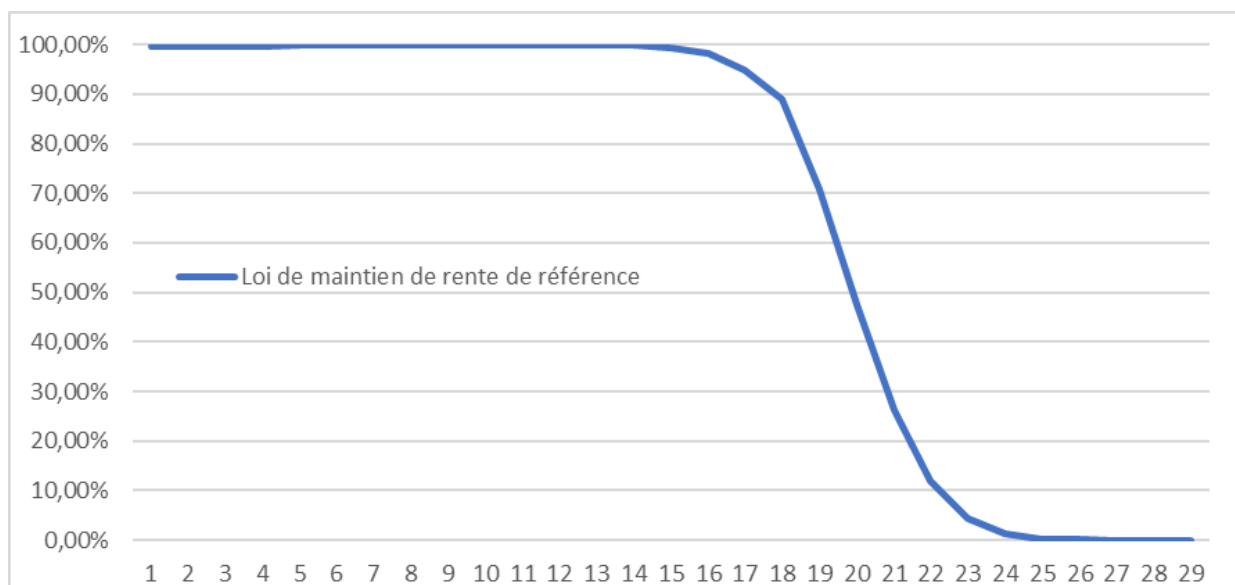


Figure 10. Loi fictive de maintien en rente éducation à partir des données d'INSEE

Chapitre 2. Construction de la base de données

L'étape de construction de la base de données est primordiale. Elle n'est pas forcément aisée car il s'agit généralement de bases contenant un nombre important d'observations. En outre beaucoup de variables sont disponibles à l'utilisateur et doivent être correctement interprétées et sélectionnées d'une manière qui permet d'éviter les informations erronées. Les données sur lesquelles le mémoire se base ont été saisies par des questionnaires dans le logiciel interne d'AXA France. Le logiciel est initialement mis en place pour la gestion de retraite mais a été étendu à l'intégralité des garanties. Cependant ce montage peut engendrer certaines complications informatiques pour le traitement des observations liées à la rente éducation. Ces points seront traités dans les paragraphes suivants.

2.1 Extraction des données

L'échantillon est construit à partir des données obtenues par la fusion à l'aide de programme WPS des trois tables suivantes : la table des mouvements des droits dus aux rentiers, la table des rentes référentielles contenant des informations personnelles sur les bénéficiaires et la table des informations générales. Uniquement les observations présentes simultanément dans les trois bases sont conservées pour la suite de l'étude.

Les données extraites pour le présent mémoire sont arrêtées à la survenance 2019. Le choix d'exclure l'année 2020 de l'échantillon pour la construction de la loi est justifié par la volonté d'avoir des données complètes sans observations manquantes. En effet les rentes éducation sont parfois déclarées tardivement et n'apparaissent dans le portefeuille que durant l'année qui suit l'évènement déclencheur de la garantie voir même plus tard.

La première étape consiste à isoler les observations qui concernent la rente éducation. Ensuite, la base de données est retraitée des contrats spécifiques incluant des rentes éducation viagères couvrant des enfants handicapés. En effet, une telle rente ne peut pas être provisionnée de manière identique à la rente éducation traditionnelle car il s'agit d'un engagement viager. La poursuite d'études n'étant pas une condition sine qua non à l'application de la rente éducation, il faut appliquer une approche spécifique pour calculer les provisions de ces contrats. Il serait judicieux, par exemple, d'utiliser une des tables réglementaires dans ce cas de figure.

Il est possible qu'un rentier ait plusieurs numéros de rente renseignés dans le système de gestion. Nombreuses sont les raisons pouvant expliquer ce phénomène. Premièrement, plusieurs rentes peuvent être ouvertes sous le même nom dans la situation où le bénéficiaire n'enverrait pas les pièces justificatives exigées pendant longtemps, ce qui justifierait la clôture de la rente. Il peut néanmoins très bien se remanifester par la suite, et dans ce cas-là, une nouvelle rente sera ouverte à son nom à la réception des documents manquants. Comme il s'agit du même individu, il est important de ne garder qu'une seule ligne par bénéficiaire. Ainsi c'est la ligne ayant la date d'effet de la rente la plus récente et la date de clôture de la rente la plus tardive qui est gardée pour chaque rentier. C'est une approche prudente qui ne génère pas de problème de sous-provisionnement.

La deuxième raison d'existence d'un même rentier avec de différents numéros de rente est la souscription de la garantie rente éducation par les deux parents. Dans ce cas, seule l'observation avec la date d'effet la plus récente est gardée pour ne pas créer un biais d'estimation à cause de la hausse artificielle du poids attribué à un seul individu.

Après ces traitements, la base de données de 16.591 observations est construite.

Elle contient les variables suivantes :

- Numéro de la rente
- Numéro du contrat (IDCORP)
- Titre du bénéficiaire (LTPAPP)
- Nom du bénéficiaire (LNPATI)
- Prénom du bénéficiaire (LPPATI)
- Sexe du bénéficiaire
- Motif de clôture de la rente
- Nature de la rente (temporelle ou viagère) (CDRENA)
- Date de naissance du bénéficiaire de la rente (DTNAIS)
- Date de décès de l'assuré (DLPAPP)
- Date de fin théorique de la rente (DTFINT)
- Date de clôture de la rente
- Date d'effet de la rente
- Code de catégorie de prestation d'assurance (CCRRCP)
- Code de catégorie de personnel (CCCORP)

2.2 Retraitement des données

Une fois la base de données extraites à l'aide d'un code WPS, le traitement de cette dernière se fait via Excel.

Suppression des rentes viagères

Les observations de la base obtenue ne concernent pas seulement les rentes temporaires mais aussi les rentes viagères. En effet, si l'enfant bénéficiaire est reconnu sous le statut handicapé, la garantie rente éducation prévoit des versements viagers. L'objectif du présent mémoire étant d'estimer la durée de la rente et le moment de sa clôture, les rentes viagères n'apportent pas d'informations nécessaires pour la modélisation. De plus, la population concernée par la rente viagère risque d'avoir un comportement statistiquement très différent par rapport à celle de référence. De ce fait, les rentes viagères doivent être retirées de la base de données. Il reste 16.330 observations dans l'échantillon après ce traitement.

Suppression d'anciennes rentes

Il est décidé par la suite de supprimer de la base de données les rentes jugées trop vieilles, notamment celles dont la date d'effet est antérieure à 1990. En effet, il est judicieux de penser que ces rentes n'apportent pas d'informations suffisamment fiables et ceci pour deux raisons. Premièrement, plus une rente est ancienne, plus le risque de tomber sur des données incomplètes ou erronées est élevé. Deuxièmement, il est préférable de ne pas conserver les contrats trop anciens si l'objectif de l'étude est d'actualiser la loi car sur une période si longue le comportement des rentiers est susceptible de changer de manière significative.

La *figure 8* du chapitre précédent montre que l'espérance de scolarisation à 2 ans varie dans l'intervalle inférieur à 1 an depuis l'année 1989. Cette observation permet de supposer que le comportement des individus en fonction de leur durée de scolarisation n'a pas évolué de manière significative durant cette période. Il est donc acceptable de garder les rentes qui prennent effet depuis 1990 sans biaiser le résultat de calcul. L'application de ce filtre permet d'obtenir l'échantillon de 14.344 observations.

Suppression des rentes sans effet

Par la suite la base de données est dépolluée des observations n'apportant pas d'informations quant à la durée de la rente des bénéficiaires. Il s'agit notamment des contrats pour lesquels la date d'effet de la rente est égale à celle de sa clôture. Ces affaires correspondent à la situation où il ne s'agit pas du versement de la rente mais plutôt du versement du capital décès en une seule fois. Ainsi 258 observations supplémentaires sont supprimées de la base.

De plus, la base de données contient 103 observations pour lesquelles la date d'effet de la rente est supérieure à celle de sa clôture et le motif de clôture n'est pas renseigné. En outre, toutes les rentes en question sont ouvertes depuis moins d'un an, ce qui permet de supposer, après échange avec l'équipe de Gestion, qu'il s'agisse d'une erreur de saisie qui sera corrigée plus tard. Ces observations doivent être supprimées de l'échantillon car elles n'apportent pas d'information sur la durée de la rente. Après ce traitement il reste 13.983 lignes dans la base.

Retraitement des motifs de clôture

Pour la suite de l'étude les données obtenues sont ventilées par motif de clôture. Le tableau de la *figure 11* reprend tous les motifs de clôture recensés et le nombre d'observations pour chaque motif. Pour la suite de l'étude tous les motifs doivent être interprétés en tenant compte de la spécificité de la garantie et de l'outil de gestion. Une nouvelle liste de motifs sera construite par la suite.

Abréviation	Motif de clôture de rente	Nombre d'observations avant le retraitement
-	En cours	4 234
DC	Décès du bénéficiaire	27
ES	Erreur de saisie	16
FS	Fin de scolarité	1 938
FT	Fin théorique de la rente	3 044
JA	Justificatif absent	3 697
MI	Migration informatique	374
OP	Changement d'options	22
RS	Résiliation de la société	108
RT	Rente temporaire	493
SE	Sans effet	30
Total général		13 983

Figure 11. Rentes éducation d'AXA France ventilées par le motif de clôture

Si le motif de clôture n'est pas renseigné alors il s'agit d'une rente qui est toujours en cours (4.234 rentes). Une nouvelle modalité *EC* (« *en cours* ») est créée pour regrouper les rentes non clôturées.

Les rentes clôturées pour le motif « *Fin théorique* » mais dont la date de clôture est postérieure au 31/12/2019 sont aussi considérées comme des observations en cours à la date de fin de la période d'observation (soit 156 rentes).

Le motif de clôture « *Fin de scolarité* » est synonyme de la sortie du bénéficiaire du portefeuille, ce qui signifie qu'il a quitté le système éducatif. La date de fin d'observation est normalement fixée au 31/12/2019. Cependant, plusieurs rentes clôturées pour motif « *fin de scolarité* » ont une date de clôture postérieure à la date d'extraction, de quelques jours, mois ou même quelques années (24 observations). Il s'agit des rentes clôturées par anticipation qui interviennent lorsque le bénéficiaire déclare en avance la date d'arrêt de son parcours. La présence de ces lignes est alors justifiée mais risque d'induire un biais puisque les autres motifs de clôture de la rente ne sont pas représentés après le 31/12/2019. Dans ce cas-là le motif de clôture pour ces observations-là devrait être modifié : ces rentes sont supposées être toujours en cours au 31/12/2019.

Les rentes clôturées pour le motif « *erreurs de saisie* » (*ES*, soit 16 observations) ne sont pas exploitables par définition et sont de ce fait supprimées de la base.

Le motif *MI*, soit « *migration* », concerne les rentes en provenance d'autres assureurs gestionnaires (dans la situation de la reprise de rente, par exemple). Ce cas peut également exister si la rente est reprise ou transmise à la concurrence. Ainsi parmi les 374 rentes clôturées suite aux migrations, 264 ont pris effet à une même date (01/09/1999) et ont toutes été clôturées 15 mois plus tard (31/12/2000). De plus ces rentes concernent la même raison sociale. D'après les retours des équipes de Gestion, il s'agirait des rentes reprises à la

concurrence puis rendues 15 mois plus tard. La durée de présence des rentes dans le portefeuille étant très courte donc, il a été décidé de supprimer ces observations de la base. Les autres rentes clôturées pour motif de migration sont également de courte durée (inférieure à 15 mois) et sont donc retirées de l'échantillon.

Le motif de clôture de la rente pour changement d'option est normalement destiné à la gestion des retraites. Ces 22 observations sont alors exclues de l'étude.

La base contient 493 observations dont le motif de clôture est « *rente temporaire* ». Comme il a déjà été indiqué auparavant, le logiciel a été initialement mis en place pour la gestion des retraites. En effet, dans le cadre d'un contrat de retraite, la rente peut être temporaire ou viagère. En revanche, la rente éducation peut être viagère uniquement dans le cas du bénéficiaire handicapé. Ainsi le motif de clôture « *rente temporaire* » est équivalent à une clôture pour le motif de « *fin théorique de rente* ». Il a été décidé alors de remplacer le motif *RT* par le motif *FT* et de garder ces 493 observations pour la suite de l'étude.

Enfin la base contient 30 lignes dont le motif de clôture est renseigné comme « *sans effet* ». Le poids de ces observations reste marginal. De ce fait elles peuvent être supprimées de la base sans que cela ait impact sur la suite de l'étude.

Le tableau de la *figure 12* donne la ventilation des rentes par motif de clôture après retraitement. La base obtenue contient désormais 13.541 observations.

Abréviation	Motif de clôture de rente	Nombre d'observations après le retraitement
DC	Décès du bénéficiaire	27
EC	En cours	4414
FS	Fin de scolarité	1914
FT	Fin théorique de la rente	3381
JA	Justificatif absent	3697
RS	Résiliation de la société	108
Total général		13541

Figure 12. Rentes éducation d'AXA France ventilées par le nouveau motif de clôture

Traitement des rentiers avec un âge atypique

Quelques nouvelles variables sont introduites pour la suite de l'étude :

- Age_effet : âge du bénéficiaire à la date d'effet de la rente
- Age_clot : âge du bénéficiaire à la date de clôture de la rente
- Age_theorique : âge du bénéficiaire à la date de fin théorique de la rente
- Age_2019 : âge du bénéficiaire de la rente au 31/12/2019

La garantie rente éducation temporaire couvre le bénéficiaire au plus tard jusqu'à ses 28 ans. De ce fait, toutes les observations pour lesquelles le bénéficiaire a strictement plus de

28 ans à la date de clôture de la rente ou à la date du 31/12/2019 si la rente est toujours en cours, sont supprimées de la base. Cela concerne 66 rentiers.

La base de données contient quelques observations avec un âge négatif au moment d'effet de la rente. Cette situation peut arriver dans le cas où l'assuré décède avant la naissance de l'enfant. Dans ce cas-là, la date d'effet de la rente ne peut pas être antérieure à 9 mois avant la date de naissance du bénéficiaire. Pour une observation, la rente prend effet 5 ans avant la naissance du rentier et la rente est clôturée un an avant sa naissance. Il s'agit à priori d'une erreur de saisie. Cette observation est donc supprimée de la base.

Pour les autres observations dans ce cas de figure, l'écart de date est inférieur à 9 mois, donc il ne s'agit pas de l'erreur de saisie. En revanche, la méthode de calcul de la rente, qui est utilisée par la suite, ne prend pas en compte de telles observations pour le calcul de l'exposition. De ce fait, il est décidé de remplacer la date d'effet de la rente par la date de naissance du bénéficiaire pour ces observations. Il s'agit de 37 lignes retraitées.

Traitement des sorties pour décès

Parmi les observations ayant pour motif de clôture « DC », 11 lignes ne contiennent pas de date de décès du bénéficiaire. En tenant compte des conditions générales de la garantie rente éducation, l'hypothèse retenue est que la date du décès de bénéficiaire est égale à celle de la clôture de la rente.

Dans 2 cas, la date de décès du bénéficiaire renseignée est postérieure à la date de clôture de la rente. Il s'agit d'une erreur de saisie et les lignes doivent être supprimées de la base.

Pour certaines rentes la date de décès est antérieure à la date de clôture. Cette situation est possible car la rente éducation peut être versée pendant un trimestre après la survenance de l'évènement ayant déclenché la clôture.

2.3 Statistiques descriptives des données

Les statistiques descriptives sont le meilleur moyen de se familiariser avec les données disponibles et d'avoir une première idée sur les tendances comportementales des rentiers du portefeuille d'AXA.

Premièrement, l'ordre de grandeur de la rente éducation versé aux bénéficiaires du portefeuille est étudié. Le tableau de la *figure 13* affiche les statistiques descriptives pour le montant de rente versé annuellement aux rentiers de la base de données.

Variable	Nombre d'observation	Moyenne	Médiane	Ecart-type	Maximum	Minimum
Montant annuel de rente (en €)	13 472	7 075	4 831	9 245	659 600	-

Figure 13. Statistiques descriptives pour le montant de la rente annuelle versé aux bénéficiaires

Deuxièmement, plusieurs variables de date sont disponibles dans la base de données mais il est difficile d'effectuer des statistiques descriptives là-dessus car une moyenne ou une médiane pour les dates n'est ni utile, ni directement interprétable. De ce fait, il est plus intéressant de travailler avec les variables d'âge ou de durée, qui sont facilement récupérables à partir des variables de date.

Le tableau de la *figure 14* reprend les principales statistiques descriptives pour les variables d'âge du bénéficiaire de rente à différents moments : à l'effet, à la clôture (si elle a eu lieu), au 31/12/2019 (si présent dans la base) et à la clôture théorique.

Variable	Nombre d'observations	Moyenne	Médiane	Ecart-type
age_effet	13 472	13,4	14,0	6,1
age_clot	9 081	22,5	23,0	3,2
age au 31/12/2019	4 388	16,7	17,3	5,3
age_theorique	13 472	25,5	26,0	2,2

Figure 14. Statistiques descriptives pour l'âge du bénéficiaire au moment d'effet, de clôture, de fin théorique de la rente et au 31/12/2019

En moyenne, la rente éducation prend effet quand le bénéficiaire a 13,4 ans et pour la moitié des bénéficiaires cela arrive après 14 ans. La clôture de la rente survient en moyenne à l'âge de 22,5 ans et après l'âge de 23 ans pour la moitié des rentiers. L'âge du bénéficiaire au moment de clôture théorique de la rente est plus élevé qu'au moment de la clôture réelle. Ce constat permet de dire qu'une partie de rentiers ne reçoivent pas la rente jusqu'à la fin théorique et sortent du portefeuille avant cette date. L'âge théorique moyen possède l'écart-type plus faible que les autres variables, ce qui montre que la partie importante de rentiers ont l'âge dans la même tranche au moment de fin théorique. Ceci est expliqué par la manière de fixer le moment de la fin théorique de rente qui est un élément contractuel. Par ailleurs, au 31/12/2019 le portefeuille d'AXA compte 4.388 rentiers en cours et leur âge moyen s'élève à 16,7 ans.

Le tableau de la *figure 15* est construit uniquement pour les rentes réellement clôturées avant la fin de la période d'observation. Selon le tableau, les rentiers d'AXA sortent du portefeuille en moyenne à l'âge de 22,5 ans. Pour la population masculine l'âge moyen de sortie est de 22,2 ans qui reste plus bas comparé à l'âge de sortie de la population féminine (22,8 ans).

Selon la *figure 4* construite à partir des données nationales, l'espérance de durée de scolarisation à l'âge de 15 ans est de 6,5 ans en 2017, ce qui revient à dire que les Français sortent des études en moyenne à l'âge de 21,5 ans. Cet âge est de 21,3 ans pour la population masculine et de 21,7 pour la population féminine. Il s'avère que l'âge de sortie des rentiers d'AXA est légèrement plus élevé que la moyenne nationale et spécifiquement pour la population féminine.

	Nombre d'observations	Moyenne	Médiane	Ecart-type
age_clot_ensemble	9 081	22,5	23,0	3,2
age_clot_garçon	4 723	22,2	22,5	3,3
age_clot_fille	4 358	22,8	23,4	3,1

Figure 15. Statistiques descriptives pour l'âge du bénéficiaire au moment de clôture de la rente

La répartition des contrats par âge de bénéficiaire est présentée par la suite. Dans cette partie, nous allons analyser l'âge moyen des rentiers au moment d'effet de la rente. Selon les données présentées sur la *figure 16*, la majorité des bénéficiaires (soit 61,7%) ont moins de 16 ans au moment d'effet de la rente. En revanche, seulement 15% d'effectifs bénéficient d'une rente après 20 ans. Ceci s'explique par le fait que lorsque le parent de l'enfant âgé de plus de 20 ans décède, ce dernier a probablement déjà fini ses études et n'a donc pas le droit au versement de la rente éducation. La croissance des effectifs rentrés en rente jusqu'aux 18 ans s'explique principalement par la hausse du taux de mortalité des parents accompagnée d'un faible nombre de sorties du système d'éducation des bénéficiaires.

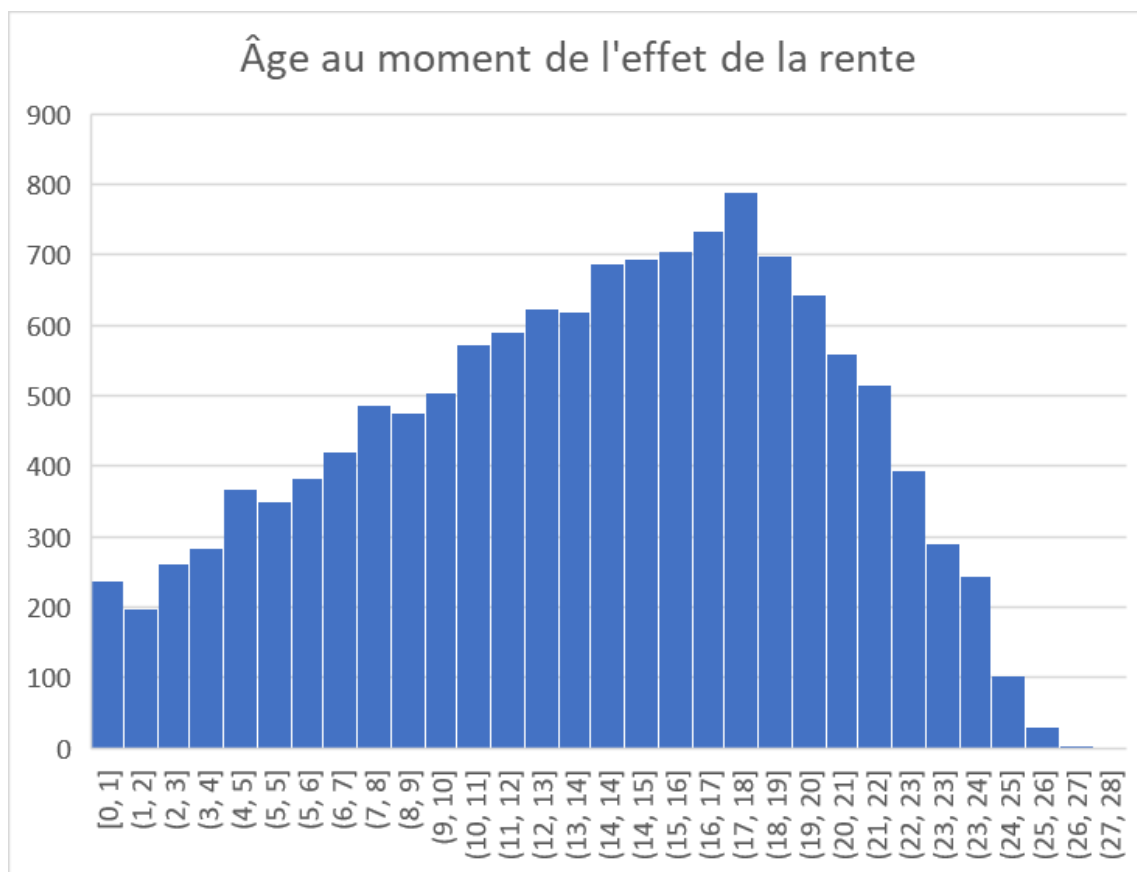


Figure 16. Ventilation des rentiers par âge au moment d'effet de la rente

L'histogramme de la *figure 17* affiche la répartition des rentiers par âge au moment de clôture de la rente. Cette variable n'est étudiée que pour les rentes clôturées avant 31/12/2019.

La part des rentes clôturées avant 15,6 ans du bénéficiaire est très faible, ce qui est plutôt logique dans le cadre législatif actuel imposant l'obligation de scolarisation jusqu'à l'âge de 16 ans. Le nombre de rentiers ayant entre 15,6 et 17,8 ans au moment de la clôture de la rente est plus important, mais reste faible. Ceci permet de supposer que la majorité des rentiers du portefeuille d'AXA arrivent à obtenir leur baccalauréat. La proportion des rentiers âgés entre 17,8 et 24,8 ans est relativement stable âge par âge. En revanche, un pic de clôtures de rentes est observé lorsque le rentier a entre 25 et 26 ans. Ceci permet de supposer qu'une partie importante de rentiers bénéficie de la garantie jusqu'à sa limite contractuelle (soit 26 ans dans la majorité des cas). Plus de 28% des rentiers ont plus de 25 ans au moment de clôture de leur rente, ce qui représente presque un tiers du portefeuille. Il s'agit donc d'une population qui a tendance à poursuivre des études longues.

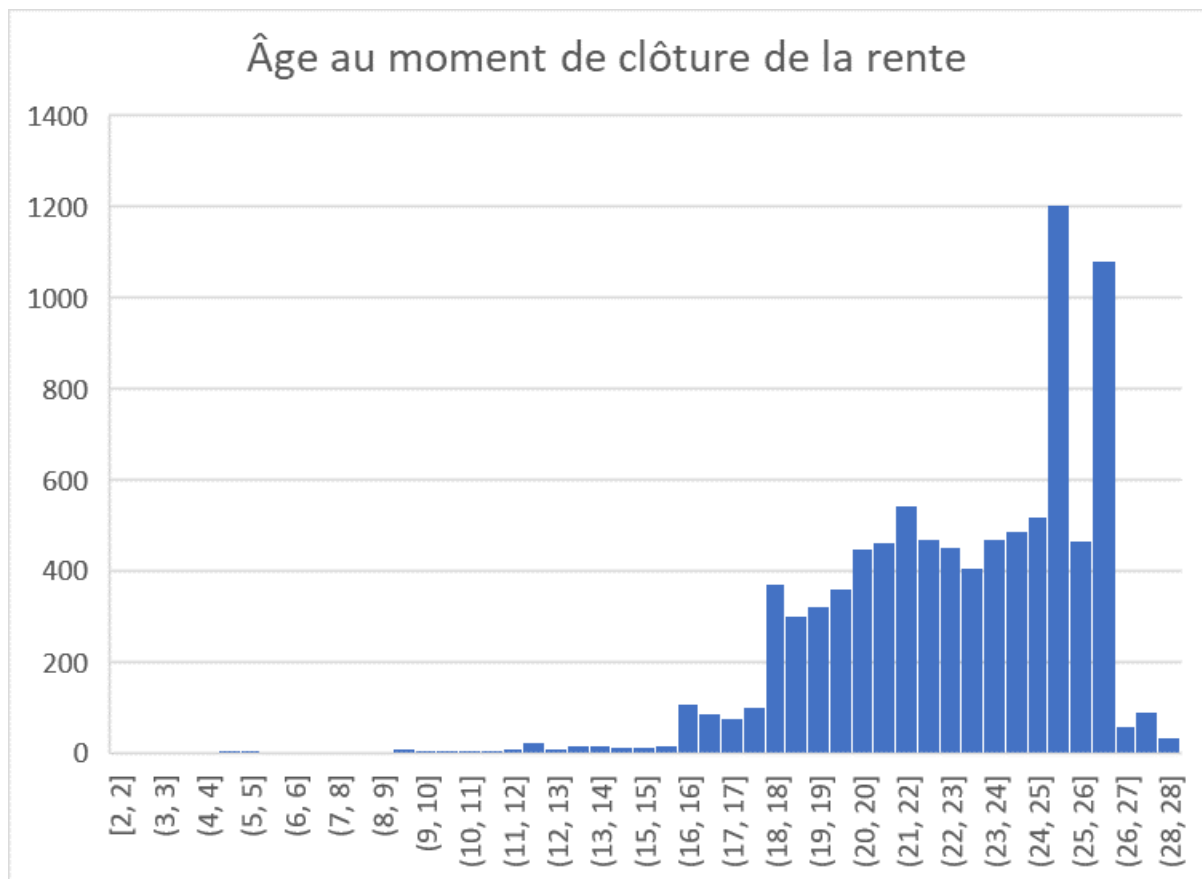


Figure 17. Ventilation des rentiers par âge au moment de clôture de la rente

L'histogramme de la figure 18 affiche l'âge au 31/12/2019 des rentiers dont les contrats sont en cours à cette date. Un peu plus que la moitié des effectifs correspondent à la tranche d'âge entre 15,7 et 23,4 ans, soit les âges pour lesquels la condition de maintien d'études est obligatoire pour le versement de la rente éducation.

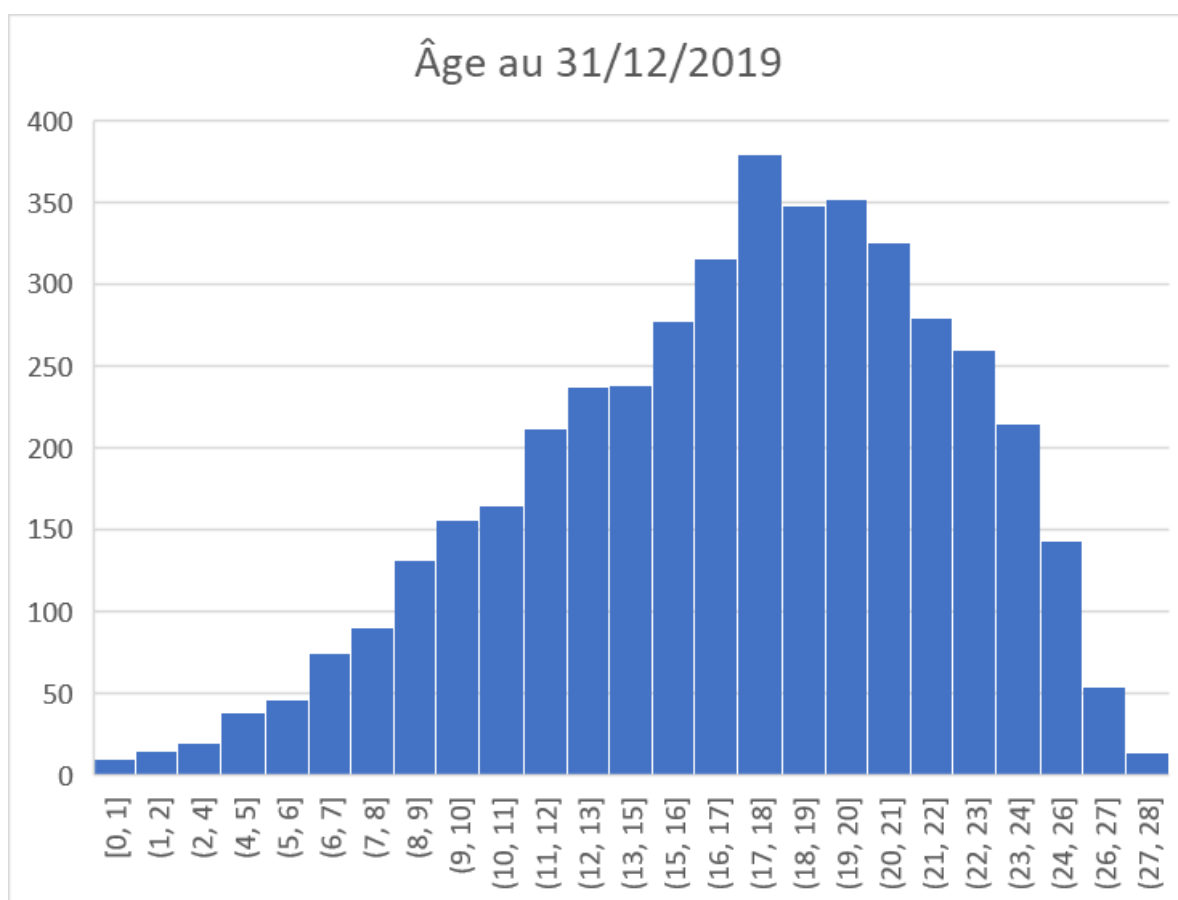


Figure 18. Ventilation des rentiers par âge au 31/12/2019

L'histogramme de la *figure 19* représente l'âge des rentiers au moment de fin théorique de la rente. La répartition des observations est très large. Par ailleurs, quelques dossiers avec l'âge allant au-delà de la limite contractuelle sont repérés. L'analyse à la maille de dossiers permet de conclure que dans ces cas-là la clôture réelle est finalement survenue au moment prévu par le terme des contrats (26 ans ou 28 ans sous certaines conditions). Il s'agirait alors des erreurs de saisi qui n'ont pas d'impact sur le calcul la durée réelle de la rente. Selon l'histogramme, la fin théorique de la majorité des rentes tombe sur une tranche d'âge comprise entre les 25 et 28 ans du bénéficiaire. Ceci est cohérent par rapport aux conditions générales de la garantie.

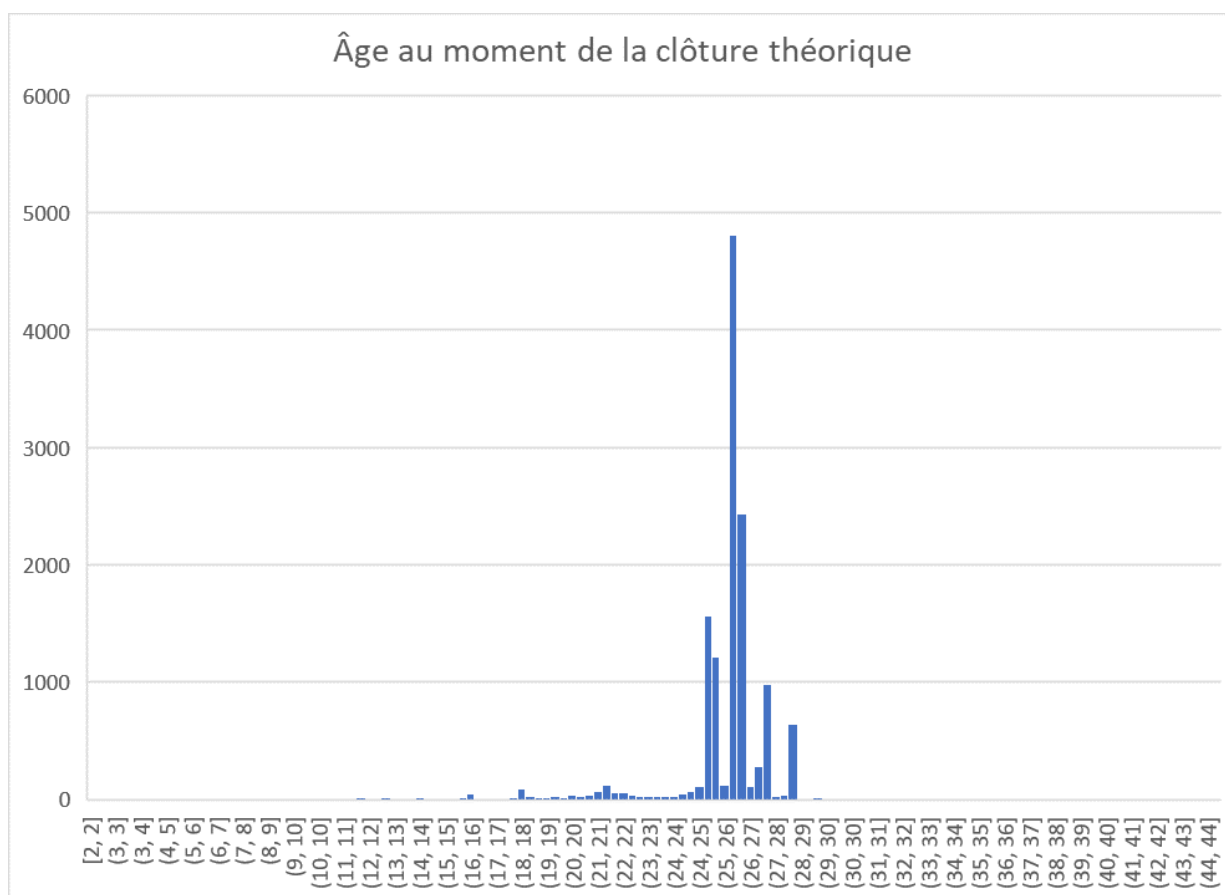


Figure 19. Ventilation des rentiers par âge au moment de la clôture théorique de la rente

Après avoir étudié les variables d'âge, il est aussi important d'analyser les variables de durée définies par les expressions suivantes :

- Durée réelle de la rente : $\text{durée_rente} = \text{âge du rentier au moment de clôture} - \text{âge du rentier à la date de prise d'effet de la rente}$
- Durée théorique de la rente : $\text{durée_théorique} = \text{âge du rentier au moment de fin théorique} - \text{âge du rentier à la date de prise d'effet de la rente}$
- Durée de la rente à la fin de la période d'observation = $\text{âge du rentier au 31/12/2019} - \text{âge du rentier à la date de prise d'effet de la rente}$

Selon le tableau de la *figure 20*, la durée théorique de la rente est plus importante que la durée réellement constatée ce qui est logiquement expliqué par le fait que certains rentiers ne bénéficient pas de leur garantie jusqu'à l'âge-limite. De plus, l'écart-type de la durée théorique de la rente est relativement important ce qui montre que les chiffres sont plutôt dispersés autour de la moyenne. Cette dispersion autour de moyenne est principalement portée par la dispersion de l'âge au moment d'effet qui est caractérisé par l'écart-type affiché dans la *figure 13*.

Selon le tableau de la *figure 20*, un rentier du portefeuille d'AXA bénéficie réellement de rente éducation pendant 7,5 années en moyenne. Et pour la moitié des rentiers cette durée est supérieure à 6,5 ans.

Variable	Nombre d'observations	Moyenne	Médiane	Ecart-type
durée_rente	9 081	7,5	6,5	5,2
durée_théorique	13 472	12,1	11,7	6,4
durée_2019	4 388	6,5	5,3	5,0

Figure 20. Statistiques descriptives pour la durée réelle de rente, la durée théorique de rente et la durée effective de rente au 31/12/2019

Les histogrammes décrivant les durées de rentes du portefeuille sont présentés dans les figures 21, 22, 23. La durée réelle de la rente affichée dans la figure 20, dépasse rarement 12 ans (seulement dans 19,3% de cas). En outre, de nombreuses rentes ont une durée assez courte (moins de 5 ans pour 39,6% de rentes). La durée théorique de rente affichée dans la figure 21 est très répartie et varie de 1 à 25 ans. Cette dispersion est surtout liée à l'écart-type assez élevé de la distribution de l'âge au moment d'effet de la rente. La figure 22 montre que quasiment la moitié des rentes en cours au 31/12/2019 sont d'une durée de moins de 5 ans à cette date précise (soit 47,7% d'effectifs), ce qui permet de supposer qu'elles vont encore être effectives pendant un certain temps.

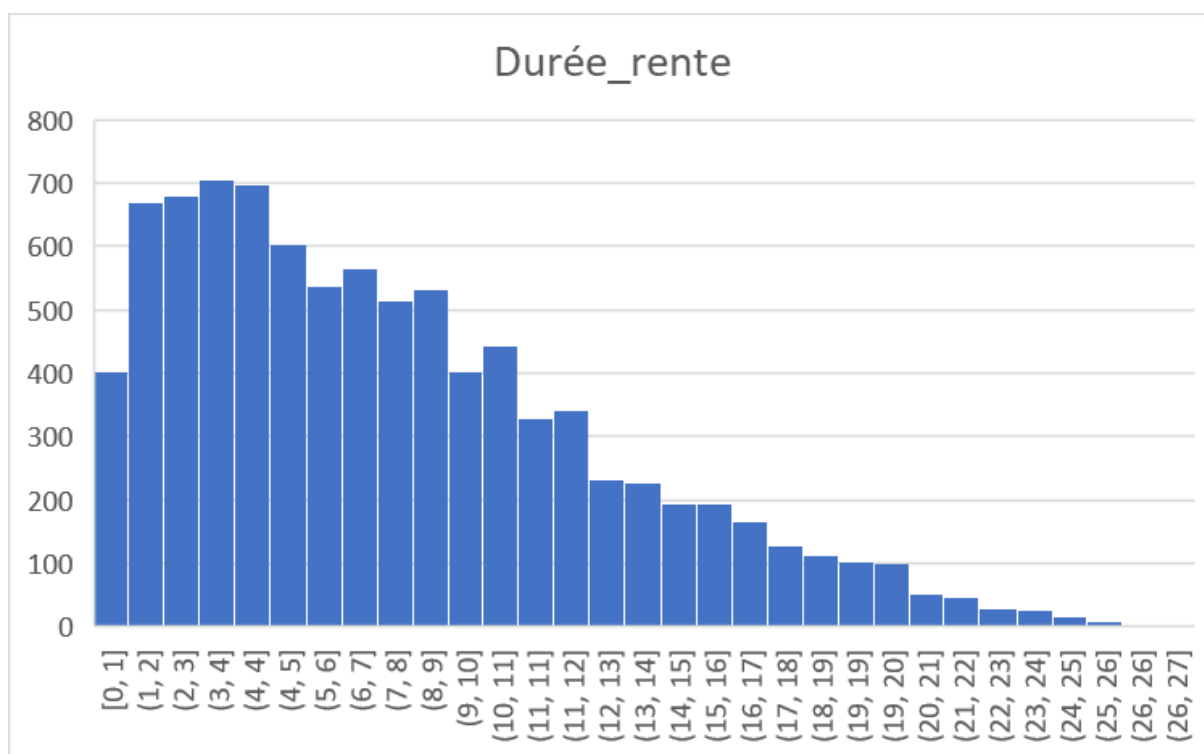


Figure 21. Histogramme pour la durée réelle de la rente

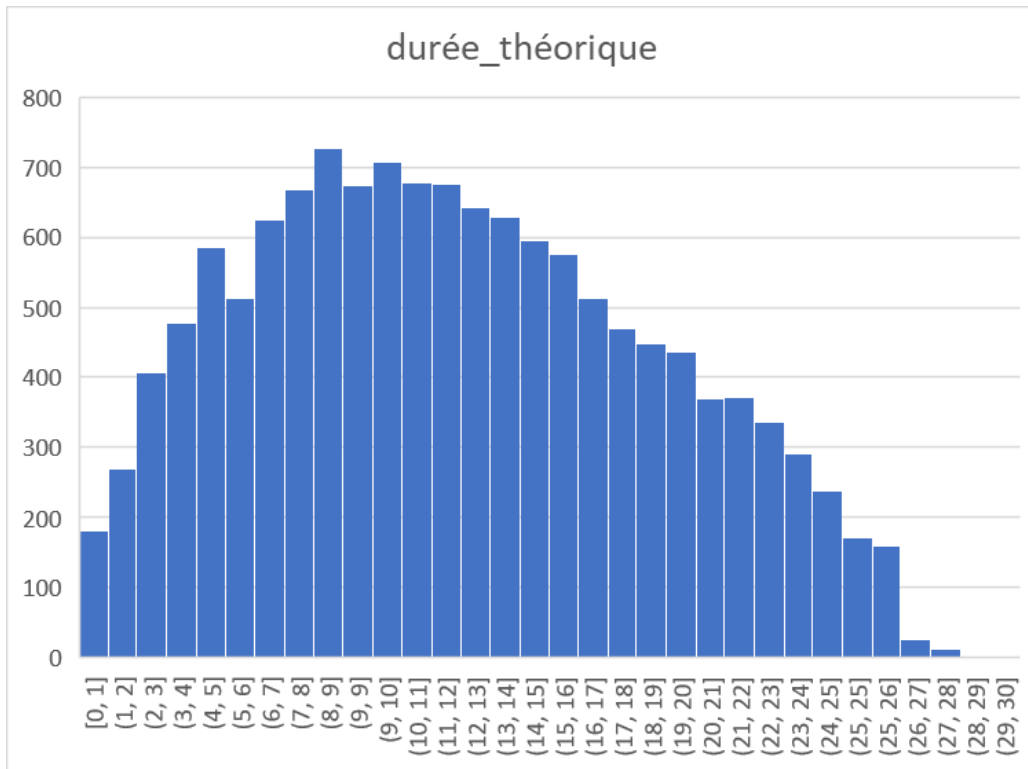


Figure 22. Histogramme pour la durée théorique de la rente

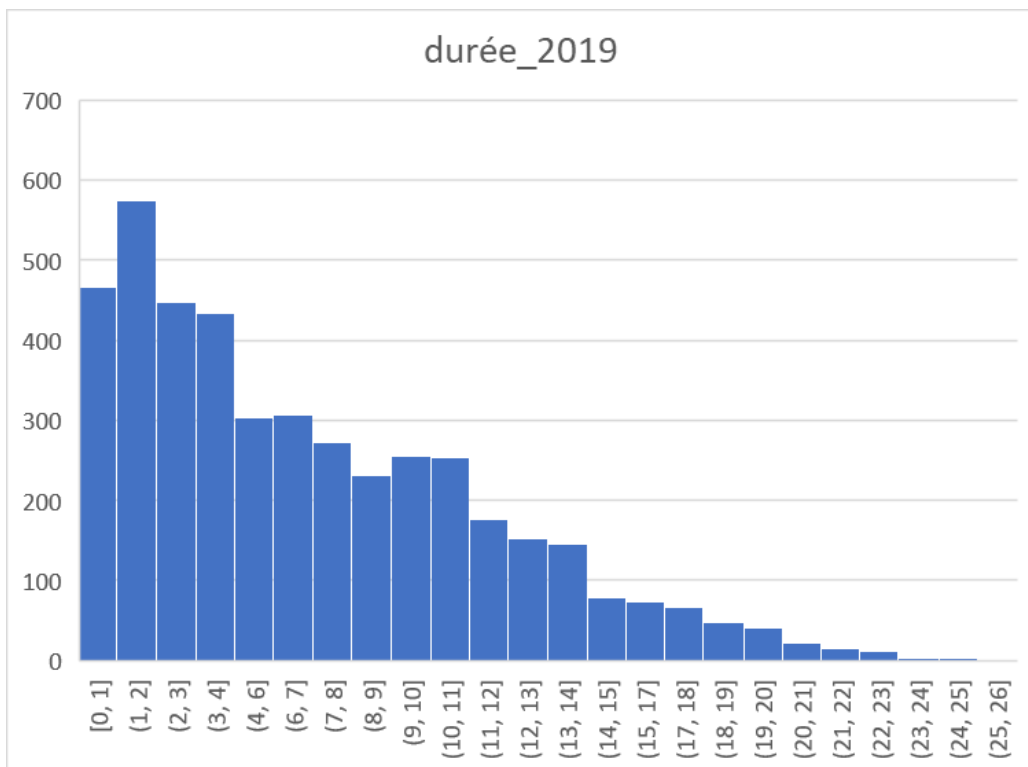


Figure 23. Histogramme pour la durée effective de la rente au 31/12/2019

Par la suite, la répartition des rentiers par sexe est étudiée. Selon le tableau de la *figure 24*, le nombre de rentiers hommes et femmes est quasiment identique dans le portefeuille d'AXA. Les hommes sont légèrement plus présents que les femmes.

Sexe du rentier	Nombre de rentiers	Poids
Homme	6861	50,9%
Femme	6611	49,1%
Total général	13472	

Figure 24. Répartition des rentiers par sexe

L'analyse de la répartition des contrats de rente éducation par motif de clôture de la rente est ensuite présentée. Le graphique de la *figure 25* montre que la majorité des rentes de l'échantillon sont toujours en cours. Concernant les rentes clôturées, le motif le plus fréquent est l'absence de justificatifs. Dans le cas général l'absence de justificatif est équivalente à une fin de scolarisation. En effet, si une rente est clôturée pour cause de documents manquants mais que le bénéficiaire poursuit ses études et présente le justificatif après la clôture de la rente, un nouveau numéro de rente sera créé et les versements vont être remis en place. Il est alors judicieux de supposer que la rente éducation est clôturée dans la majorité des cas pour cause de fin de scolarisation. Le deuxième motif le plus important est la fin théorique de la rente. Il s'agit des rentiers qui poursuivent leurs études jusqu'à l'âge-limite de la garantie rente éducation.

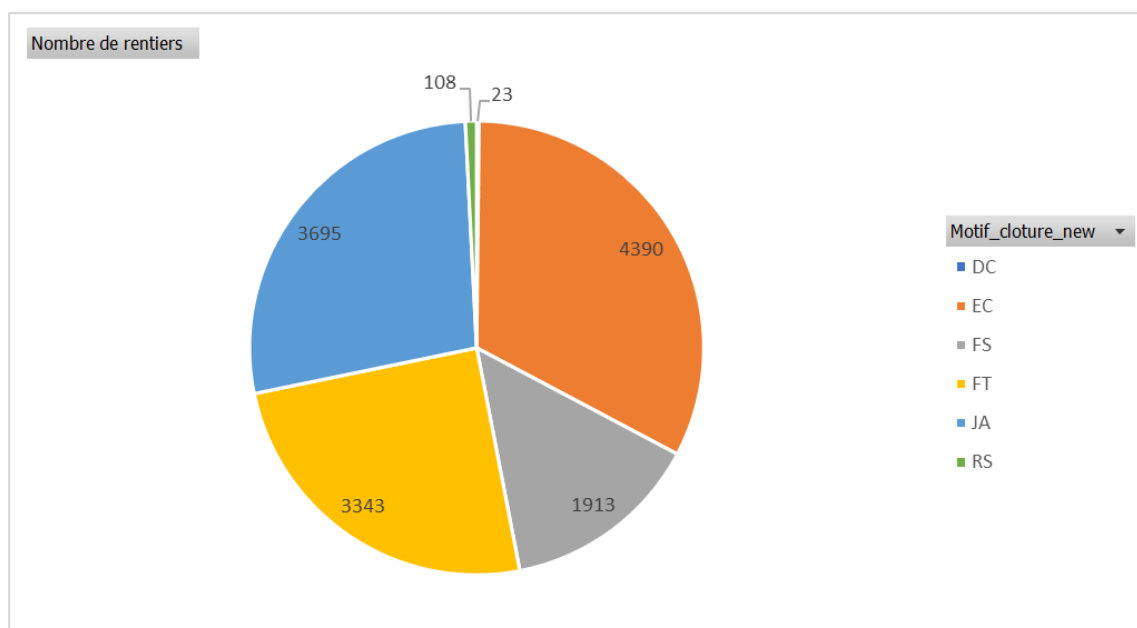


Figure 25. Répartition des bénéficiaires de la rente éducation en fonction du motif de clôture

Il est également important de comprendre si le motif de clôture de la rente a une certaine corrélation avec le sexe du rentier. Selon le graphique de la *figure 26*, les rentes des hommes et des femmes sont clôturées dans la majorité des cas pour cause d'absence de justificatifs. Dans l'échantillon des deux groupes de bénéficiaires la majorité des rentes restent toujours en cours au 31/12/2019. En outre, les rentes de la population féminine sont plus présentes dans celles en cours et moins présentes dans celles clôturées pour tous les motifs. Même si le nombre de rentes résiliées pour le décès du bénéficiaire reste très faible, la

population masculine présente un poids plus important dans cette catégorie (plus de 60% du total).

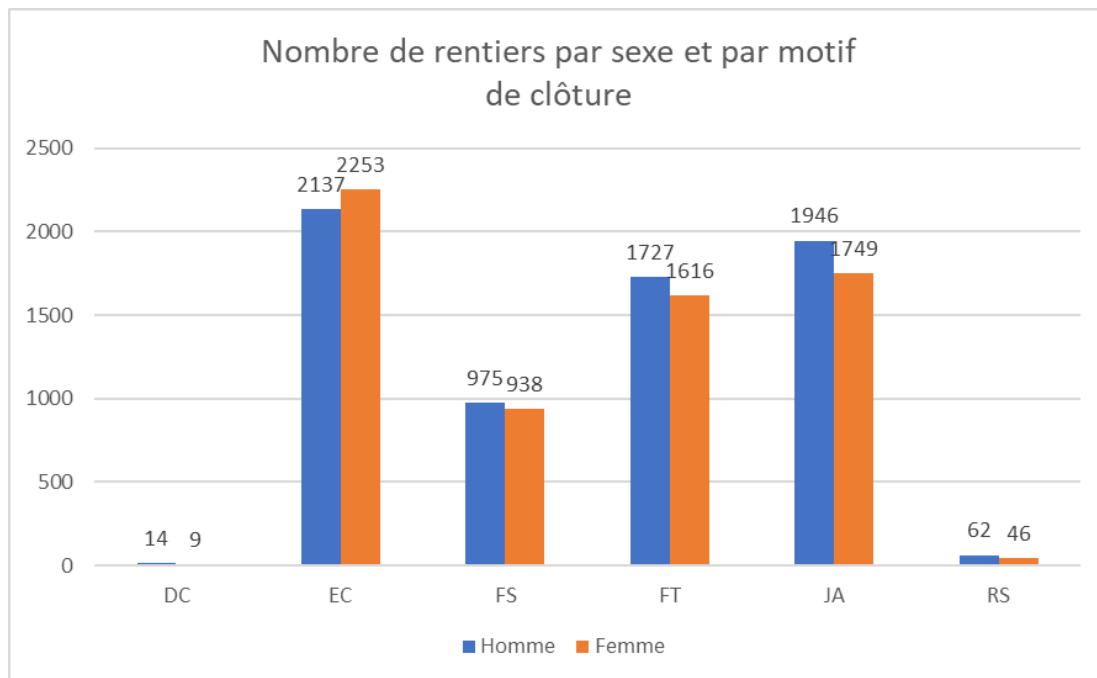


Figure 26. Répartition des bénéficiaires de la rente éducation en fonction du sexe et du motif de clôture

Chapitre 3. Construction de la loi de maintien d'expérience

3.1 Principales étapes du calcul de la loi de maintien

Après avoir construit une base de données fiable, il convient de procéder à la construction de la table de maintien. Etant donné que la taille de l'échantillon est relativement faible, les taux de sortie sont estimés uniquement par âge, sans distinction par année de survénance. En effet, un nombre limité d'observations ne permettrait pas d'obtenir les taux par année suffisamment fiables.

Plusieurs méthodes de modélisation des taux bruts peuvent être utilisées : paramétriques, semi-paramétriques et non paramétriques. Le choix de méthode est une étape très importante pour la suite de l'étude. Dans un premier temps, les méthodes non paramétriques sont utilisées pour l'estimation car elles permettent de ne pas faire d'hypothèses à priori à propos de la forme de la loi de maintien. Il faut donc chercher à estimer cette fonction directement à partir des données à disposition. L'application de l'approche non paramétrique inclut plusieurs étapes : calcul de l'exposition au risque, calcul des taux de sortie, lissage des taux obtenus, contrôle de la fiabilité des taux obtenus. Dans un second temps, une approche semi-paramétrique est utilisée pour la construction de la loi de maintien. Cette dernière permet d'étudier l'impact de certaines variables discriminantes sur la durée de la rente.

3.1.1 Définitions et notations

La modélisation de la loi d'expérience nécessite d'introduire quelques indicateurs dont la définition est présentée ci-dessous.

- ✓ Exposition au risque de sortie de la rente – E_n .

L'exposition au risque de sortie à l'âge n est un effectif présent dans le portefeuille à l'âge n . Autrement dit, il faut recenser des bénéficiaires de la rente éducation, qui sont susceptibles de sortir du portefeuille durant la période étudiée. La notion d'exposition au risque est liée à celles de censure et de troncature.

- ✓ Censure et troncature

Tous les rentiers ne sont pas forcément présents dans le portefeuille durant toute l'année calendaire de leur âge n . En effet, si un rentier sort du portefeuille pour une raison quelconque à l'âge n , la sortie ne survient pas obligatoirement le jour de son anniversaire, mais peut avoir lieu à n'importe quel moment de cette année. Il s'agit alors de la présence des données incomplètes, c'est-à-dire des données censurées ou tronquées.

La notion de censure intervient quand une observation sort du champ d'étude pour un motif différent de celui modélisé. L'information qui existe sur cette observation, c'est que l'évènement étudié n'a pas eu lieu pour cet individu avant un certain âge (celui qu'il a au moment de sortie du champ de l'étude). La date réelle de l'occurrence de l'évènement modélisé n'est donc pas connue. Soit (T_1, T_2, \dots, T_k) l'ensemble des durées des rentes

éducation du portefeuille et $C_i > 0$ le seuil de censure pour le rentier i . Ainsi l'ensemble de variables (T_1, T_2, \dots, T_k) n'est pas directement observable. Une autre variable, notée X_i vient pour la remplacer :

$$X_i = \min (T_i, C_i)$$

Si la rente du rentier i est clôturée avant la fin de la période d'observation – alors la durée exacte de cette rente est connue. S'il ne sort pas avant la fin de la période d'observation – sa rente reste en cours au moins jusqu'à cette date-là et la durée exacte de sa rente n'est pas connue.

La notion de troncature à gauche est utilisée quand la variable étudiée n'est observable qu'à partir d'un certain seuil, noté c_i . La différence avec la censure est telle qu'aucune information sur l'observation n'est disponible avant son entrée dans le champ d'observation. La variable observée de durée devient donc $T_i | T_i > c_i$.

Il faut également rajouter les notions de censure et de troncature aléatoire ou non aléatoire. La censure (ou la troncature) est non aléatoire s'il est possible de l'anticiper. Par exemple, dans le cadre de la table de mortalité, il s'agit de la censure aléatoire, car l'évènement étudié est le décès et ce dernier n'est pas prévisible.

Plusieurs cas de figures possibles peuvent avoir lieu dans le cadre de l'étude du portefeuille de rente éducation :

- Un rentier i est sorti du périmètre étudié durant la période d'observation pour le motif de migration → il s'agit de la censure à droite aléatoire
- La rente du rentier i prend effet avant le début de la période d'observation → il s'agit de la troncature à gauche non aléatoire
- Un rentier i sort du portefeuille à l'âge n après la fin de la période d'observation → il s'agit de la censure à droite non aléatoire
- ✓ Nombre de sorties de la rente éducation – N_n

Il s'agit simplement du nombre de rentiers sortis du portefeuille ou autrement dit du nombre de rentes clôturées avant la date de fin de période d'observation.

- ✓ Taux de sortie de la rente – q_n

Il s'agit du quotient du nombre de sorties survenues reporté à l'exposition au risque de sortie à l'âge n :

$$q_n = \frac{N_n}{E_n}$$

- ✓ Loi de maintien en rente – L_n

Une fois que les taux de sortie sont obtenus, la loi de maintien en rente peut être déduite : il suffit de calculer L_n qui représente le nombre de rentiers dont les rentes sont encore effectives à l'âge n du bénéficiaire :

$$L_n = (1 - q_{n-1}) * L_{n-1}$$

3.1.2 Limites de la méthode

Dans le cadre d'un portefeuille de données réelles, il n'est pas aussi évident de calculer l'exposition au risque à l'âge n qui est un âge discret.

Le premier problème qui peut être rencontré durant la modélisation est lié au caractère continu de la variable de durée de la rente. En effet, le nombre d'années n'est quasiment jamais un chiffre rond. Par exemple, une rente peut prendre effet à l'âge de 13 ans et 6 mois du bénéficiaire et être clôturée à l'âge de 19 ans et 4 mois du bénéficiaire. Le calcul de l'exposition n'est pas évident dans ce cas de figure. Cette problématique sera étudiée en détail dans le paragraphe 3.1.4.

Le deuxième point délicat de la modélisation de la loi de maintien d'expérience est lié aux données censurées. En effet, la grande partie de rentes du portefeuille d'AXA est encore en cours au 31/12/2019 (soit 32,6% d'observations). En outre, plusieurs rentes sont clôturées pour des motifs qui ne permettent pas d'émettre une conclusion ferme sur la durée réelle de la rente. Ces observations ne peuvent pas être comptées comme sorties de la rente mais ne doivent pas non plus être supprimées de l'échantillon car elles apportent une certaine information. Elles permettent notamment de dire qu'à un âge n , des rentiers ne sont toujours pas sortis du portefeuille, même si la date exacte de leur sortie n'est pas connue. Nous devons donc les compter dans l'exposition, car ils ont bien été observés et bénéficient d'une rente au moment de la fin de la période d'observation.

3.1.3 Détection des observations censurées

Pour la suite de l'étude, il faut identifier les rentiers dont la rente peut être considérée comme clôturée de manière définitive. Cette analyse est réalisée en s'appuyant sur les échanges avec l'équipe de Gestion. La variable *mot_clot_0* extraite de l'entrepôt de données peut prendre la modalité « *Fin de scolarité* » qui indique naturellement qu'il s'agit de la rente clôturée pour motif de fin d'études. Selon l'avis d'un expert technique, le motif de clôture « *Justificatif absent* » peut être considéré comme une fin de scolarisation car il est compliqué d'imaginer la situation où un rentier refuserait d'envoyer les documents justificatifs lui permettant de recevoir la rente s'il en dispose. Enfin, le motif de clôture « *Fin théorique* » peut aussi être interprété comme une sortie définitive de la rente pour les observations dont la date théorique de clôture est éloignée de plus de 3 mois (soit 92 jours) de la date réelle de clôture, parce que la majorité des rentes est versée trimestriellement. Dans le cas contraire, la clôture de la rente théorique est vue comme une censure car la raison réelle de sa clôture n'est pas correctement renseignée.

Il est aussi judicieux de penser que les rentes clôturées pour le motif « *Justificatif absent* » et dont la clôture ne date pas de plus de 6 mois, peuvent encore être rouvertes car le bénéficiaire peut mettre du temps à obtenir les documents nécessaires et à les transférer à l'assureur. De plus, le motif « *Justificatif absent* » pour un rentier de moins de 16 ans semble être incohérent par rapport à la réglementation actuelle. Il ne s'agit donc pas de clôture réelle mais plutôt d'un dossier en attente et donc d'une observation censurée.

Il faut classer toutes les observations de l'échantillon dans deux catégories : les rentiers dont la sortie de la rente est confirmée et définitive, puis ceux qui sont censurés. Si la rente est toujours en cours au moment de la fin de la période d'observation, alors il s'agit de la censure à droite aléatoire car le moment exact de clôture de la rente n'est pas connu. De plus, si un rentier sort du portefeuille avant la fin de période d'observation mais pour un motif ambigu ou erroné, il est préférable de parler de la censure. Une nouvelle variable *mot_clot* est alors créée et définie comme suit :

- ✓ Cens → sortie censurée :
 - Dont les rentes en cours au 31/12/2019
 - Dont les rentes clôturées pour « *Résiliation de la société* »
 - Dont les rentes clôturées pour « *Fin théorique* » si la date de clôture de la rente est éloignée de moins de 92 jours de sa date de fin théorique
 - Dont les rentes clôturées pour « *Justificatifs absents* » si le rentier a moins de 16 ans ou si sa rente est clôturée il y a moins de 6 mois par rapport à la date de la fin de période d'observation (soit moins de 183 jours)
- ✓ Sortie → clôture définitive de la rente éducation :
 - Dont les rentes clôturées pour « *Fin de scolarité* »
 - Dont les rentes clôturées pour le décès du bénéficiaire
 - Dont les rentes clôturées pour « *Fin théorique* », si la date de clôture de la rente est éloignée de plus de 91 jours de sa date de fin théorique
 - Dont les rentes clôturées pour « *Justificatifs absents* » si le rentier a plus de 16 ans ou si sa rente est clôturée il y a plus de 6 mois par rapport à la date de la fin de période d'observation (soit plus de 182 jours)

Le graphique de la *figure 27* affiche la répartition des observations de la base de données par le nouveau motif de clôture de la rente. La majorité des rentes censurées sont toujours en cours au 31/12/2019. Et parmi les rentes clôturées la majorité est interrompue pour la raison de l'absence de justificatifs. En outre, plus de la moitié des observations (7.959 rentiers, soit 59,1% de la base) sont censurées.

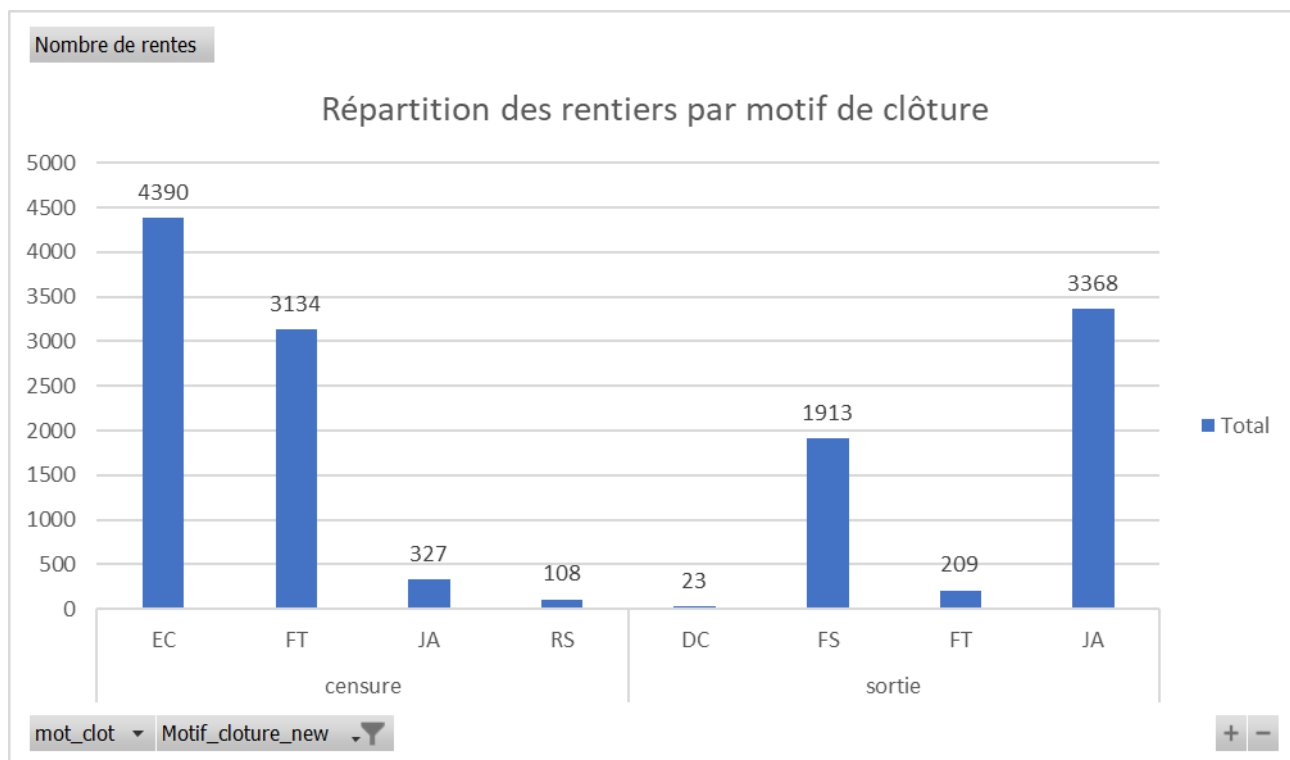


Figure 27. Nombre de sorties et de censures dans le portefeuille des rentiers

3.1.4 Calcul d'exposition au risque

Dans ce mémoire, deux méthodes de calcul de l'exposition au risque sont proposées. Le choix de méthode impacte directement les taux de sortie modélisés par la suite.

Première méthode

Pour la première méthode il est convenu de se placer successivement sur chaque année entre 1990 et 2019. Soit un rentier du portefeuille dont l'âge arrondi vaut n au 31/12/ X .

Plusieurs cas de figures sont possibles :

- si le bénéficiaire reste en rente durant toute l'année X , alors il comptera pour 1 dans l'exposition à l'âge n
- si le bénéficiaire sort du portefeuille pour un motif de censure durant l'année X , il comptera pour $\frac{1}{2}$ dans l'exposition de l'âge n
- si le bénéficiaire sort du portefeuille pour un motif bien défini durant l'année X , alors son exposition à l'âge n vaudra 1

Cette méthode a des limites. Premièrement, il est possible de se retrouver avec les taux de sortie supérieurs à 1. Deuxièmement, l'exposition calculée par cette méthode est approximative puisque lorsqu'une rente prend effet, il est supposé que cela arrive au milieu de l'année (autrement dit, l'exposition vaut $\frac{1}{2}$). En plus, un bénéficiaire contribue à la hauteur de $1\frac{1}{2}$ à l'exposition de son âge de sortie de la rente. Si cet individu sort de rente durant l'année X , alors il comptera pour 1 dans l'exposition à l'âge n pour l'année $(X-1)$ mais aussi

pour $\frac{1}{2}$ dans l'exposition de l'âge n pour l'année X . Etant donné que les expositions de toutes les années sont additionnées à la fin, l'exposition est surestimée et le taux de sortie d'éducation est systématiquement sous-estimé.

De plus, si un bénéficiaire sort du portefeuille durant l'année X ce n'est pas son âge exact au moment de sortie qui est calculé mais plutôt son âge au 31 décembre de l'année précédente, ce qui crée un biais pour le nombre de sorties observées à chaque âge.

Deuxième méthode

La deuxième méthode a pour objectif de calculer la contribution exacte de chaque bénéficiaire à l'exposition de chaque âge. Un exemple précis est détaillé ci-dessous :

- Date de naissance du rentier : 01/04/1991
- Date d'effet de la rente : 01/07/1998 (7 ans 3 mois)
- Date de clôture de la rente : 01/05/2014 (23 ans 1 mois)

Le bénéficiaire compte pour 1 dans les expositions des âges de 9 à 22 ans, pour 0,75 dans l'exposition de l'âge de 7 ans et pour 0,083 dans l'exposition de l'âge de 23 ans. Comme il s'agit d'une clôture pour motif connu, l'observation comptera pour 1 dans le nombre de sorties à l'âge de 23 ans. Il est alors théoriquement possible de se retrouver dans une situation où le taux de sortie serait supérieur à 1.

Comparaison des méthodes

Le graphique de la *figure 28* montre le résultat de calcul de l'exposition par les deux méthodes décrites ci-dessus.

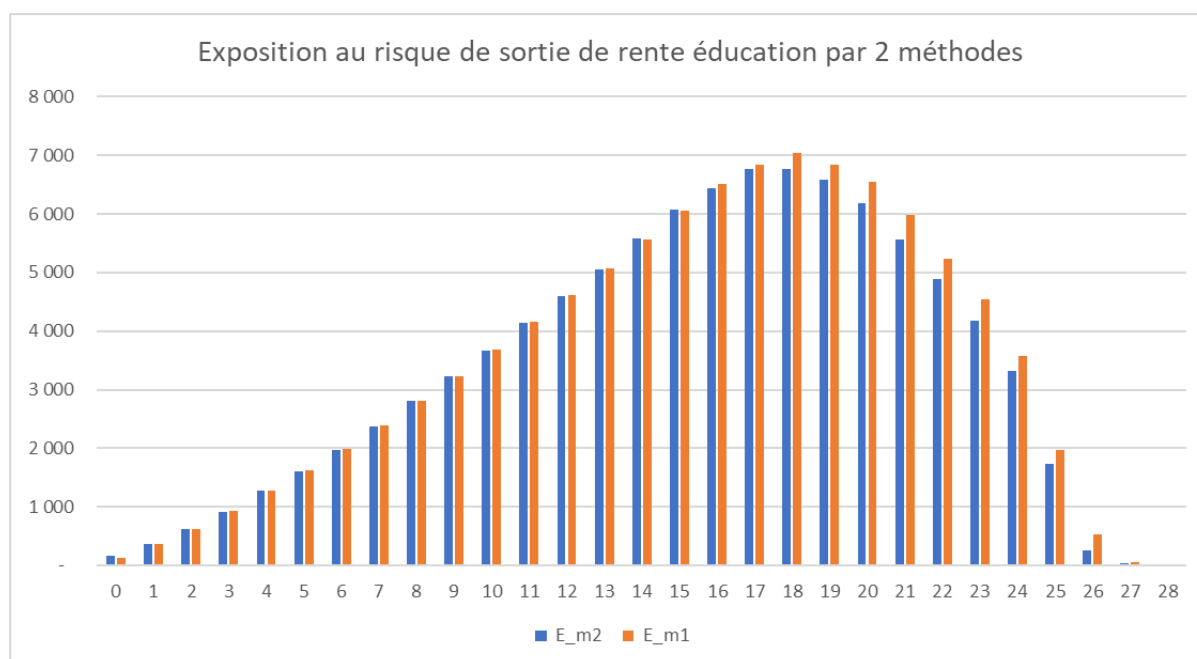


Figure 28. Exposition au risque de sortie du portefeuille calculée par 2 méthodes

Jusqu'à l'âge de 16 ans les expositions calculées par les deux méthodes sont très proches avec des montants légèrement supérieurs pour quasiment tous les âges pour la méthode 2. En revanche, la différence des expositions calculées par les deux approches devient ensuite de plus en plus visible, avec les expositions plus importantes pour la méthode 1 cette fois-ci. La différence qui existe est environ de ½ année pour chaque âge. Ceci est dû à la méthodologie retenue pour la première approche qui approxime toutes les entrées et les sorties en milieu de l'année et qui arrondit les âges au 31 décembre de l'année considérée.

Il est préférable de retenir la méthode 2 car la méthode 1 surestime les expositions pour les âges à partir des 16 ans, soit pour la période qui représente l'intérêt principal de la présente étude.

3.2 Estimation des taux bruts de sortie par les méthodes non paramétriques

La modélisation des taux de sortie de la rente éducation est réalisée par deux modèles non paramétriques : estimateur de taux de sortie de Hoem et de la fonction de survie de Kaplan Meier. Le choix de ces approches est principalement justifié par la présence des données incomplètes dans l'échantillon de données de taille réduite.

3.2.1 Estimateur de Hoem

Cadre théorique

Dans un premier temps, les taux bruts de sortie de la rente éducation à l'âge n sont calculés par la méthode de Hoem, à l'aide de l'expression suivante :

$$\hat{q}_n = \frac{N_n}{E_n}$$

Il s'agit du rapport entre le nombre de clôtures de rentes et l'exposition au risque à l'âge n . Avec l'exposition calculée par la méthode 2 définie dans les paragraphes précédents, les censures à droite sont bien prises en compte.

L'estimateur de Hoem est pertinent et sans biais. De plus, son avantage est dans l'interprétation facile des résultats obtenus.

En revanche, les taux de sortie bruts calculés par la méthode de Hoem représentent souvent des irrégularités qui sont expliquées par des fluctuations d'échantillonnage. L'étude des taux bruts ainsi obtenus n'est donc pas suffisante et il est important de trouver des indicateurs complémentaires. Il convient de faire appel aux intervalles de confiance qui permettent de visualiser la volatilité des taux calculés. Cependant, pour les utiliser, il faut analyser la taille de l'échantillon de données et choisir la méthode appropriée.

Deux approches peuvent être utilisées. Si la base de données est suffisamment grande, la loi Normale peut servir pour l'approximation des taux de sortie sous réserve que les conditions ci-dessous soient respectées :

$E_n * \hat{q}_n > 5$ et $E_n * (1 - \hat{q}_n) > 5$, alors $Q_n \sim N(q_n; \sigma_n)$

$$\text{Avec } \sigma_n = \sqrt{\frac{q_n * (1 - q_n)}{E_n}}$$

L'intervalle de confiance est donc évalué comme suit :

$$\hat{q}_n \pm u_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_n * (1 - \hat{q}_n)}{E_n}}$$

Où $u_{\frac{\alpha}{2}}$ représente le quantile $\frac{\alpha}{2}$ de la loi Normale centrée réduite.

Si l'exposition n'est pas suffisamment grande, il est possible d'utiliser l'intervalle de confiance modélisé par distance finie. Il faut alors trouver la valeur de m_α tel que :

$$P(\hat{q}_n - m_\alpha \leq \hat{q}_n + m_\alpha) \geq 1 - \alpha$$

La probabilité est calculée par multiplication des termes de l'inégalité par l'exposition à l'âge n :

$$\sum_{k=E_n * (\hat{q}_n - m_\alpha)}^{E_n * (\hat{q}_n + m_\alpha) + 1} P(D_n = k) \geq P(\hat{q}_n - m_\alpha \leq q_n \leq \hat{q}_n + m_\alpha) \geq 1 - \alpha$$

Ensuite l'algorithme itératif permet de calculer la probabilité $P(D_n = k)$ pour $k = E_n * \hat{q}_n$ et de la comparer à $(1 - \alpha)$. Si $(1 - \alpha)$ est plus élevé, alors le calcul se poursuit pour $\sum_{k=E_n * \hat{q}_n - j}^{E_n * \hat{q}_n + j} P(D_n = k)$ et s'arrête lorsque $m_\alpha = \frac{j}{E_n}$.

(Kamega A., Planchet F., 2010, Mesure du risque d'estimation associé à une table d'expérience.)

Dans l'échantillon de données très peu de sorties de la rente sont observées avant l'âge de 16 ans du bénéficiaire. C'est-à-dire que la condition qui permet d'approcher des taux bruts par la loi normale n'est pas respectée pour les âges de 0 à 16 ans. En revanche, la taille de l'échantillon permet d'approximer l'intervalle de confiance par la loi normale pour les âges allant de 16 à 28 ans. Etant donné le faible nombre de rentes clôturées avant 16 ans et l'obligation de scolarisation imposée par la législation, il n'est pas nécessaire de construire un intervalle de confiance pour la tranche d'âge entre 0 et 16 ans.

Taux bruts de sortie de la rente

Après avoir calculé l'exposition au risque et le nombre de sorties non censurées, le taux de sortie de la rente peut être calculé.

Le graphique de la *figure 29* affiche les taux bruts de sortie estimés par la méthode de Hoem, ainsi que les intervalles de confiance construits par l'approximation de la loi normale

pour la tranche d'âge de 16 à 28 ans. Les taux de sortie bruts sont très proches de 0 jusqu'à l'âge de 16 ans, ce qui se justifie par l'obligation d'instruction au niveau national et par un faible taux de mortalité des bénéficiaires. A partir de l'âge de 16 ans, le taux de sortie de la rente augmente progressivement jusqu'à l'âge de 26 ans qui est l'âge-limite contractuel de la garantie pour la plupart des contrats (28 ans pour de nouveaux contrats).

En outre, la vitesse de sortie de la rente éducation a tendance à augmenter avec l'âge des bénéficiaires. Le taux de sortie est de 100% à l'âge de 28 ans, ce qui est lié aux conditions du contrat et à la suppression des rentes viagères et des dossiers aberrants. La probabilité de clôture de la rente est élevée à l'âge de 27 ans avec une volatilité très importante qui s'explique par un faible nombre d'observations. Ceci montre qu'une partie non négligeable de rentes sont clôturées à l'âge-limite de garantie parce que les rentiers poursuivent des études de longue durée.

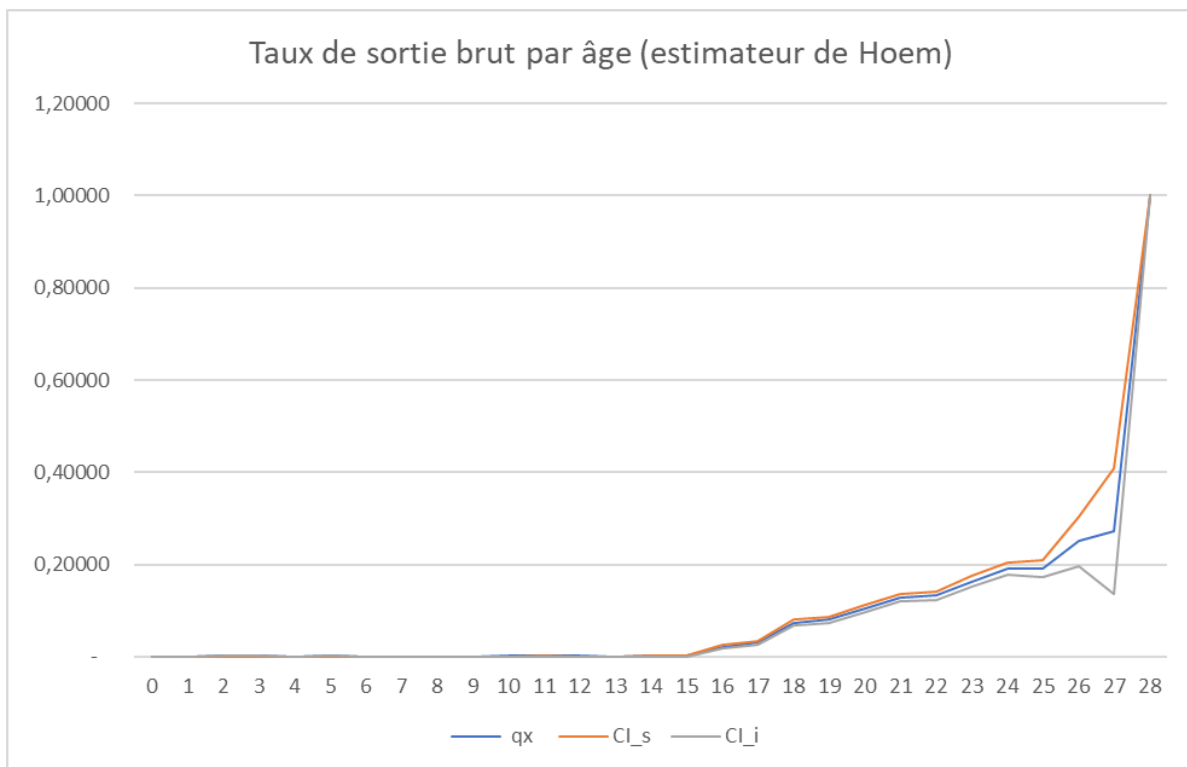


Figure 29. Taux bruts de sortie de la rente éducation (estimateur de Hoem)

3.2.2 Estimateur de Kaplan Meier

Cadre théorique

Dans la suite du mémoire, l'approche de Kaplan Meier est utilisée. Il s'agit de l'estimateur de la fonction de survie, à la différence de celui de Hoem, qui est l'estimateur des taux de sortie. La méthode de Kaplan Meier s'appuie sur le fait qu'être encore présent dans un échantillon à un moment t implique forcément la présence juste avant t . Autrement dit, si un rentier est toujours dans le portefeuille au moment t , ceci sous-entend qu'il a été en étude (et donc en rente) jusqu'au moment t . Formellement, le principe s'écrit comme suit :

$$S_t = Pr(T > t) = Pr(T > t, T > s) = Pr(T > t | T > s) * Pr(T > s) = Pr(T > t | T > s) * S_s$$

avec $s < t$ et S_t fonction de maintien.

Si les moments t sont choisis par critère de l'occurrence de l'évènement étudié, alors les probabilités de maintien en portefeuille entre T_i et T_{i+1} se calculent par la formule ci-dessous :

$$p_i = Pr(T > T_{i+1} | T > T_i)$$

Soit D_i variable indicatrice qui prend la valeur 1 si l'évènement se produit et 0 sinon (observation censurée). Avec ces notations, l'estimateur de maintien (de survie) de Kaplan Meier se calcule comme suit :

$$\hat{S}_t = \prod_{T_i \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{D_i}$$

Si les rentiers censurés sont précédés par ceux non censurés, alors la fonction de maintien s'écrit :

$$\hat{S}_t = \prod_{T_i \leq t} \left(1 - \frac{d_i}{E_i}\right)^{D_i}$$

Avec E_i – nombre de rentiers sous risque, et d_i – nombre de sorties de la rente éducation réellement observé.

Il faut préciser que l'estimateur de Kaplan Meier est celui du maximum de vraisemblance généralisé. Il est biaisé positivement, convergent, cohérent et asymptotiquement gaussien.

La variance de l'estimateur de Kaplan Meier, qui permet de construire les intervalles de confiance (asymptomatiques), peut être calculée à l'aide de l'estimateur de Greenwood dont la formule s'écrit :

$$\hat{V}(\hat{S}_t) = \hat{S}_t^2 \sum_{T_i \leq t} \frac{d_i}{r_i(r_i - d_i)}$$

(Planchet F., version Septembre 2020, Statistique des modèles non paramétriques)

Taux bruts de sortie de la rente

Dans le cadre du présent mémoire, l'approche de Kaplan Meier a été appliquée à l'aide du logiciel R avec le package Survival.

Le graphique de la *figure 30* affiche les taux bruts de sortie de la rente éducation calculés avec la méthode de Kaplan Meier. La courbe est quasiment plate jusqu'à l'âge de 15 ans avec une probabilité de sortie proche de 0. Ce constat est cohérent avec l'obligation légale de scolarisation jusqu'à l'âge de 16 ans. En effet, la rente éducation ne peut logiquement pas être suspendue pour le motif de fin d'études avant le 16^{ème} anniversaire du bénéficiaire. La sortie de la rente peut tout de même avoir lieu pour d'autres motifs (par exemple, justificatif absent) ou bien si le rentier décède. Etant donné le faible taux de mortalité pour cette tranche d'âge, la probabilité de sortie de rente proche de 0 semble être naturelle. A partir de l'âge de 15 ans, le taux de sortie commence à augmenter. La vitesse de croissance des taux est assez stable. Quelques pics sont cependant observés sur le graphique : ils correspondent à une hausse relativement plus importante constatée à 18, 24, 26 et 28 ans. Les âges de 18 et 24 ans sont généralement ceux de la transition vers un autre cycle d'enseignement (obtention du baccalauréat à 18 ans et fin des études supérieures à 24 ans). Il est alors naturel de constater un surplus du nombre de sorties de la rente à ses âges. En revanche, les pics à 26 et à 28 ans sont plutôt expliqués par les conditions contractuelles de la garantie : la majorité des rentes prennent effet au 26^e anniversaire du bénéficiaire et certaines peuvent aller jusqu'à son 28^e anniversaire. Etant donné que le présent mémoire ne traite que le cas de rente temporaire, l'intégralité des rentiers doit sortir du portefeuille à l'âge de 28 ans, d'où le taux de sortie égale à 1.

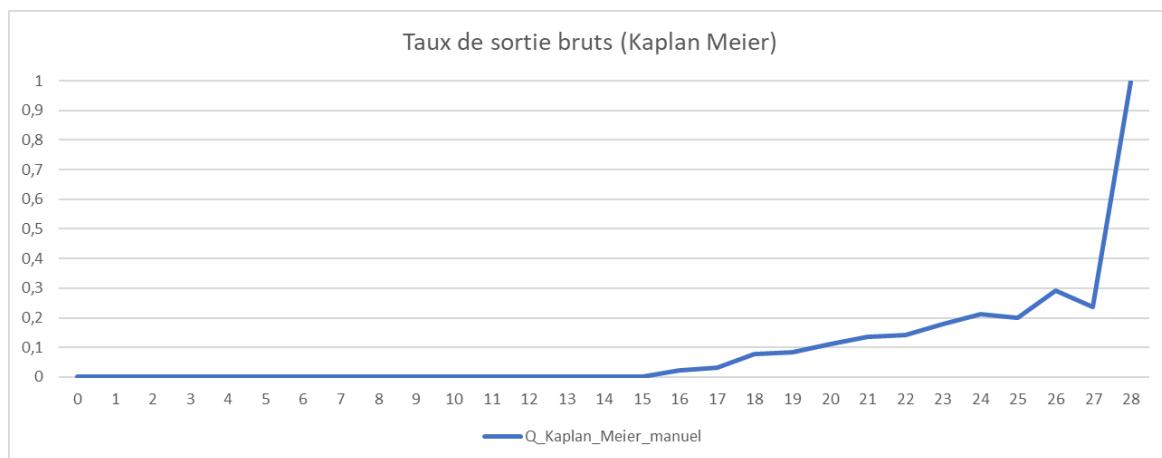


Figure 30. Taux bruts de sortie de la rente éducation (estimateur Kaplan Meier)

Fonction de maintien en rente

La *figure 31* fait apparaître la courbe de maintien en rente éducation obtenue avec l'utilisation de l'estimateur Kaplan Meier à l'aide du logiciel R (Samsom E., 2008, Identification des marqueurs pronostiques chez les patients atteints d'un cancer de la tête et du cou). La probabilité de maintien en rente ne baisse que très légèrement jusqu'à l'âge de 15 ans. A partir de l'âge de 16 ans, la courbe se met à décroître au fur et à mesure pour atteindre la survie nulle vers 28 ans. La vitesse de la décroissance de la courbe de maintien est plus importante pour les âges plus élevés, ce qui s'explique à la fois par la hausse du nombre de

rentiers sortant du système éducatif, par la diminution de la taille de l'exposition et aussi par la hausse du taux de mortalité des bénéficiaires.

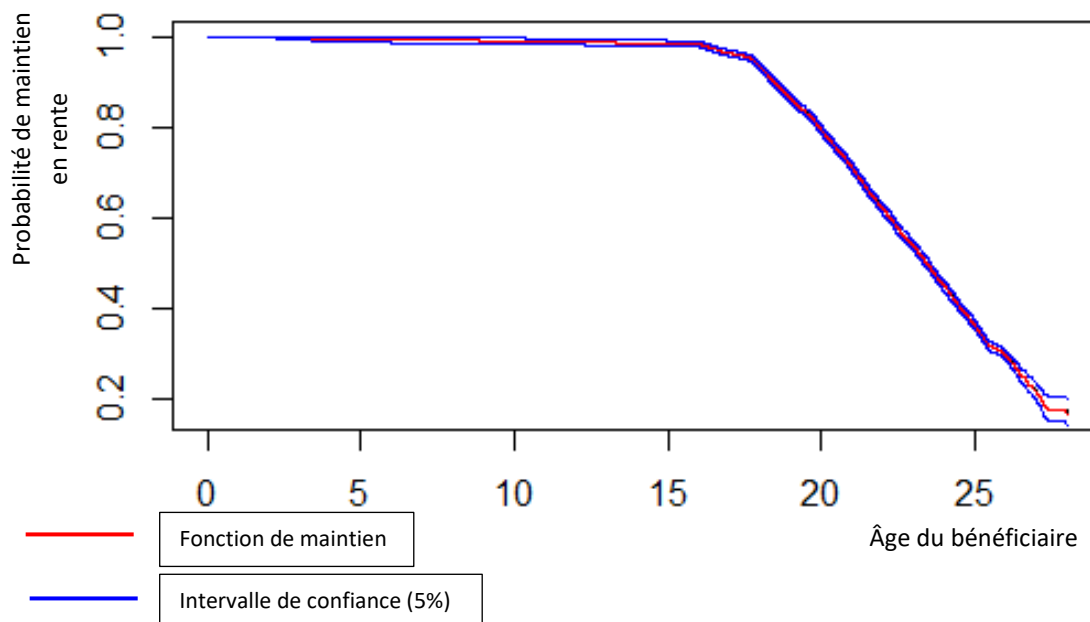


Figure 31. Fonction de maintien en rente (estimateur Kaplan Meier)

3.3 Lissage des taux bruts de sortie

Dans les paragraphes précédents, les taux bruts de sortie de la rente ont été calculés par deux méthodes non paramétriques. Il s'avère que ces taux présentent une certaine irrégularité qui est liée à une imperfection de la méthode de calcul de l'exposition et aux fluctuations de l'échantillon de données. L'irrégularité des courbes obtenues augmente la volatilité de la durée estimée et donc la probabilité d'erreur dans le calcul des provisions.

Afin d'améliorer la qualité du modèle, il faut procéder au lissage des taux bruts. Il s'agit de l'ajustement des valeurs brutes obtenues afin de se rapprocher plus de la loi réelle de maintien qui n'est pas connue. Concrètement, il faut calculer des valeurs \hat{q}_n en minimisant le montant de l'erreur $e_n = \hat{q}_n - q_n$ avec q_n qui définit un vrai taux de sortie de la rente. La courbe des taux \hat{q}_n ainsi obtenue sera plus lisse que celle des taux bruts.

Dans ce mémoire, les résultats sont lissés avec la méthode de Whittaker-Henderson en dimension 1. Cette approche permet d'ajuster les taux sans chercher leur représentation paramétrique qui est compliquée dans le cadre d'un échantillon de données de petite taille.

L'approche consiste à prendre en compte deux critères : celui de fidélité et celui de régularité.

- Critère de fidélité $F = \sum_{i=1}^k w_i (q_i - \hat{q}_i)^2$

Le critère de fidélité est responsable des écarts entre des valeurs brutes et des valeurs lissées. Concrètement, l'objectif est de mesurer la distance quadratique entre le taux estimé pondéré par un poids w_i et le taux lissé. Ce chiffre doit donc être minimisé afin de retrouver la courbe pas très éloignée des données d'origine.

- Critère de régularité : $S = \sum_{i=1}^{k-z} (\Delta^z q_i)^2$

Le critère de régularité répond au besoin initial d'obtenir la courbe plus régulière. Ce chiffre doit lui aussi être minimisé.

Le but de la méthode est de trouver un \hat{q} qui minimise le nombre $M = F + h * S$, qui représente l'agrégation des deux critères. Il est donc nécessaire de choisir les paramètres z et h et la méthode d'attribution des poids qui répondent au mieux aux besoins de l'étude.

Le nombre z apparaît dans le critère de régularité et détermine le degré du polynôme. Ainsi, $z=1$ revient à retrouver la distance euclidienne entre deux valeurs consécutives. La valeur de z entre 2 et 5 permet d'inclure plus d'informations consécutives mais augmente la complexité de la recherche de solution. Il est décidé de retenir $z=2$ qui est le cas le plus souvent utilisé.

L'objectif est de lisser la valeur de taux q_i par rapport aux valeurs de $q_{i+1}, q_{i+2}, \dots, q_{i+z}$. Soit $\Delta q_i = q_{i+1} - q_i$, d'où par itération :

$$\Delta^z q_i = \sum_{j=1}^z \binom{z}{j} (-1)^{z-j} q_{i+j}$$

Minimiser la valeur S signifie donc à minimiser les écarts entre les q_i et q_{i+1} , les q_{i+1} et q_{i+2} , et ainsi de suite jusqu'à q_{i+z} (Planchet F., 2020, Modèles de durée. Présentation des méthodes de lissage et d'ajustements).

Le paramètre h permet de quantifier la préférence pour un des deux critères. En choisissant un h plus élevé, l'utilisateur accorde plus d'importance au critère de régularité, ce qui pousse à sacrifier une partie de la précision de la courbe par rapport aux données d'origine afin d'obtenir une courbe plus lisse.

Pour choisir le paramètre h optimal, le lissage des taux bruts a été réalisé pour trois valeurs : $h=1, 5, 10$.

Enfin, le dernier paramètre à choisir pour réaliser le lissage est le poids w_i . Ce paramètre permet d'indiquer quelle importance est attribuée à chaque observation, soit à chaque âge. Trois approches ont été testées pour l'attribution de poids dans la présente étude : le poids identique pour chaque âge, le poids pondéré par l'exposition au risque, le poids attribué de manière analytique. La dernière approche s'appuie sur l'idée que les âges

différents n'ont pas la même importance pour la construction de la loi de maintien en rente. Selon les données du portefeuille, la sortie de la rente éducation est un évènement rare pour les individus de moins de 16 ans. De plus, il s'agit des âges pour lesquels la scolarisation est obligatoire. De ce fait, ces observations ne rapportent pas d'information très importante pour le provisionnement et il est décidé de leur attribuer le poids minimal, soit 1. En outre, certains âges semblent être plus importants pour la construction de la loi car il s'agit des âges pivot quant au comportement des rentiers : 18 et 26 ans. L'âge de 26 ans reflète les conditions contractuelles de la majorité des contrats. L'âge de 18 ans correspond à l'âge d'obtention du baccalauréat et du passage au cycle d'enseignement suivant. De ce fait, le poids de 20 est attribué à ces observations. L'âge de 28 ans est peu important, car très peu présent dans l'échantillon et correspond à l'âge-limite pour le reste des contrats, alors le poids de 1 y est attribué. Enfin le poids de 10 est attribué à tous les âges restants.

Une fois les taux lissés obtenus, il faut s'assurer qu'ils représentent les taux bruts de manière suffisamment fidèle afin d'éviter les biais d'estimation de la loi de sortie. Plusieurs méthodes de validation de lissage existent, deux parmi elles sont utilisées dans le cadre du présent mémoire. Tout d'abord, le lissage sera contrôlé par l'analyse graphique qui consiste à comparer les courbes des taux bruts avec celles des taux lissés. En revanche, la méthode reste assez subjective et la conclusion sur la qualité de lissage n'est pas toujours facile à établir. De ce fait, il est nécessaire de renforcer la prise de décision par une approche quantitative qui serait plus objective. Le critère d'information Akaike qui est un test pertinent semble être bien adapté pour le choix des paramètres de lissage. En effet, le critère Akaike, connu sous abréviation *AIC*, permet non seulement d'évaluer la bonne adéquation d'un lissage mais aussi de comparer plusieurs configurations entre eux afin de n'en retenir qu'une.

La statistique du test de fidélité se calcule de la manière suivante :

$$AIC = -2\ln L + 2k$$

Avec k – nombre de paramètre du modèle et L – vraisemblance du modèle lissé

Le lissage le plus adapté pour l'ajustement des taux bruts est celui qui minimise la valeur de l'information Akaike (Guibert Q., Planchet F., 2014, Construction de lois d'expérience en présence d'évènements concurrents : Application à l'estimation des lois d'incidence d'un contrat dépendance).

3.3.1 Lissage de l'estimateur de Hoem

Dans un premier temps, les taux bruts de Hoem sont lissés avec les trois modalités du paramètre h et les trois méthodes d'attribution du poids.

Les figures 32, 34, 36 affichent les courbes des taux bruts de Hoem lissés avec trois différentes valeurs de h et trois méthodes d'attribution du poids respectivement : poids identiques, poids proportionnels à l'exposition et poids analytiques. Pour faciliter la lecture des graphiques, les courbes sont affichées uniquement pour l'intervalle d'âge allant de 16 à

28 ans, car les taux avant l'âge de 16 ans sont proches de 0. Les figures 33, 35, 37 présentent les valeurs de critère *AIC* calculées pour chaque configuration de lissage appliqué aux taux bruts de Hoem.

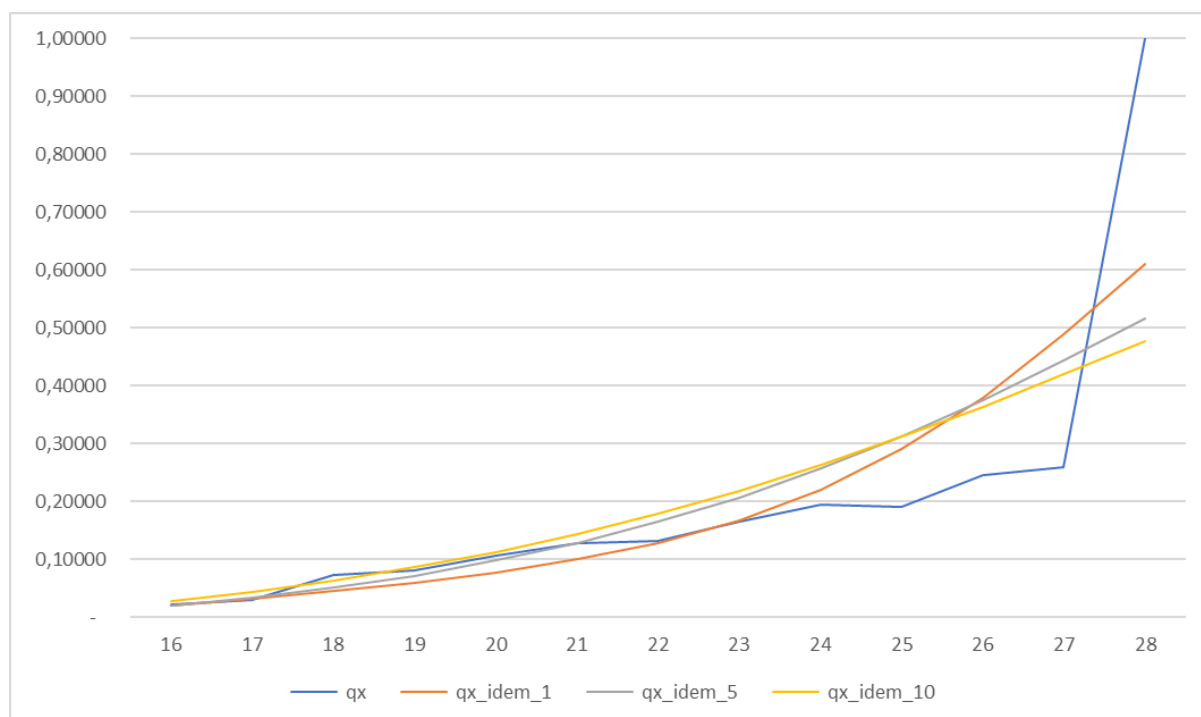


Figure 32. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids identiques (estimateur de Hoem) pour l'intervalle de 16 à 28 ans

Selon la *figure 32*, le lissage avec les poids identiques attribués à toutes les observations ne reflète pas bien la courbe des taux bruts de Hoem. Les écarts entre les courbes lissées et la courbe des taux bruts sont les plus importants pour la tranche d'âge entre 23 et 27 ans. La courbe du lissage avec $h=1$ sous-estime les taux bruts sur l'intervalle entre 17 et 23 ans tout comme pour l'âge de 28 ans et surestime les taux entre 23 et 27 ans. La même tendance est observée pour le lissage avec un $h=5$ mais les écarts semblent être moins prononcés. La courbe des taux lissés avec un $h=10$ sous-estime les taux pour les âges de 18 ans et de 28 ans. La statistique *AIC* présentée dans la *figure 33* vient compléter l'analyse graphique et indique de retenir le lissage avec un $h=1$.

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Hoem)	identique	1	-51,6495
Taux bruts (Hoem)	identique	5	-43,8635
Taux bruts (Hoem)	identique	10	-41,1522

Figure 33. Valeurs de *AIC* pour le lissage des taux bruts de Hoem pour $z=2$, poids identiques et $h=1, 5, 10$

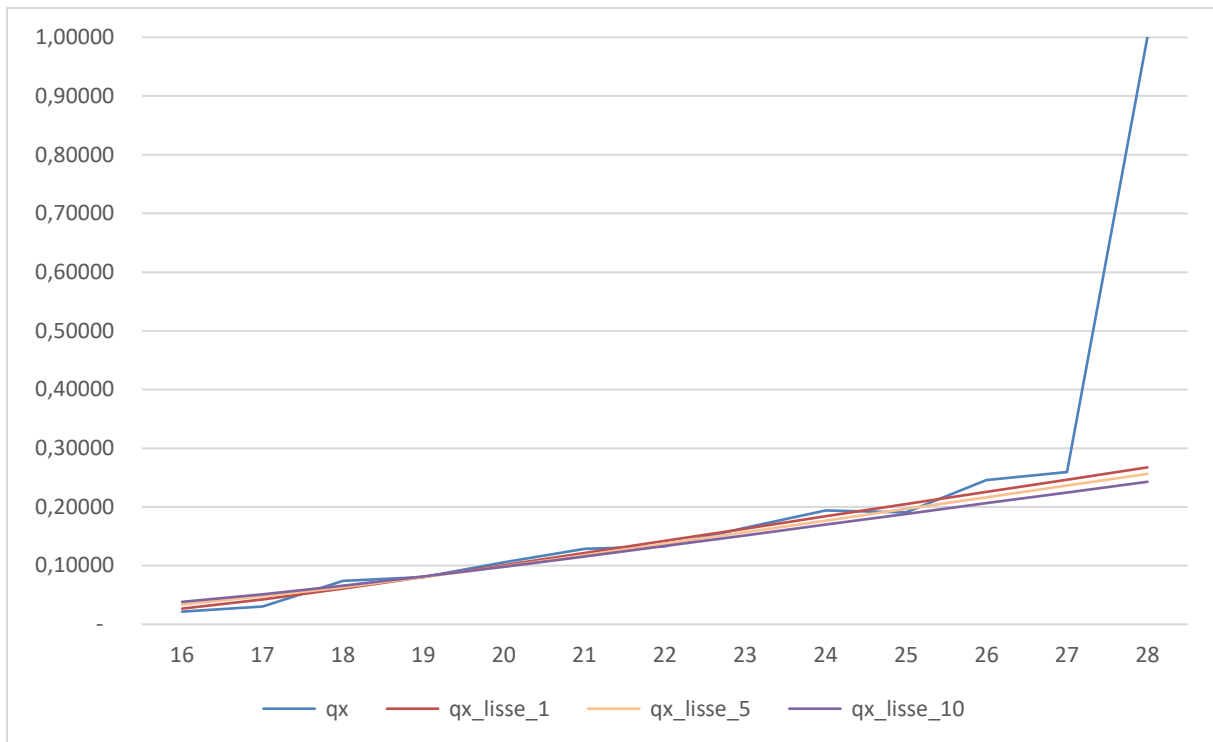


Figure 34. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids proportionnels à l'exposition (estimateur de Hoem) pour l'intervalle de 16 à 28 ans

La figure 34 permet de comparer les lissages de taux pour 3 paramètres différents de h dans le cas de l'attribution des poids proportionnels à l'exposition à chaque âge. Les 3 courbes des taux lissés ne font pas apparaître de gros écarts par rapport à celle de taux bruts. La manière d'attribuer les poids semble donc être mieux adaptée que dans le paragraphe précédent. Le lissage avec un $h=1$ donne les taux plus élevés que celui avec un $h=10$ pour les âges supérieurs à 19 ans. La courbe avec un $h=5$ se positionne entre les deux autres courbes pour tous les âges. Globalement, les taux lissés avec cette attribution des poids surestiment les taux de sortie entre 13 et 17 ans. La courbe avec un $h=1$ surestime aussi les taux de sortie pour 22 et 25 ans mais donne les taux plus proches des taux bruts à partir des 26 ans.

De nouveau, il est préférable de vérifier les valeurs du test quantitatif pour déterminer le meilleur paramètre de h dans le cas des poids proportionnels à l'exposition. Selon le tableau de la figure 35, la valeur de h minimisant le critère d'information AIC est de 1.

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Hoem)	proportionnel à l'exposition	1	-35,0656
Taux bruts (Hoem)	proportionnel à l'exposition	5	-34,4166
Taux bruts (Hoem)	proportionnel à l'exposition	10	-33,1411

Figure 35. Valeurs de F pour le lissage des taux bruts de Hoem pour $z=2$, poids proportionnels à l'exposition et $h=1,5,10$

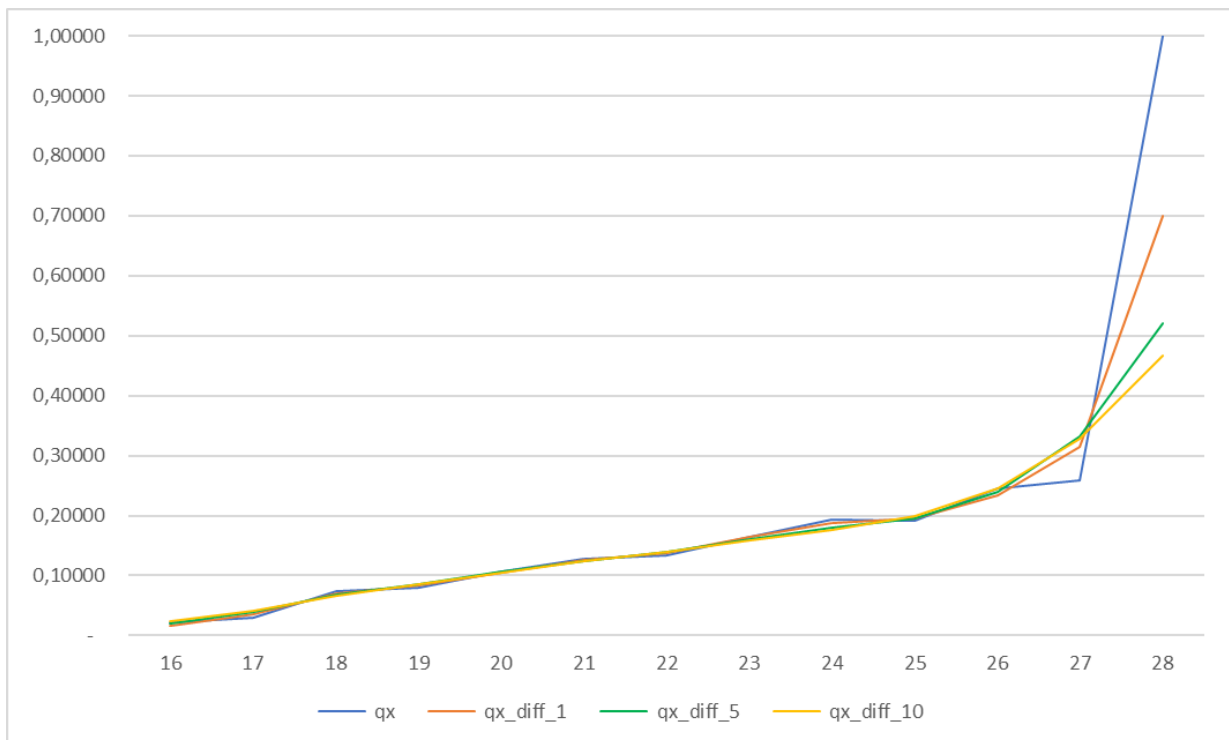


Figure 36. Taux de sortie lissés pour $h=1,5,10$ avec des poids choisis analytiquement (estimateur de Hoem) pour l'intervalle de 16 à 28 ans

Enfin, la *figure 36* permet de comparer les courbes des taux lissés construites pour les poids attribués de manière analytique pour les modalités de h égales à 1, 5 ou 10. Les trois courbes semblent bien respecter les fluctuations de la courbe des taux bruts. Les écarts assez significatifs apparaissent seulement pour les âges supérieurs à 26 ans. Le lissage avec un $h=10$ surestime le taux de sortie à 27 ans et sous-estime celui à 28 ans de manière plus prononcée que les autres lissages. Cette observation se confirme par le critère quantitatif qui apparaît dans le tableau de la *figure 37*. Selon le critère d'information AIC, le lissage retenu pour l'estimateur de Hoem avec l'attribution analytique des poids est celui avec le paramètre h égal à 1.

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Hoem)	analytique	1	-89,2919
Taux bruts (Hoem)	analytique	5	-72,4689
Taux bruts (Hoem)	analytique	10	-65,1845

Figure 37. Valeurs de F pour le lissage des taux bruts de Hoem pour $z=2$, poids analytiques et $h=1, 5, 10$

Une fois la meilleure modalité du paramètre h définie pour chaque méthode d'attribution des poids, il faut procéder à la comparaison des trois lissages retenus. L'objectif est de déterminer quelle méthode d'attribution des poids aux observations donne les taux lissés les plus fidèles aux taux bruts. La *figure 38* permet de comparer graphiquement les trois courbes obtenues. L'attribution des poids identiques à chaque âge donne un lissage non optimal car les taux lissés pour les âges entre 22 et 27 ans semblent être fortement surestimés par rapport aux taux bruts. Ceci crée un risque de sous-estimation des provisions nécessaires.

La méthode de pondération par exposition au risque donne des taux lissés relativement proches des taux bruts. En revanche, la courbe ainsi obtenue semble être plus lisse que les autres : les taux lissés sont surestimés entre 14 et 16 ans et sous-estimés à partir de 26 ans. Selon le graphique, l'attribution des poids analytique semble donner les taux lissés les plus fidèles aux taux bruts. Cette hypothèse se confirme aussi par les valeurs des statistiques AIC présentées dans les *figures 33, 35 et 37*. Finalement, les meilleurs paramètres pour le lissage pour les taux bruts de Hoem sont : $z=2$, $h=1$, attribution analytique des poids.

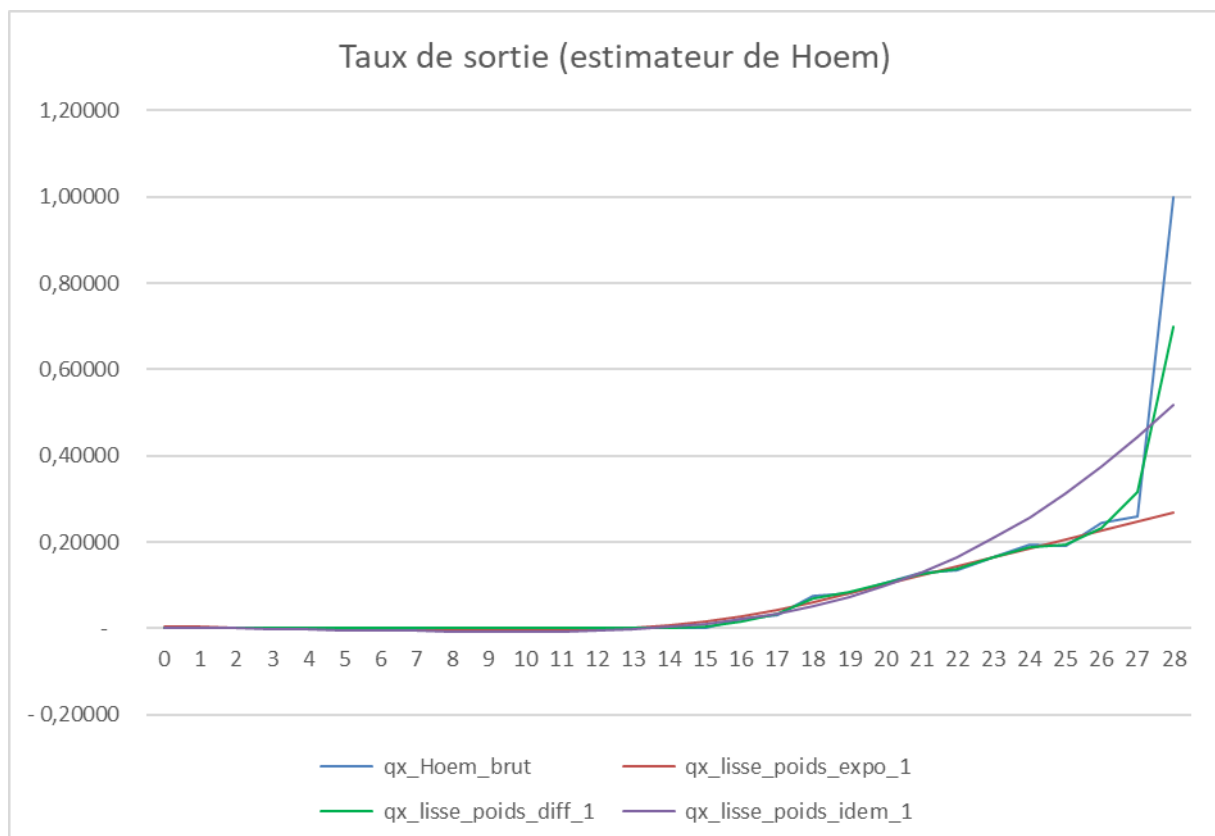


Figure 38. Taux de sortie lissés pour $z=2$ avec des poids et le h attribués selon les 3 approches (estimateur de Hoem)

Une fois le lissage sélectionné, il faut accorder l'attention au fait que pour certains âges (0, 13 et 14 ans) les taux lissés sont négatifs, ce qui n'est pas admissible dans le contexte de la présente étude. De ce fait, il a été décidé de remplacer les taux lissés négatifs par les taux bruts des âges correspondants. De plus, le taux lissé pour l'âge de 27 ans semble être trop élevé et risque de créer un biais de sous-provisionnement, alors le taux finalement retenu pour cet âge est celui brut et non pas celui lissé. La *figure 39* affiche les taux bruts, les taux lissés et les taux finalement retenus pour l'estimateur de Hoem.

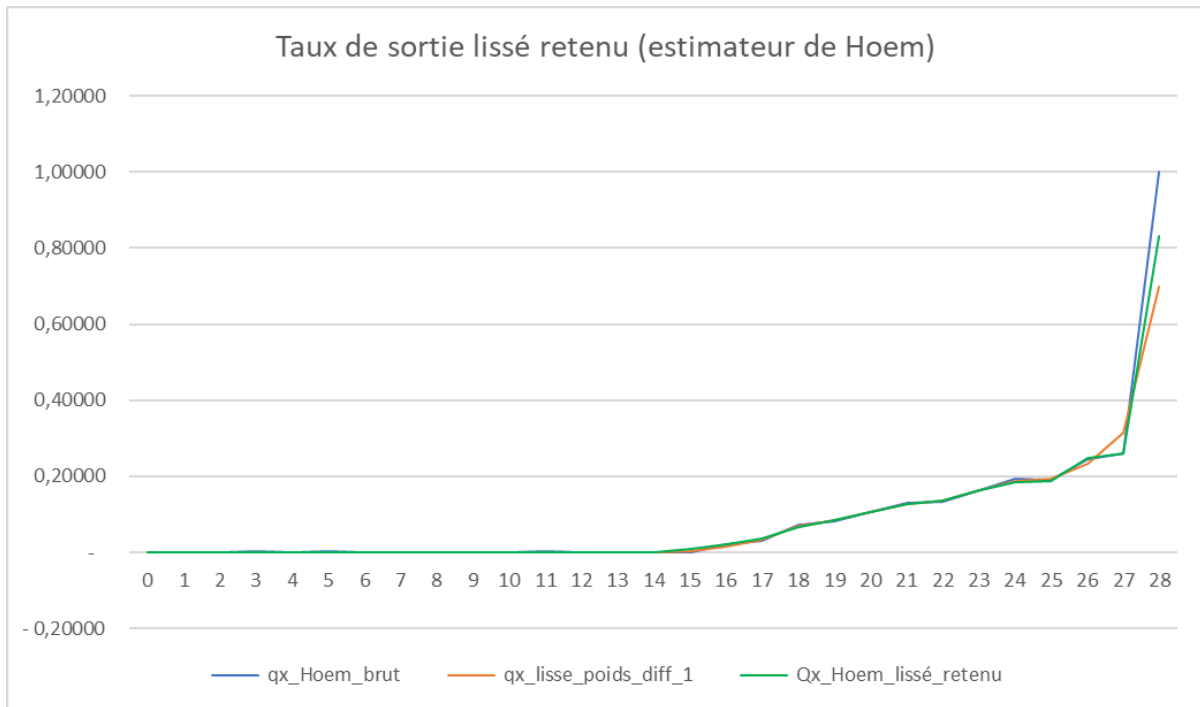


Figure 39. Taux de sortie bruts, lissés et retenus (estimateur de Hoem)

3.3.2 Lissage de l'estimateur de Kaplan Meier

Dans un second temps, le lissage de Whittaker-Henderson est appliqué aux taux bruts de sortie obtenus à l'aide de l'approche de Kaplan Meier. Les graphiques des figures 40, 42, 45 font apparaître les courbes des taux bruts et lissés avec le paramètre h prenant les modalités 1, 5 et 10 pour les trois approches d'attribution des poids respectives. De nouveau, les graphiques ne font apparaître que l'intervalle d'âge entre 16 et 28 ans, parce que les taux pour les âges inférieurs à 16 ans ne sont pas très différents de 0. Les valeurs des statistiques AIC pour chaque méthode d'attribution des poids sont affichées dans les tableaux des figures 41, 43 et 45.

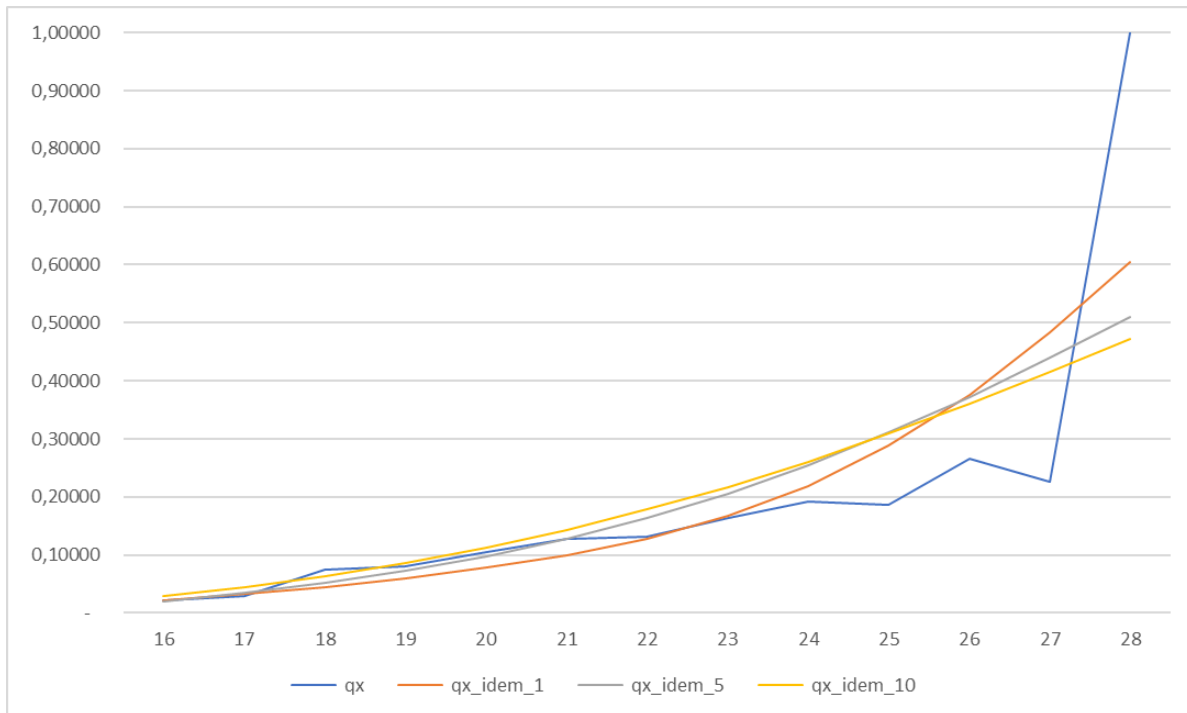


Figure 40. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids identiques (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans

Selon la *figure 40*, le lissage avec des poids identiques semble donner les taux lissés relativement éloignés des taux bruts, surtout pour les âges supérieurs à 23 ans. Les lissages avec $h=5$ et 10 surestiment significativement les taux de sortie à partir de l'âge de 21 ans et sous-estiment le taux pour 27 et 28 ans. Le lissage avec un $h=1$ donne les taux lissés inférieurs aux bruts pour les âges allant de 17 à 22 et de 27 à 28 ans. La tendance s'inverse pour l'intervalle de 23 à 26 ans. Selon le test de fidélité affiché dans la *figure 41*, le lissage avec un $h=1$ minimise la statistique F .

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Kaplan Meier)	identique	1	-50,0308
Taux bruts (Kaplan Meier)	identique	5	-42,8113
Taux bruts (Kaplan Meier)	identique	10	-40,2699

Figure 41. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids identiques et $h=1, 5, 10$

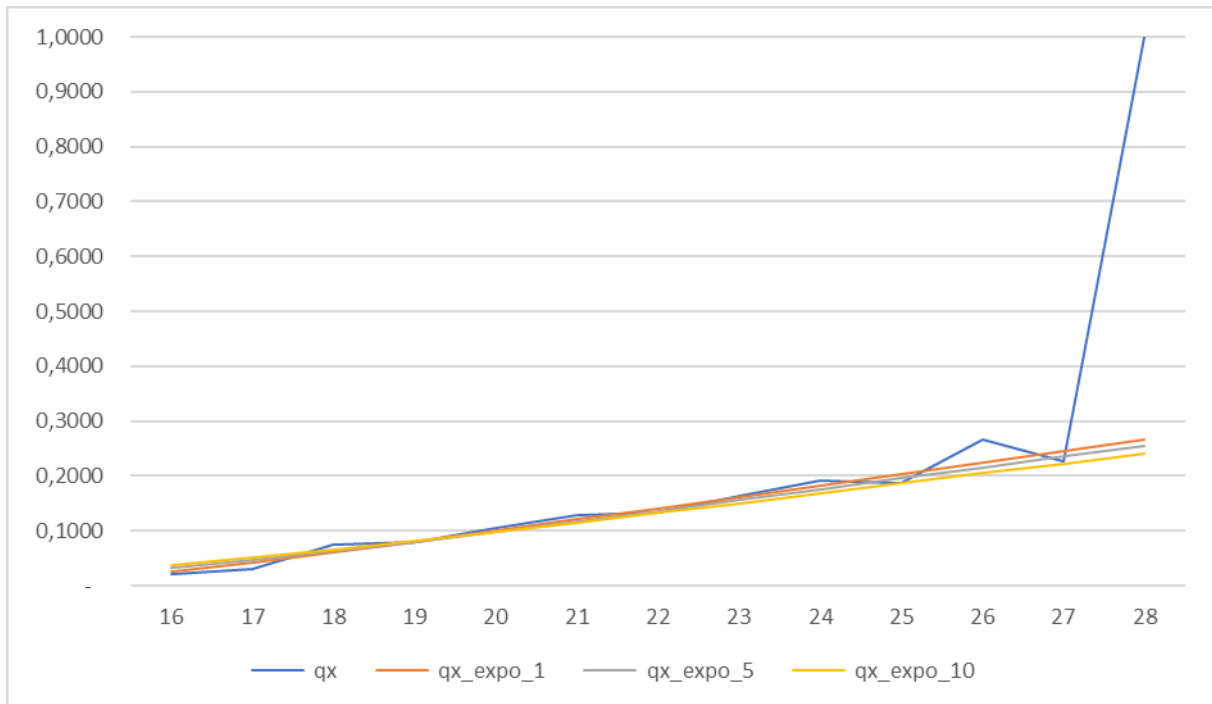


Figure 42. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids proportionnels à l'exposition (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans

La figure 42 affiche les courbes des taux lissés avec des poids proportionnels à l'exposition. La courbe du lissage avec un $h=10$ est plus plate que les autres. De ce fait, les taux lissés avec cette configuration sont plus élevés pour les âges supérieurs à 19 ans et plus bas pour les âges inférieurs à 18 ans. Les montants de critère d'information AIC indiqués dans la figure 43 permettent de conclure que le lissage le plus fidèle aux données brutes est celui avec un $h=1$.

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Kaplan Meier)	proportionnel à l'exposition	1	-34,5268
Taux bruts (Kaplan Meier)	proportionnel à l'exposition	5	-33,9016
Taux bruts (Kaplan Meier)	proportionnel à l'exposition	10	-32,6767

Figure 43. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids proportionnels à l'exposition et $h=1, 5, 10$

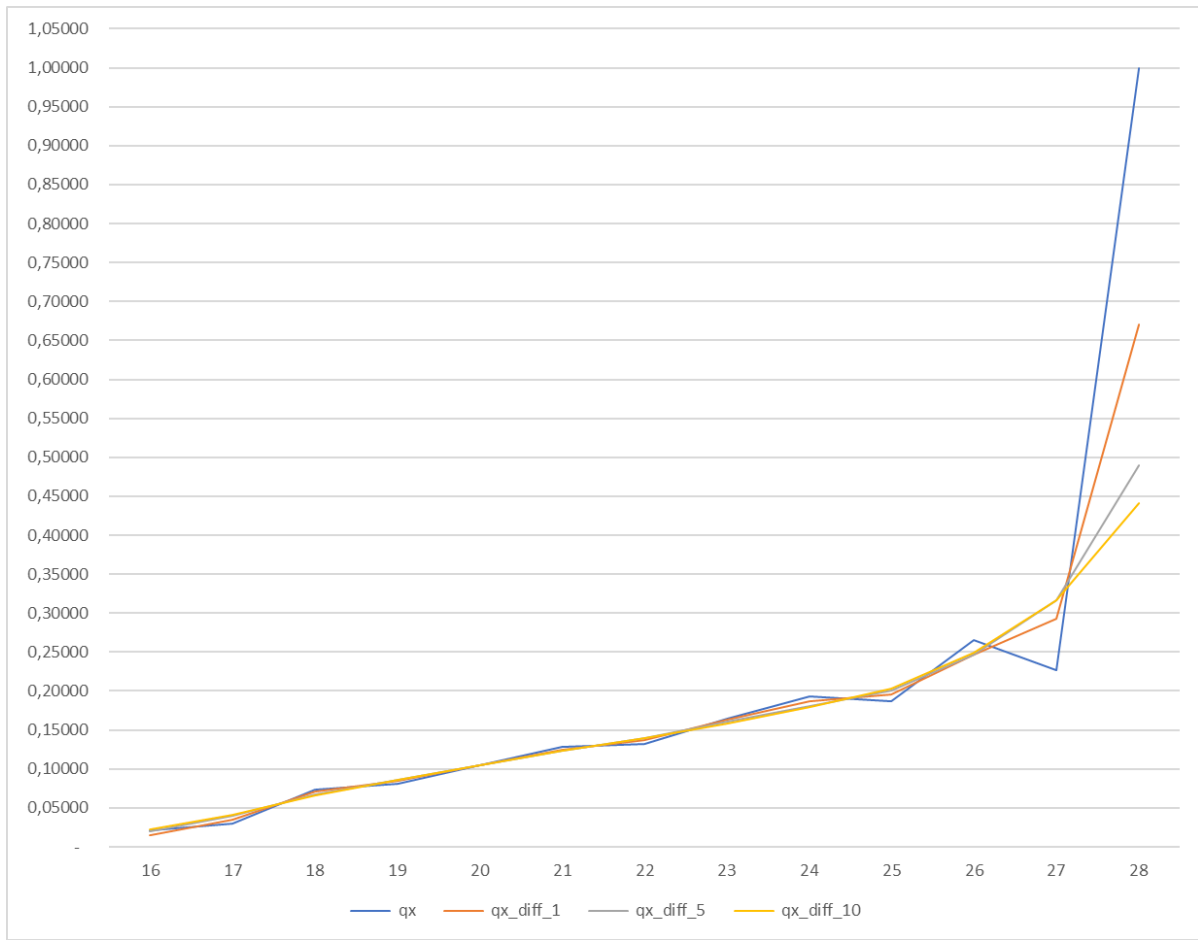


Figure 44. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids choisis analytiquement (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans

Enfin la *figure 44* fait apparaître les courbes des taux lissés avec l'attribution analytique des poids pour 3 modalités de h . Les trois lissages semblent donner les taux assez fidèles aux données brutes à l'exception des taux à 27 et 28 ans. À ces âges, le lissage avec un $h=10$ est celui qui donne des taux les plus éloignés des taux bruts. Il est difficile de faire le choix entre les lissages avec $h=1$ et $h=5$ en s'appuyant sur les graphiques car les courbes lissées sont assez proches. La comparaison des valeurs de critère *AIC* affichée dans le tableau de la *figure 45* permet de faire le choix de manière objective. Ainsi, le lissage avec un $h=1$ est celui qui minimise le montant de *AIC*, il sera donc retenu pour la suite de l'étude.

Lissage Whittaker-Henderson	Poids	h	AIC
Taux bruts (Kaplan Meier)	analytique	1	-81,9168
Taux bruts (Kaplan Meier)	analytique	5	-67,5746
Taux bruts (Kaplan Meier)	analytique	10	-61,6829

Figure 45. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids analytiques et $h=1, 5, 10$

Le graphique de la *figure 46* permet de comparer les courbes des taux bruts et lissés calculés par la méthode de Kaplan Meier pour les trois approches de détermination des poids avec le meilleur choix de h pour chaque configuration. Selon le graphique, le lissage avec des poids identiques pour toutes les observations ne semble pas être correct pour approcher des

taux bruts. En effet, une forte surestimation des taux est observée pour les âges entre 24 et 27 ans. Ce lissage ne peut donc pas être retenu. La méthode des poids proportionnels à l'exposition au risque donne des taux lissés plus fidèles aux taux d'origine, mais elle propose une courbe très lisse. Finalement, la méthode d'attribution des poids de manière analytique semble être la meilleure pour le lissage des taux bruts de Kaplan Meier car elle apporte un équilibre entre la précision et la régularité. L'analyse graphique est certainement subjective, il est judicieux alors de comparer des statistiques *AIC* affichées dans les *figures 41, 43 et 45*. Les chiffres observés confirment l'hypothèse issue du test visuel et c'est bien le lissage avec les poids attribués analytiquement avec les paramètres $z=2$ et $h=1$ qui est retenu.

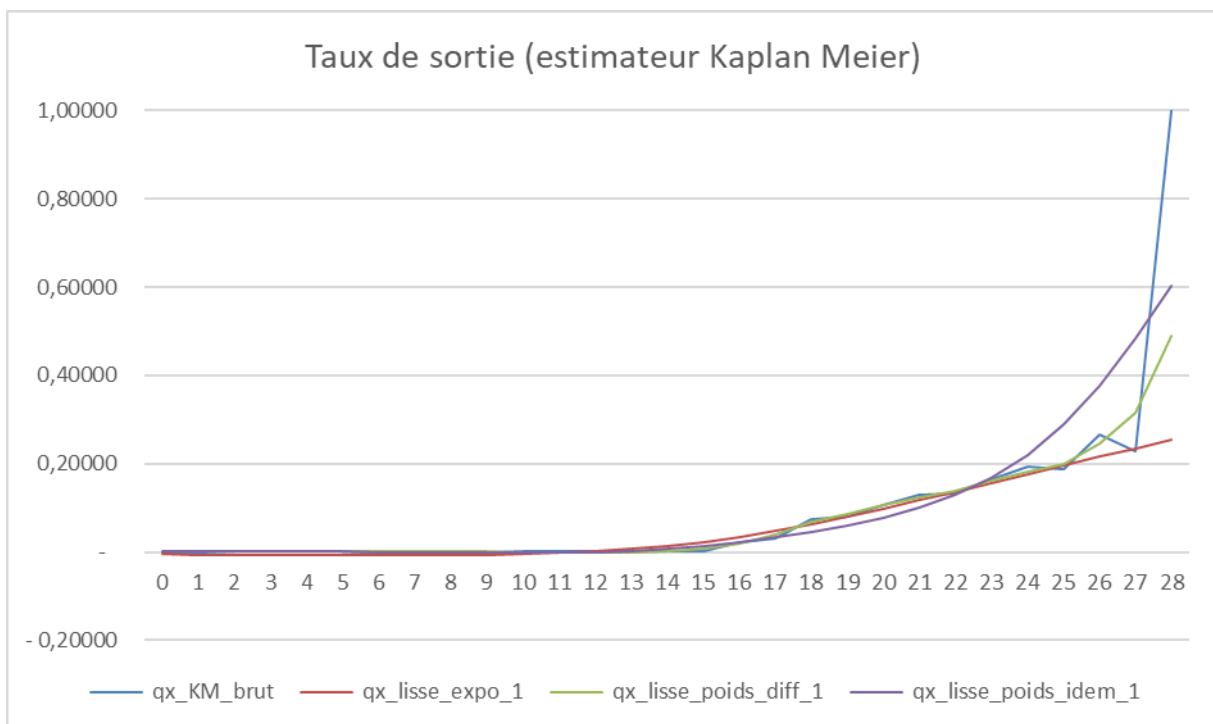


Figure 46. Taux de sortie lissés pour $z=2$ avec des poids et le h attribués selon les 3 approches (estimateur de Kaplan Meier)

Une fois le lissage sélectionné, il est important de corriger les incohérences rencontrées lors de son application. En effet, les taux lissés sont négatifs pour les âges de 0, 13 et 14 ans ce qui n'est pas possible par définition de taux dans le cadre de la présente étude. Les taux lissés sont alors remplacés par les taux bruts pour ces âges. De plus, le taux de sortie de 27 ans semble être trop élevé par rapport au taux brut, ce qui s'explique par un petit nombre d'observations. Sachant qu'il s'agit du contexte de provisionnement qui doit respecter le principe de prudence, il est judicieux de remplacer le taux lissé pour 27 ans par le taux brut du même âge.

Le graphique de la *figure 47* est une démonstration de la courbe finale ainsi que celle des taux bruts et des taux lissés. Visiblement la courbe retenue reflète bien celle des taux bruts et rajoute un peu de régularité par rapport à la présentation initiale.

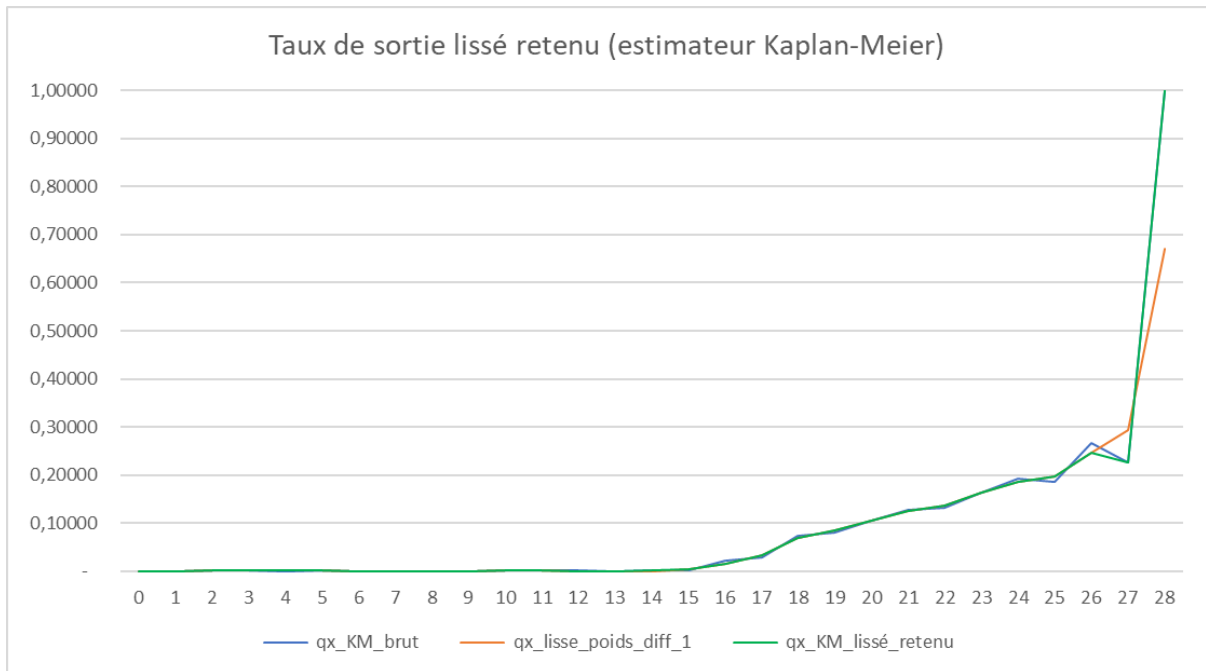


Figure 47. Taux de sortie bruts, lissés et retenus (estimateur de Kaplan Meier)

3.3.3 Comparaison des résultats

Dans les paragraphes précédents deux méthodes non-paramétriques ont été appliquées pour la construction de la loi de maintien en rente éducation à partir des données du portefeuille d'AXA. Les résultats de deux approches peuvent à présent être comparés afin de ne retenir qu'une seule loi pour l'estimation du montant des provisions nécessaires. La *figure 48* permet de comparer les taux de sortie de la rente retenus selon les deux méthodes. L'intervalle entre 0 et 16 ans semble être moins important pour la prise de décision car le nombre de sorties est très faible pour ces âges et les deux estimateurs donnent les résultats quasiment identiques.

Les deux courbes des taux de sortie semblent être très proches sur l'intervalle entre 16 et 25 ans. Les écarts apparaissent uniquement pour les âges de 26, 27 et 28 ans : le taux de sortie estimé par l'approche de Hoem est plus élevé à 27 et 28 ans et plus bas à l'âge de 26 ans.

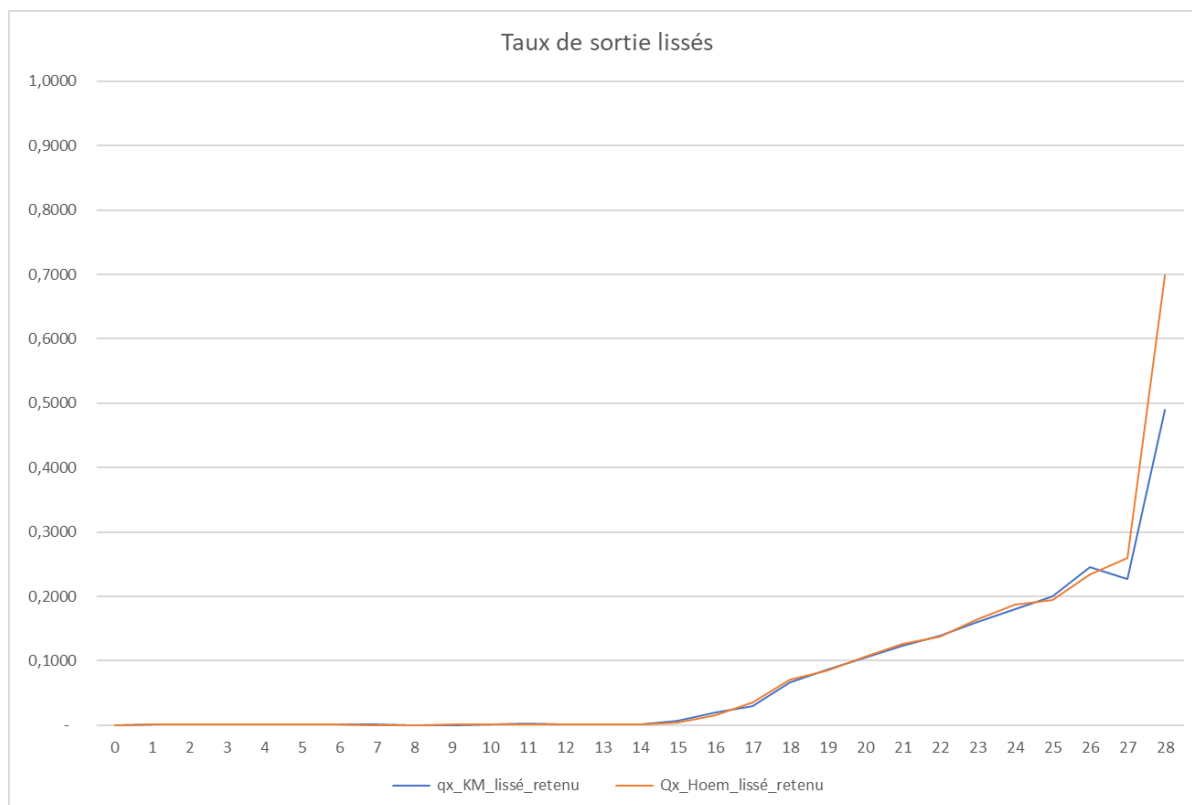


Figure 48. Taux de sortie lissés retenus (estimateur de Hoem et de Kaplan Meier)

L’histogramme de la *figure 49* présente la comparaison de la loi de maintien en rente éducation obtenue par les méthodes de Hoem et de Kaplan Meier. De nouveau, les deux lois semblent être très proches. Les écarts n’apparaissent que vers les âges de 15, 24, 27 et 28 ans : le taux de survie estimé par l’approche de Kaplan Meier est plus bas à 15 ans et est plus important aux âges élevés par rapport à l’approche de Hoem.

L’objectif du présent mémoire est de calculer le montant de provisions nécessaire pour respecter les engagements de la garantie Rente éducation. L’analyse graphique ainsi que l’étude des chiffres obtenus permet de conclure que la loi de Kaplan Meier donne une approche plus prudente avec des taux de sorties plus faibles pour les âges d’intérêt. L’estimateur de Kaplan Meier semble finalement être mieux adaptée pour répondre aux besoins du provisionnement, même si celui de Hoem donne les résultats assez proches et satisfaisants.

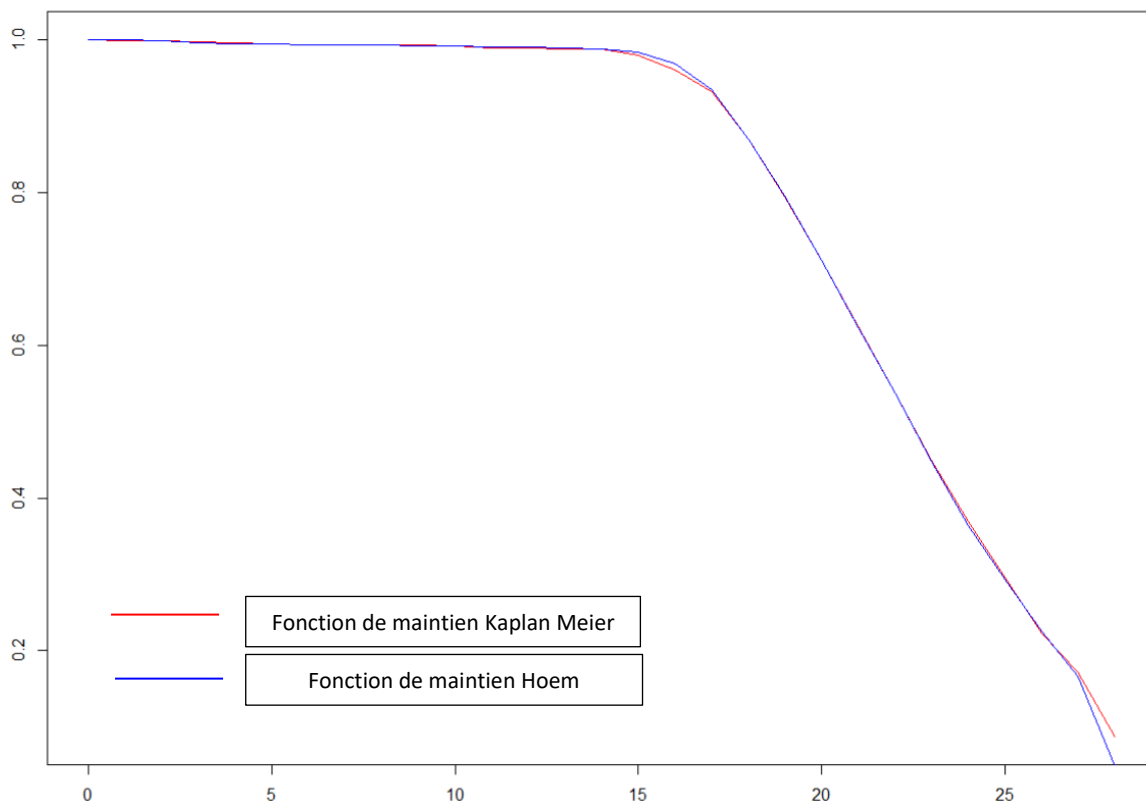


Figure 49. Fonctions de maintien en rente éducation (estimateur de Hoem et de Kaplan Meier)

3.4 Méthode semi-paramétrique : analyse d'hétérogénéité

L'objectif de cette partie du chapitre est de faire appel à une approche alternative pour la modélisation de la loi de sortie des bénéficiaires de la rente éducation. En effet, dans les paragraphes précédents du présent mémoire, des méthodes non paramétriques de l'analyse de durée ont été utilisées. Il est possible par ailleurs de s'intéresser à la méthode qui permet de construire une certaine hiérarchie des sous-populations. Pour le faire, la fonction de maintien peut être modélisée par un des modèles à hasard proportionnel. Il s'agit des approches qui supposent l'existence du hasard de base, qui pour chaque sous-population est multiplié par un certain coefficient représentant la sensibilité au risque de cette sous-population en fonction des variables explicatives. La méthode permet ainsi de classer les sous-populations par rapport à leur niveau de hasard et ceci en présence d'un nombre modeste de données.

Trois types de modèles à hasard proportionnel existent :

1. Les modèles à fonction de hasard de base connue. Il s'agit du positionnement du hasard d'une sous-population par rapport à celle de référence. Il est possible, par exemple, de construire un groupe de référence en utilisant les données disponibles au niveau national. Le défaut de cette méthode est lié à la difficulté à exprimer le hasard instantané de la population de référence.

2. Les modèles à fonction de hasard de base paramétrique. Dans cette approche le hasard de base est supposé suivre une loi paramétrique connue. Le problème de la méthode est qu'il s'avère parfois restrictif de supposer que la fonction de hasard suit une loi paramétrique précise.

3. Les modèles à fonction de hasard de base non spécifiée. Il s'agit de l'approche qui ne cherche pas à spécifier la forme de la fonction de hasard de base en supposant qu'elle n'apporte pas d'utilité particulière à l'analyse de hasard de la population.

Le modèle de Cox qui fait partie du troisième groupe des modèles à hasard proportionnel permet de modéliser la loi de maintien dans le cadre de présence des données censurées. Il peut être utilisé pour construire la loi de maintien en rente éducation.

3.4.1 Modèle de Cox : modèle à risques proportionnels

Cadre théorique

Tout d'abord, la description brève du cadre théorique de l'approche de Cox est présentée.

Une des idées centrales du modèle de Cox est l'absence d'hypothèse particulière sur la forme de la fonction de sortie. La loi de sortie de base est donc considérée comme paramètre de nuisance du modèle. Cette hypothèse est un point fort du modèle car il est effectivement compliqué de supposer l'existence du cadre paramétrique propre dans l'échantillon des données réelles. Le fait de devoir faire une hypothèse stricte sur la forme de la fonction de sortie de base semble être assez restrictif. Dans le modèle de Cox, la spécification de la loi pour la population de référence n'apporte pas de valeur ajoutée au modèle parce que l'utilisateur s'intéresse uniquement à l'effet des covariables. Le principe du modèle de Cox est d'établir un lien entre une ou des variables explicatives et la survenance de l'évènement étudié. L'objectif est de positionner de différents groupes d'observations les uns par rapport aux autres et non pas de modéliser le taux de sortie absolu.

Le principe de l'approche de Cox est en quelque sortes comparable à un modèle de régression classique car on cherche à relier un évènement modélisé à certaines variables explicatives. L'avantage du modèle c'est la possibilité de l'utiliser en présence des données censurées. Pour rappel, dans le cadre du présent mémoire les observations sont censurées s'il s'agit de la rente qui est en cours au moment de la fin de la période d'observation ou bien si la rente est clôturée pour un motif non identifié. C'est donc non pas la durée de la rente réelle de chaque bénéficiaire qui est observée, mais plutôt deux nouvelles variables définies comme suit :

$$X_i = T_i \wedge C_i, \text{ pour } i \in [1, k]$$

Et

$$D_i = \begin{cases} 1, & \text{si } T_i \leq C_i \\ 0, & \text{si } T_i > C_i \end{cases} \text{ pour } i \in [1, k]$$

Avec (T_1, \dots, T_k) un échantillon de durées des rentes éducation et (C_1, \dots, C_k) un deuxième échantillon composé des variables de censures qui est indépendant du premier.

Le modèle de Cox peut être formalisé de la manière suivante :

$$q(t|s; \theta) = \exp(-s'\theta)q_0(t)$$

Avec $q_0(t)$ – taux de sortie de base instantané qui est inconnu, s – vecteur de variables explicatives et θ – vecteur de paramètres qui représentent l'effet des covariables sur le taux de sortie instantané. C'est ce dernier paramètre qui doit être modélisé pour étudier le positionnement des groupes d'observations les uns par rapport aux autres. Dans l'expression ci-dessus le risque de sortie du portefeuille se décompose en deux termes : le premier est le risque de base qui ne dépend que du temps, et le deuxième est celui qui dépend des variables explicatives. Ce deuxième terme est supposé être indépendant du temps ce qui est une des hypothèses principales du modèle de Cox (Planchet F., 2020, Modèles de durée. Introduction – concept de base. Support de cours 2020-2021).

Interprétation du modèle

Soit s_i est une variable qualitative qui prend deux modalités en fonction de l'appartenance du rentier à un groupe défini. Formellement :

$$s_i = \begin{cases} 1, & \text{si rentier } i \text{ appartient au groupe X} \\ 2, & \text{si rentier } i \text{ appartient au groupe Y} \end{cases}$$

Le rapport suivant pour les rentiers numéro 1 et 2 (en supposant que le rentier numéro 1 appartient au groupe X et le rentier numéro 2 au groupe Y) peut être écrit :

$$\frac{q_0(t)\exp(s_2 * \theta)}{q_0(t)\exp(s_1 * \theta)} = \frac{q_0(t)\exp(2 * \theta)}{q_0(t)\exp(1 * \theta)} = \exp(\theta)$$

Le paramètre $\exp(\theta)$ représente le rapport de risque de sortie du portefeuille entre deux groupes de bénéficiaires.

Plusieurs cas de figures sont possibles en fonction de la valeur du paramètre obtenu :

- Si $\exp(\theta) < 1 \rightarrow \theta < 0$, alors le taux de sortie du portefeuille est plus important pour les individus appartenant au groupe X que pour ceux appartenant au groupe Y
- Si $\exp(\theta) = 1 \rightarrow \theta = 0$, alors le risque de sortie du portefeuille est identique pour les deux groupes de rentiers et la variable étudiée n'explique pas la variation du risque

- Si $\exp(\theta) > 1 \rightarrow \theta > 0$, alors le taux de sortie du portefeuille est moins important pour les individus appartenant au groupe X que pour ceux appartenant au groupe Y

3.4.2 Hypothèses du modèle

L'utilisation de l'approche de Cox nécessite de poser quelques hypothèses sur les paramètres du modèle :

- Le ratio entre les risques instantanés des groupes d'observations étudiés est constant, autrement dit indépendant du temps. Le rapport qui doit être calculé n'évolue pas dans le temps. Il s'agit de l'hypothèse des risques proportionnels qui donne son nom à cette famille des modèles semi-paramétriques.
- Le logarithme du taux de sortie absolu est une fonction linéaire du vecteur des variables explicatives. Il s'agit de l'hypothèse de log-linéarité qui s'écrit formellement comme suit :

$$\log(q(t|s; \theta)) = \log(q_0(t)) - s'\theta$$

En complément des hypothèses sur les paramètres du modèle, il faut également les poser sur les données censurées ce qui permettra par la suite de simplifier l'expression de la vraisemblance :

- La censure existante est indépendante de la variable étudiée conditionnellement aux covariables. Formellement, X_i est indépendant de C_i conditionnellement à s
- La censure est non informative, autrement dit la loi de C_i ne dépend pas des paramètres du modèle θ et $q_0(t)$.

Il est important de vérifier l'hypothèse des risques proportionnels pour utiliser le modèle de Cox sur un échantillon de données réelles car il s'agit de l'hypothèse principale de l'approche. Si le risque n'est pas proportionnel, il sera difficile d'interpréter correctement les résultats des tests d'adéquation et de valider le modèle obtenu. L'hypothèse peut être vérifiée à l'aide de plusieurs méthodes existantes.

Si les variables explicatives choisies sont qualitatives avec un nombre faible de modalités, il est possible de faire appel à la méthode log-log. Il s'agit de l'approche qui permet d'analyser la proportionnalité graphiquement.

Il faut construire un modèle de Cox stratifié par rapport à la variable s et représenter sur un même graphique les courbes $\log(-\log(\hat{S}_j(t)))$ pour $j=1,2,\dots$ l'indicateur d'appartenance du rentier à un groupe d'observations. \hat{S}_j définit la fonction de maintien en rente estimée pour le groupe j . Si les courbes semblent être parallèles, alors l'hypothèse de proportionnalité est respectée et le modèle de Cox peut être utilisé pour l'échantillon de

données en question (Despieres E., 2010, Construction et validation de tables d'expérience pour le provisionnement d'un portefeuille de rentes éducation)..

L'inconvénient de cette méthode de vérification est son focus sur la seule analyse graphique qui est liée à une forte subjectivité. De ce fait, le test est conseillé pour l'utilisation uniquement pour un nombre faible de groupes d'observations.

3.4.3 Vraisemblance du modèle de Cox

Avec les notations définies dans les paragraphes précédents, l'expression de la vraisemblance complète s'écrit comme suit :

$$L(\theta, q_0) = \prod_{i=1}^k [q_0(t_i) \exp(-\theta' s_i) \exp(-Q_0(t_i) \exp(-\theta' s_i))]^{D_i} [\exp(-Q_0(t_i) \exp(-\theta' s_i))]^{(1-D_i)}$$

Dans l'approche de Cox, le taux de sortie de base est considéré comme un paramètre de bruit. Cependant, il intervient bien dans la formule de la vraisemblance complète de deux manières : directement et à travers la fonction de sortie cumulée. Si aucune restriction n'est imposée à la fonction de sortie de base, alors le maximum de vraisemblance n'existera pas. Pour résoudre le problème qui s'impose, Cox a proposé dans son modèle de faire appel à une vraisemblance partielle, dans laquelle le taux de sortie de base n'intervient pas. Pour le faire, il faut décomposer la vraisemblance complète afin d'isoler le terme de nuisance dont l'incidence peut être négligée (Dupuy J.F., 2002, Modélisation conjointe de données longitudinales et de durées de vie). Ci-dessous l'expression de la vraisemblance partielle ainsi obtenue :

$$L^{Cox}(\theta) = \prod_{i=1}^k \left[\frac{\exp(-\theta' s_i)}{\sum_{j=1}^n \exp(-\theta' s_j) 1_{\{T_i \leq T_j\}}} \right]^{D_i}$$

Cette expression permet de confirmer que dans le modèle de Cox ce n'est pas le niveau de sortie de base qui est étudié, mais plutôt le positionnement du niveau de hasard pour des groupes d'observations distingués par un ou plusieurs facteurs de risque.

L'estimateur se calcule alors de la manière suivante :

$$\hat{\theta} = \operatorname{argmax}_{\theta} L^{Cox}(\theta)$$

Le taux de hasard de base cumulé s'écrit comme suit :

$$\widehat{Q}_o(t) = \sum_{i=1}^k \frac{D_i 1_{\{T_i \leq t\}}}{\sum_{j=1}^k \exp(-\hat{\theta}' s_j) 1_{\{T_i \leq T_j\}}}$$

Le processus $\widehat{Q}_o(t)$ de risque de base cumulé converge en loi vers un processus centré et dispose d'une variance qui peut être calculée.

(Planchet F., Septembre 2020, Statistique des modèles paramétriques et semi-paramétriques)

3.4.4 Choix des variables et mise en place du modèle

Sexe du bénéficiaire

Dans un premier temps, le modèle de Cox est appliqué sur la base de données d'AXA en partant de l'hypothèse que l'échantillon de données représente une hétérogénéité par rapport à la variable « sexe du bénéficiaire ». En effet, selon le tableau de la *figure 17* (construit à partir des données d'AXA), l'âge moyen à la clôture de la rente est supérieur pour les filles que pour les garçons. De ce fait, il est judicieux de supposer que l'appartenance à la population féminine ou masculine puisse avoir un impact non négligeable sur la durée de maintien en rente. La variable « sexe du bénéficiaire » est donc retenue en tant que variable discriminante pour la construction du modèle de Cox.

Selon la *figure 24*, le nombre de garçons dépasse légèrement le nombre de filles dans l'échantillon de données, c'est donc la fonction de hasard de la population masculine qui est retenue comme celle de hasard de base.

Le modèle à construire s'écrit donc comme suit :

$$q(t|sexe_i; \hat{\theta}) = \exp(-sexe_i \hat{\theta}) q_0(t)$$

Avec $sexe_i = \begin{cases} 1, & \text{si reniter } i \text{ est un homme} \\ 2, & \text{si reniter } i \text{ est une femme} \end{cases}$ et $\hat{\theta}$ le coefficient estimé à l'aide de la vraisemblance partielle.

La modélisation est faite à l'aide du logiciel R et son package Survival (Geffray S., 2009, Analyse des durées de vie avec le logiciel R). Le résultat de l'estimation du coefficient est présenté dans la *figure 50*.

Likelihood ratio test=27.56 on 1 df		p=2e-07	
Wald test=27.52 on 1 df		p=2e-07	
Score (logrank) test=27.57 on 1 df		p=2e-07	
	chisq	df	p
Sexe	84.8	1	<2e-16
GLOBAL	84.8	1	<2e-16

	Coeff	Exp(coef)	se(coef)	z	Pr(> z)
Sexe F	-0.14175	0.86784	0.02702	-5.246	1.55e-07 ***

Figure 50. Résultats de l'estimation du modèle de Cox pour la variable « sexe de bénéficiaire » issus de R

Selon la première table, l'hypothèse nulle de non-significativité globale du modèle peut être rejetée avec moins d'1% de risque d'erreur. Le deuxième tableau indique que le coefficient $\hat{\theta}$ estimé pour le taux de sortie de la population féminine est différent de 0 avec

une probabilité supérieure à 99%. Il s'agit donc de la variable discriminante ayant un impact non négligeable sur le phénomène étudié. Le paramètre $\exp(\hat{\theta}) = 0,86784$ peut être interprété de la manière suivante : le risque de sortie du portefeuille de rente éducation est 13,2% plus faible pour une fille que pour un garçon pour chaque âge, ce qui confirme les hypothèses posées après l'étude des statistiques descriptives de la *figure 8*.

Pour pouvoir appliquer le modèle de Cox pour le provisionnement de la garantie rente éducation, il est important de s'assurer que l'hypothèse forte de proportionnalité du hasard de base est bien respectée. La vérification sera faite de deux manières :

- par la méthode log-log qui permet de comparer les deux courbes $\log(-\log(\hat{S}_j(t)))$ avec la fonction de survie obtenue à l'aide de l'estimateur Kaplan Meier et avec $j=1$ pour les hommes et $j=2$ pour les femmes
- par la comparaison des courbes des taux de sortie obtenues pour la population féminine grâce à l'application de l'approche Kaplan Meier et de celle de Cox

La *figure 51* affiche les courbes log-log des fonctions de maintien obtenues par la méthode Kaplan Meier pour les hommes et les femmes de l'échantillon. Si le hasard de base (celui de la population masculine dans ce cas-là) est proportionnel, alors les deux courbes doivent être parallèles. Selon le graphique, les deux lignes ne semblent pas être parallèles, ce qui permet de déduire que l'hypothèse de proportionnalité du hasard de base n'est pas respectée et le modèle de Cox ne peut pas être utilisé pour l'échantillon de données en question.

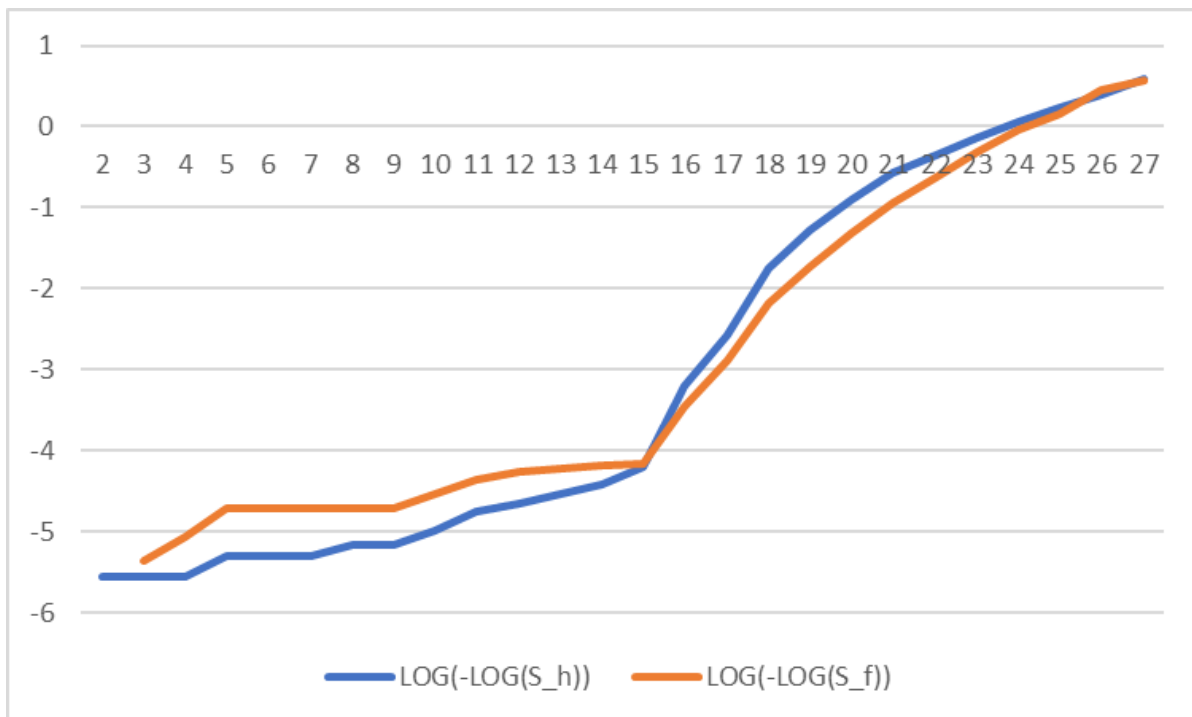


Figure 51. Courbes Log-log des fonctions de maintien (estimateur Kaplan Meier) pour les hommes et les femmes

La *figure 52* fait apparaître les taux de sortie de la rente pour les filles estimés selon la méthode de Kaplan Meier et selon le modèle de Cox construit ci-dessus. Il est facile de voir que les deux lignes ne se comportent pas de manière similaire, ce qui confirme la non-validation de l'hypothèse de proportionnalité du hasard de base.

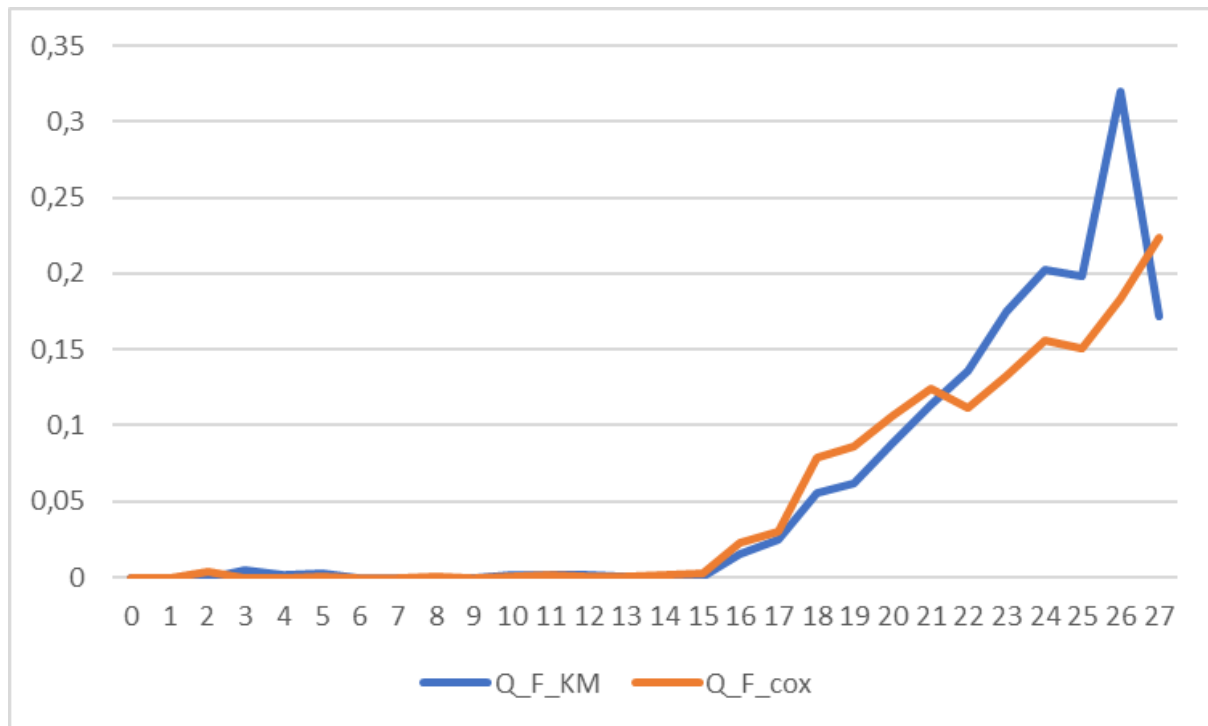


Figure 52. Taux de sortie de la rente éducation pour la population féminine selon Kaplan Meier et Cox

Malgré le fait que l'hypothèse de nullité du coefficient du modèle ait été rejetée, il n'est pas possible d'utiliser le modèle obtenu pour la modélisation de la loi de maintien en rente éducation.

Age d'entrée en rente

Dans un second temps, il est intéressant d'appliquer le modèle de Cox pour une autre variable potentiellement discriminante pour l'estimation de la durée de la rente, notamment l'âge au moment d'effet de la rente. La rente éducation se déclenche au moment du décès de l'assuré, soit d'un parent du rentier. Un tel événement peut potentiellement avoir un impact sur la décision du rentier de poursuivre ou d'arrêter ses études. Il est aussi possible que le fait de bénéficier de la rente pousse le rentier à poursuivre ses études pendant plus longtemps. De ce fait, la variable « âge au moment d'effet de la rente » semble être importante pour sa durée. Il s'agit de la variable continue mais il est possible de la transformer en une variable qualitative en créant des groupes d'âge et en ventilant les rentiers entre ses groupes.

Pour la construction du modèle de Cox, trois groupes d'âge au moment d'effet de la rente ont été créés :

Modalité	Description	Nombre de rentiers
Groupe 1	0 – 14 ans	6.751
Groupe 2	14 – 18 ans	3.235
Groupe 3	18 – 27 ans	3.486

Les groupes sont construits à partir de la distribution des rentiers en fonction de l'âge au moment d'effet de la rente qui apparaît sur la *figure 15*. L'âge de 14 ans correspond généralement à la fin de la période d'adolescence, et l'âge de 18 ans à celui où les individus obtiennent leur baccalauréat et prennent une décision concernant la poursuite d'études. Le *groupe 1* contient le plus d'effectifs, de ce fait il est retenu comme celui de référence pour la construction du taux de sortie de base.

Le modèle de Cox s'écrit donc comme suit :

$$q(t) = \left(\exp(-\text{groupe_âge}_2 * \hat{\theta}_2) + \exp(-\text{groupe_âge}_3 * \hat{\theta}_3) \right) * q_0(t)$$

Avec groupe_âge_i – variable indicatrice de l'appartenance de l'individu au groupe d'âge i et $\hat{\theta}_2$ et $\hat{\theta}_3$ les coefficients estimés à l'aide de la vraisemblance partielle de Cox.

La *figure 53* affiche les résultats des tests de significativité globale du modèle ainsi que les coefficients estimés pour chaque groupe d'âge.

Likelihood ratio test=71.8 on 2 df		p=3e-16	
Wald test=70.24 on 2 df		p=6e-16	
Score (logrank) test=70.59 on 2 df		p=5e-16	
	chisq	df	p
Groupe	9.17	2	0.01
GLOBAL	9.17	2	0.01

	Coef	Exp(coef)	se(coef)	z	Pr(> z)
Groupe 2	-0.07478	0.92794	0.03212	-2.328	0.0199 *
Groupe 3	-0.29273	0.74623	0.03517	-8.322	<2e-16 ***

Figure 53. Résultats de l'estimation du modèle de Cox pour la variable « groupe d'âge au moment d'effet de la rente » issus de R

Selon le premier tableau, l'hypothèse de la significativité globale du modèle peut être acceptée au seuil de 1%. Les tests de rapport de vraisemblance, de Wald et de Score confirment également la significativité globale du modèle. Le deuxième tableau fait apparaître les coefficients estimés pour les deux groupes d'âge ainsi que le résultat du test de leur non-nullité. L'hypothèse nulle est rejetée pour le coefficient estimé pour *le groupe 2* au seuil de 2% et pour *le groupe 3* au seuil inférieur à 1%. L'interprétation de l'exponentielle de chaque coefficient permet de dire que le fait d'appartenir au *groupe 2* réduit la probabilité de sortir de la rente de 7,2%. L'appartenance au *groupe 3* réduit la chance de sortir de la rente de 25,4%. La définition des groupes permet de conclure qu'une rente prenant effet plus tardivement incite le rentier à rester en étude pendant plus longtemps afin de bénéficier d'un

financement supplémentaire. Cette tendance est plus prononcée pour la tranche d'âge à partir de 18 ans.

Même si l'hypothèse de nullité des coefficients du modèle est rejetée, il est important de vérifier l'hypothèse fondamentale de proportionnalité du hasard de base. De nouveau, deux méthodes de validation décrites dans les paragraphes précédents sont utilisées.

La *figure 54* affiche les courbes Log-log des fonctions de survie selon l'estimateur de Kaplan Meier pour les trois groupes d'âge. Pour l'intervalle jusqu'aux 25 ans les courbes des *groupes 1* et *2* semblent être parallèles mais la tendance disparaît à partir de l'âge de 25 ans. La courbe du *groupe 3* n'est parallèle avec aucune autre courbe. L'hypothèse de proportionnalité de hasard de base semble ne pas être respectée.

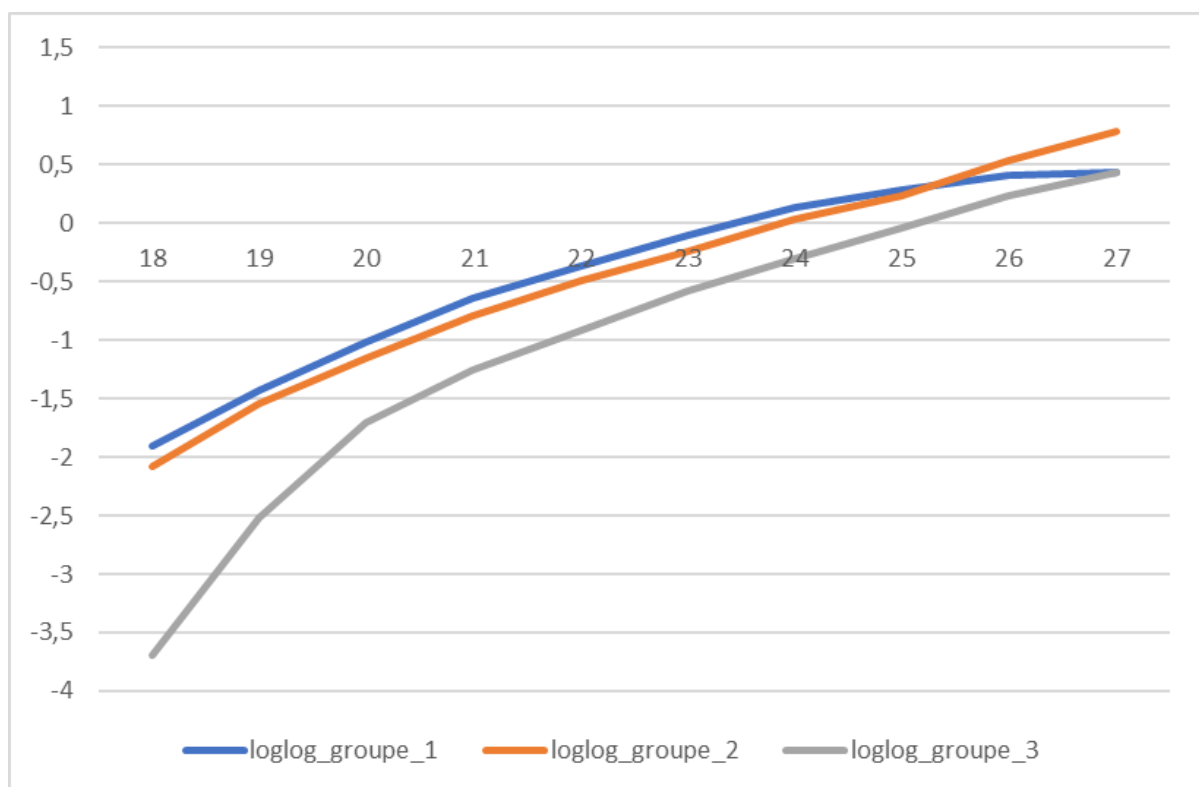


Figure 54. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 3 groupes d'âge au moment d'effet de la rente

Pour s'assurer que la proportionnalité de hasard de base n'est pas validée, les courbes des taux de sortie estimées par Kaplan Meier et par le modèle de Cox sont tracées pour les *groupes 2* et *3*. Les *figures 55* et *56* affichent les courbes pour les *groupes 2* et *3* respectivement. Il est évident que dans les deux cas, les lignes n'ont pas la même allure. Le modèle de Cox n'est donc pas applicable dans ce contexte à cause de non-validation de l'hypothèse fondamentale de la proportionnalité du hasard de base.

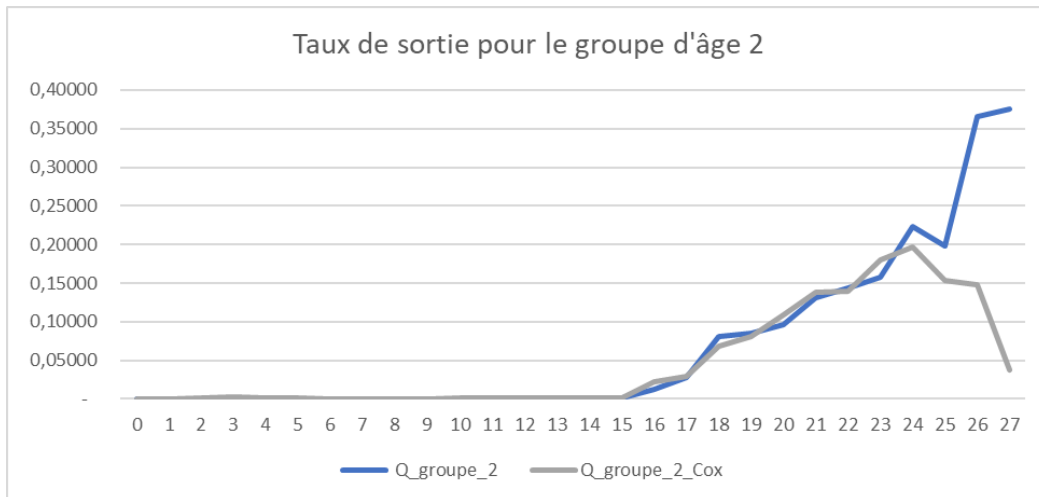


Figure 55. Taux de sortie de la rente éducation pour le groupe d'âge 2 selon Kaplan Meier et Cox

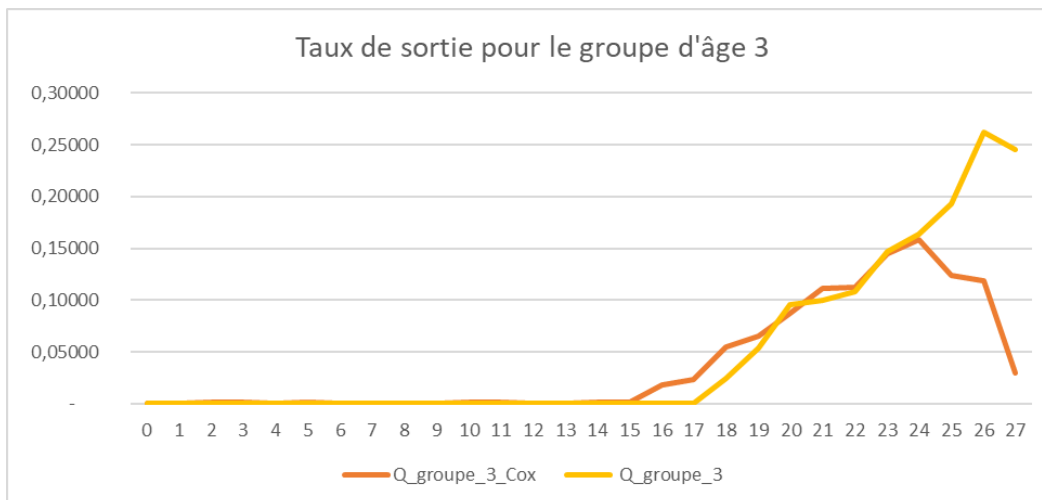


Figure 56. Taux de sortie de la rente éducation pour le groupe d'âge 3 selon Kaplan Meier et Cox

Modèle combiné

A présent le modèle de Cox a été appliqué aux données du portefeuille en se basant sur l'idée qu'une seule variable détermine l'appartenance du rentier à un groupe de risque. Dans la partie suivante, les bénéficiaires de la rente seront répartis par groupes caractérisés par deux facteurs discriminants simultanément, soit le sexe et l'âge au moment d'effet de la rente. En effet, selon les résultats du modèle de Cox appliqué pour un seul facteur de risque, il est judicieux de supposer que le sexe et l'âge d'entrée en rente ont un impact sur le risque de sortir du portefeuille. Il s'avère intéressant à vérifier, si la répartition des rentiers par groupes caractérisés par la combinaison de deux facteurs assure la validation de l'hypothèse de proportionnalité du hasard de base.

Les rentiers du portefeuille d'AXA sont alors repartis en 6 groupes dont la description et le nombre d'effectifs sont présentés dans le tableau de la figure 57.

Critères	Nombre de rentiers	Nom de groupe combiné
F_groupe d'âge d'effet 1	3 244	Gr_comb_2
F_groupe d'âge d'effet 2	1 596	Gr_comb_4
F_groupe d'âge d'effet 3	1 771	Gr_comb_6
M_groupe d'âge d'effet 1	3 507	Gr_comb_1
M_groupe d'âge d'effet 2	1 639	Gr_comb_3
M_groupe d'âge d'effet 3	1 715	Gr_comb_5
	13 472	

Figure 57. Répartition des rentiers par groupe en fonction du sexe et de l'âge au moment d'effet de la rente

Ainsi, le groupe incluant les garçons du groupe d'âge d'effet 1 (soit inférieur à 14 ans) semble être le plus nombreux, de ce fait il est retenu en tant que population de référence dont le hasard est considéré comme celui de base.

Les résultats de l'application du modèle de Cox sont présentés dans les tableaux de la figure 58. Selon les tests du ratio de vraisemblance, de Wald et de score, l'hypothèse de nullité des coefficients estimés peut être rejetée au seuil de moins de 1%. En revanche, selon le test de Wald appliqué pour chaque coefficient estimé individuellement, le coefficient correspondant au *groupe 3* n'est pas significativement différent de 0 au seuil de 5% (en effet l'hypothèse nulle ne peut être rejetée qu'au seuil de 10%). La conclusion est telle que le comportement du *groupe 3* vis-à-vis au risque de sortie du portefeuille n'est pas différent de celui du groupe de référence. Les autres coefficients sont significatifs, ce qui montre que le taux de hasard varie en fonction de l'appartenance du rentier à un groupe. Par exemple, le fait d'appartenir au *groupe 2* diminue le risque de sortie du portefeuille de rente éducation de 19,3%. La définition des groupes permet de conclure que parmi les rentes qui prennent effet au moment où le bénéficiaire a entre 0 et 14 ans, la probabilité qu'un rentier de sexe féminin sorte du portefeuille est inférieure de 19,3% par rapport à son homologue masculin.

Likelihood ratio test=116.6 on 5 df			p=<2e-16
Wald test=117.6 on 5 df			p=<2e-16
Score (logrank) test=118.4 on 5 df			p=<2e-16
	chisq	df	p
Groupes combinés	79.1	5	1.3e-15
GLOBAL	79.1	5	1.3e-15

	Coeff	Exp(coef)	se(coef)	z	Pr(> z)
Groupe combiné 2	-0.21546	0.80617	0.04102	-5.253	1.50e-07
Groupe combiné 3	-0.07876	0.92427	0.04379	-1.798	0.0721
Groupe combiné 4	-0.27859	0.75685	0.04532	-6.148	7.86e-10
Groupe combiné 5	-0.41841	0.65809	0.04831	-8.661	< 2e-16
Groupe combiné 6	-0.37756	0.68553	0.04729	-7.985	1.41e-15

Figure 58. Résultats de l'estimation du modèle de Cox pour les variables « sexe » et « groupe d'âge au moment d'effet de la rente » issus de R

Comme déjà précisé, le modèle de Cox ne peut être utilisé que si l'hypothèse de proportionnalité de hasard de base est validée. Le graphique de la *figure 59* montre les courbes $\log(-\log(\hat{S}_j(t)))$ construites pour l'estimateur de survie de Kaplan Meier pour chaque groupe d'individus étudié. Visiblement, les courbes ne sont pas parallèles entre elles, car elles se croisent toutes à un moment donné. Ceci indique que l'hypothèse de proportionnalité de hasard de base est violée.

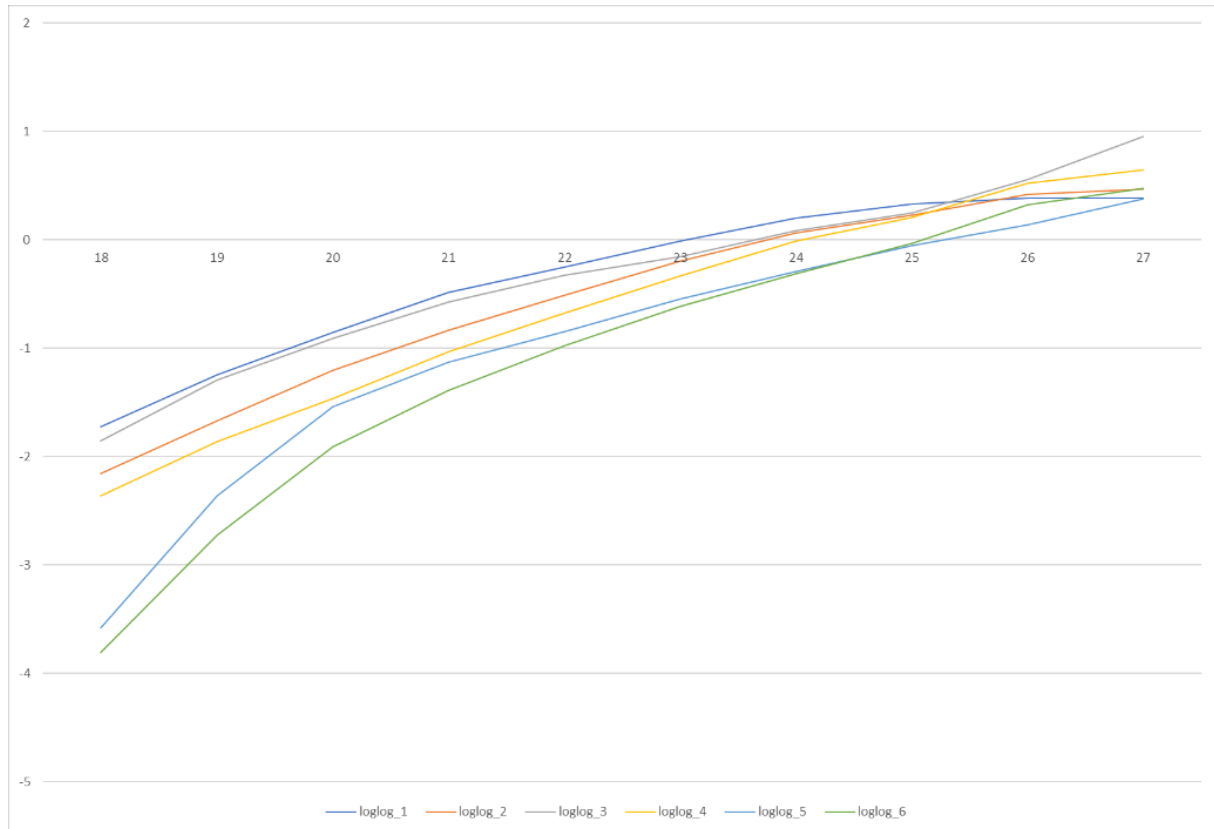


Figure 59. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 6 groupes d'âge au moment d'effet de la rente

Pour confirmer la non-proportionnalité du hasard de base, les courbes de hasard construites à partir des coefficients de Cox estimés ci-dessus sont comparées avec celles obtenues grâce à l'approche de Kaplan Meier. Elles sont présentées sur les graphiques de la *figure 60* pour chaque groupe individuellement. Les courbes n'ont pas la même allure pour tous les groupes de rentiers. Des écarts importants apparaissent surtout pour la tranche d'âge entre 24 et 27 ans, qui est l'âge proche de la fin contractuelle de la rente. Il s'agit aussi de l'intervalle d'âge avec le nombre réduit d'effectifs, ce qui rend l'application du modèle de Cox plus complexe, surtout en présence du découpage de la population en groupes.

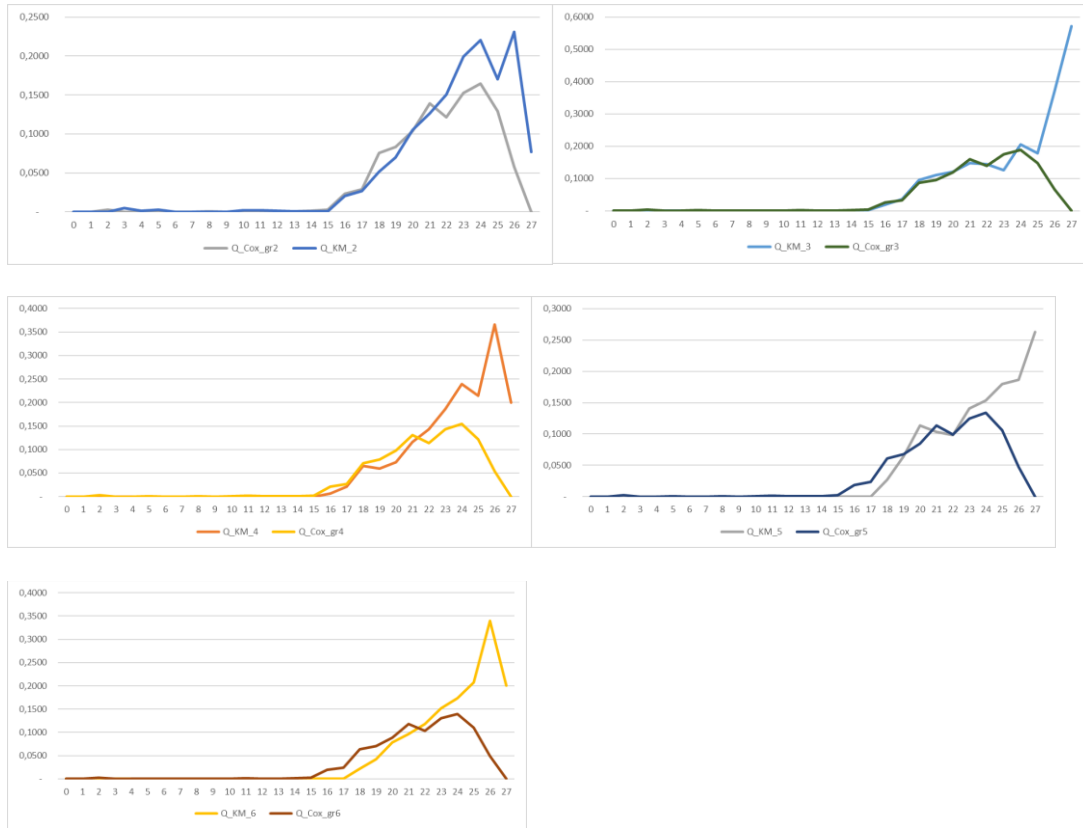


Figure 60. Taux de sortie de la rente éducation pour les groupes combinés « sexe/âge au moment d'effet » selon Kaplan Meier et Cox

Il est cependant intéressant d'appliquer le modèle de Cox dans la même configuration mais en réduisant le nombre de catégories de rentiers. En effet, selon les données de la *figure 58*, le hasard du *groupe 3* n'est pas significativement différent de celui du groupe de référence, alors ces deux catégories peuvent n'en former qu'une. Ainsi, le tableau de la *figure 61* représente une nouvelle ventilation des rentiers par groupes en fonction des variables « sexe » et « groupe d'âge au moment d'effet de rente ».

Critères	Nombre de Identifiant	Nom de groupe combiné
F_groupe d'âge d'effet 1	3 244	Gr_comb_2
F_groupe d'âge d'effet 2	1 596	Gr_comb_4
F_groupe d'âge d'effet 3	1 771	Gr_comb_6
M_groupe d'âge d'effet 1+2	5 146	Gr_comb_1
M_groupe d'âge d'effet 3	1 715	Gr_comb_5
	13 472	

Figure 61. Répartition des rentiers par groupe en fonction du sexe et de l'âge au moment d'effet de la rente

Selon les chiffres de la *figure 62*, les tests de ratio de vraisemblance, de Wald et de score permettent de rejeter l'hypothèse de nullité globale des coefficients du modèle au seuil de moins d'1%, ce qui confirme la significativité des variables sélectionnées et valide le choix

des groupes d'observations. Le test de Wald individuel pour chaque coefficient du modèle permet de conclure que tous les coefficients sont différents de 0 avec une certitude supérieure à 99%. Tous les coefficients estimés sont inférieurs à 0, alors leurs exponentielles sont inférieures à 1, ce qui indique que le risque de sortir du portefeuille est le plus élevé pour le groupe de référence (soit pour des garçons ayant entre 0 et 18 ans au moment d'effet de la rente). Par exemple, pour le *groupe 6*, le risque de clôture de la rente est 29,15% plus faible par rapport au *groupe 1*. Autrement dit, pour les rentiers garçons, si la rente prend effet entre les 18 et 27 ans du bénéficiaire, alors le risque de sortie est 29,15% moins fort par rapport au cas où la rente prend effet entre 0 et 18 ans du bénéficiaire.

Likelihood ratio test=113.3 on 4 df		p=<2e-16	
Wald test=113.9 on 4 df		p=<2e-16	
Score (logrank) test=114.7 on 4 df		p=<2e-16	
	chisq	df	p
Groupe combiné	80.2	4	<2e-16
GLOBAL	80.2	4	<2e-16

	Coeff	Exp(coef)	se(coef)	z	Pr(> z)
Groupe combiné 2	-0.18346	0.83238	0.03715	-4.938	7.90e-07
Groupe combiné 4	-0.24615	0.78180	0.04175	-5.896	3.72e-09
Groupe combiné 5	-0.38542	0.68017	0.04485	-8.593	< 2e-16
Groupe combiné 6	-0.34456	0.70853	0.04375	-7.876	3.38e-15

Figure 62. Résultats de l'estimation du modèle de Cox pour les variables « sexe » et « groupe d'âge au moment d'effet de la rente » avec le nombre de catégories réduit issu de R

De nouveau, il est nécessaire de s'assurer que l'hypothèse de proportionnalité de hasard de base soit vérifiée. La *figure 63* représente les courbes Log-log construites pour les 5 groupes des rentiers. La courbe de couleur bleu clair correspondante au *groupe 5* semble se croiser à un moment avec chacune des autres courbes. En revanche, les couples de courbes 1 avec 2 et 4 avec 6 semblent ne jamais se croiser sur l'intervalle étudié. Il s'agit alors de prétendants potentiels sur la proportionnalité de hasard. Cependant, la distance entre les courbes 1 et 2 ainsi qu'entre les courbes 4 et 6 ne paraît pas être constante durant tout l'intervalle. Il n'est donc pas évident de faire une conclusion sur la proportionnalité de risque avec seulement le test Log-log. Il faut faire appel au deuxième contrôle.

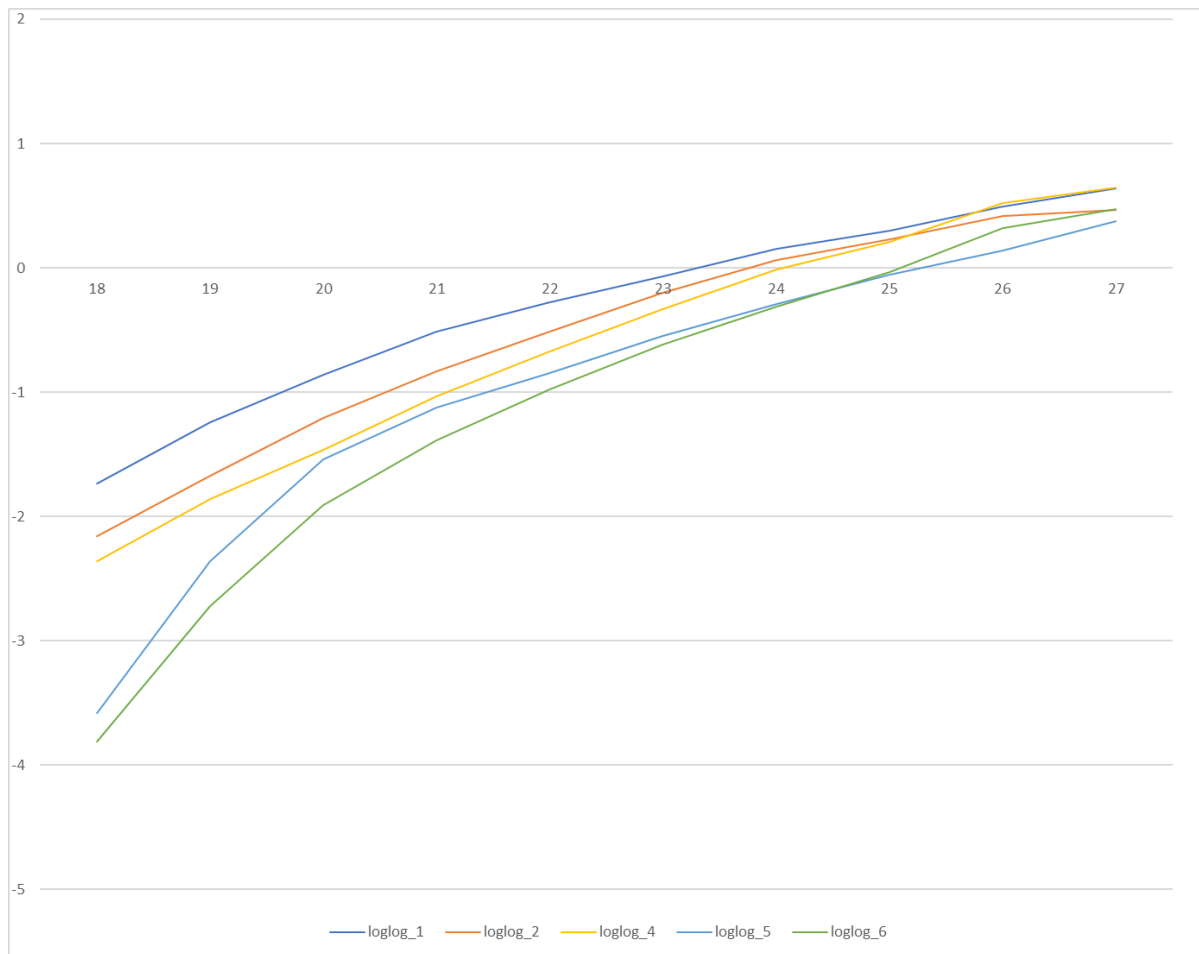


Figure 63. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 5 groupes d'âge au moment d'effet de la rente

Les graphiques de la *figure 64* affichent les courbes des taux de sortie construites par les modèles de Kaplan Meier et de Cox pour chaque groupe de rentiers à l'exception de celui de référence. Pour tous les groupes les taux de sortie semblent être très différents entre les deux méthodes, la différence est surtout importante pour l'intervalle d'âge entre 21 et 27 ans. En effet, le risque de sortie modélisé par le modèle de Cox ne capte pas le pic de clôtures de rente correspondant à l'âge de 26 ans qui est l'âge-limite contractuel pour la majorité des contrats du portefeuille. Même si le provisionnement de la garantie en assurance doit respecter le principe de prudence, il est tout de même important de ne pas surestimer le risque et de ne pas constituer un stock de provisions trop important. De ce fait, il est judicieux de ne pas utiliser le modèle de Cox pour la construction de la loi de maintien en rente éducation.

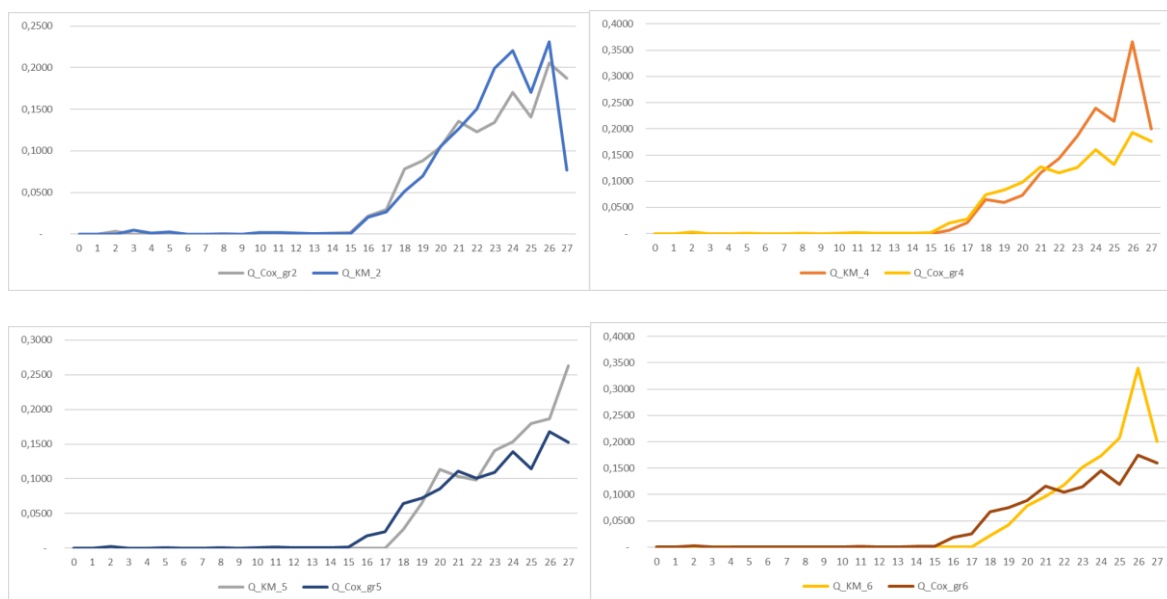


Figure 64. Taux de sortie de la rente éducation pour de nouveaux groupes combinés « sexe/âge au moment d'effet » selon Kaplan Meier et Cox

3.5 Loi de maintien segmentée

Dans le paragraphe précédant l'approche semi-paramétrique de Cox a été appliquée aux données issues du portefeuille d'AXA dans le but de construire le modèle permettant de tenir compte d'existence des groupes de risque homogènes à l'intérieur de l'échantillon de rentiers. Malgré le fait que la loi obtenue ne peut pas être retenue à cause de la non-validation de l'hypothèse fondamentale du hasard de base, elle permet de confirmer que le sexe du bénéficiaire a un impact non négligeable sur la durée de la rente. Il est donc judicieux de tenir compte de ce facteur discriminant dans la loi de maintien.

Afin de prendre en compte l'impact du sexe d'un bénéficiaire sur sa durée de rente, il est possible de construire un modèle segmenté, ou autrement dit, de construire deux modèles distincts en séparant l'échantillon initial en 2 sous-populations. Le critère de construction des sous-populations est le facteur discriminant, le sexe du rentier. La modélisation de la loi peut se faire à l'aide d'une méthode non paramétrique qui a déjà été utilisée dans les paragraphes précédents. Selon le chapitre 3.3.3 l'estimateur de Kaplan-Meier semble être bien adapté pour le calcul de la loi de survie pour l'échantillon global, de ce fait il peut de nouveau être utilisé pour la construction des lois distinctes.

En s'appuyant sur les chiffres affichés dans la *figure 24*, deux échantillons de données sont construits à partir de celui initial : l'échantillon de 6.861 observations pour la population masculine et de 6.611 pour celle féminine.

Ainsi, la *figure 65* affiche les taux de sortie bruts pour les populations féminine et masculine obtenus à l'aide de l'estimateur Kaplan Meier. Ils sont comparés entre eux et avec les taux retenus dans le paragraphe 3.3.3 pour la totalité de l'échantillon.

Les taux sont très proches pour les âges inférieurs à 16 ans. Sur l'intervalle entre 15 et 22 ans les taux de sortie de la population masculine sont plus élevés que ceux de la population féminine, en revanche cette tendance s'inverse sur l'intervalle entre 22 et 26 ans. Les taux lissés construits pour la totalité de l'échantillon se situent généralement entre les taux des deux sous-populations distinctes, ce qui est logique étant donné que l'échantillon global est une fusion des deux sous-populations. En tenant compte de la taille réduite de deux sous-échantillons utilisés pour la construction de la loi segmentée, il semble inutile de lisser les taux bruts du modèle segmenté. En effet, les pics de sortie observés pour les deux groupes de rentiers pour certains âges ont une signification importante pour l'analyse de la loi de sortie. Le fait de lisser les taux bruts segmentés crée le risque de rendre la courbe de taux moins fidèles aux données d'origine, et ce risque devient plus important avec la diminution de la taille de l'échantillon. Il est donc décidé de garder les taux de sortie bruts pour le modèle segmenté.

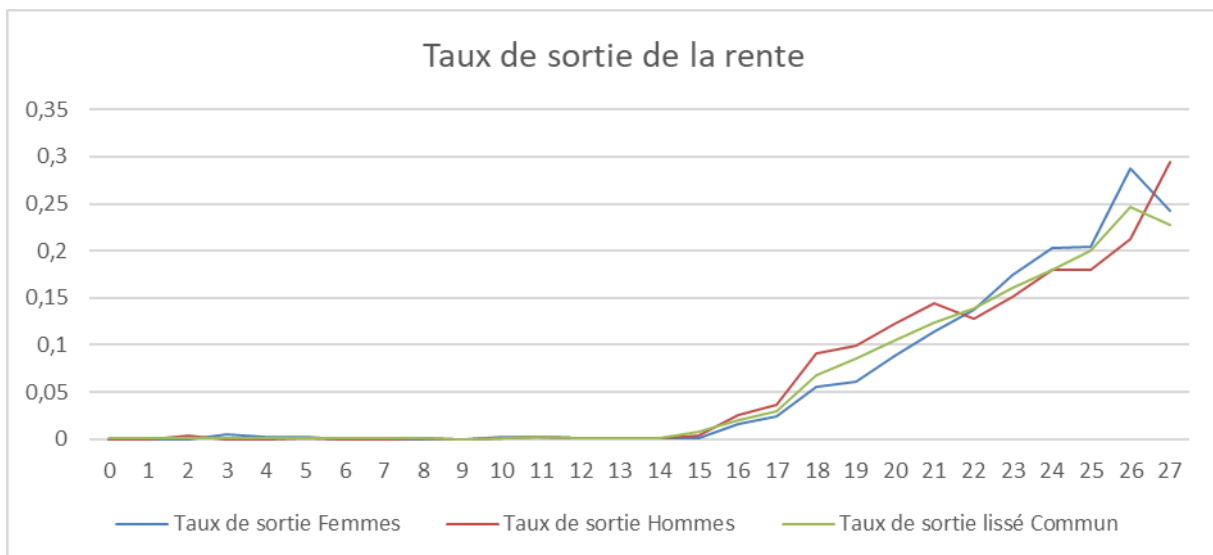


Figure 65. Taux de sortie bruts pour les hommes et les femmes séparément et taux de sortie lissés pour la totalité de l'échantillon

La figure 66 fait apparaître les lois de maintien en rente pour les hommes et les femmes du portefeuille. Les deux courbes sont proches pour les âges inférieurs à 16 ans. Sur l'intervalle entre 16 et 27 ans le taux de maintien en rente de la population féminine est supérieur à celui de la population masculine ce qui confirme la conclusion faite à l'aide du modèle de Cox. Le fait de prendre en compte les différences comportementales de deux groupes d'observations permet de trouver un équilibre entre le provisionnement prudent et réaliste. Il est donc judicieux de retenir le modèle segmenté par sexe de bénéficiaire pour construire la table de maintien en rente pour le besoin des travaux de provisionnement.

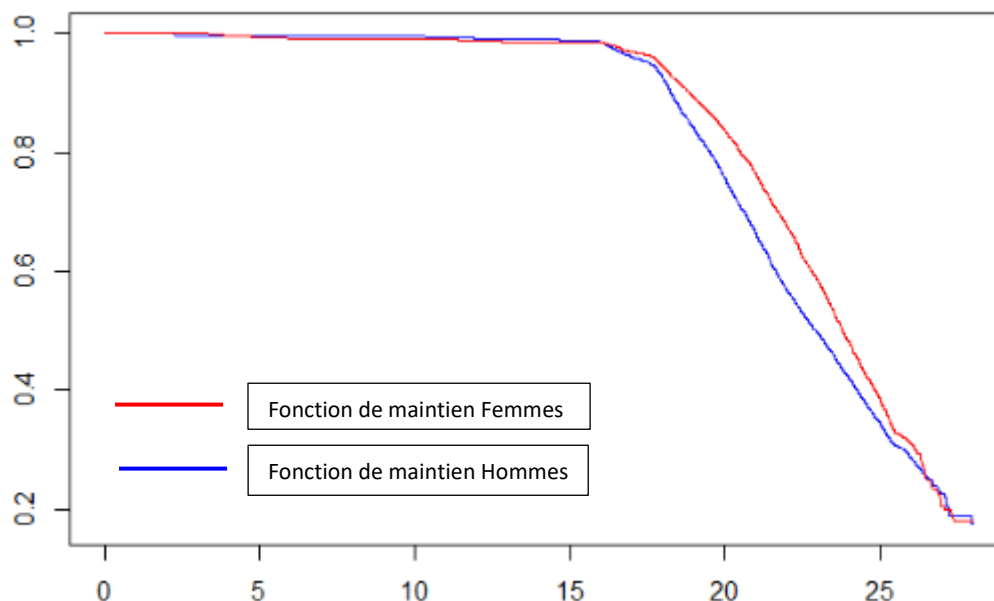


Figure 66. Lois de maintien en Rente pour les hommes et les femmes de l'échantillon

Pour terminer la partie sur la construction de la loi de maintien d'expérience, il est intéressant de vérifier, dans quelle mesure la loi de maintien construite à partir des données d'AXA diffère de celle fictive construite à partir des données de l'INSEE. La courbe de maintien construite dans le paragraphe 1.3.2 est commune pour la totalité de la population nationale. Pour comparer la loi retenue dans l'étude avec la loi de maintien nationale fictive, il est nécessaire de la segmenter en fonction du sexe de bénéficiaire. La *figure 67* fait apparaître les courbes de maintien d'expérience par sexe construites à l'aide de l'estimateur Kaplan Meier et les courbes fictives par sexe caractérisant la population française. L'analyse du graphique met en évidence des différences significatives entre les taux de maintien de la population nationale par rapport à ceux des rentiers d'AXA. Les 4 courbes sont assez proches pour la population âgée de moins de 15 ans, ce qui se justifie par l'obligation de scolarisation jusqu'à l'âge de 16 ans et par un faible taux de mortalité à ces âges. En revanche, à partir des 16 ans le taux de maintien en rente obtenu avec les données nationales pour les deux populations chute rapidement pour tomber quasiment à 0 vers 24 ans. Enfin, selon les chiffres de l'INSEE quasiment la totalité de l'effectif sortirait de la population entrant dans le champ d'application de la garantie rente éducation vers 24 ans. Ce n'est pourtant pas le cas pour la population du portefeuille d'AXA. L'étude des courbes permet de conclure que les rentiers du portefeuille ont un comportement significativement différent par rapport à celui de la population française. Il s'avère alors que les tables disponibles au niveau national ne peuvent pas répondre aux exigences de l'assureur en matière de prudence.

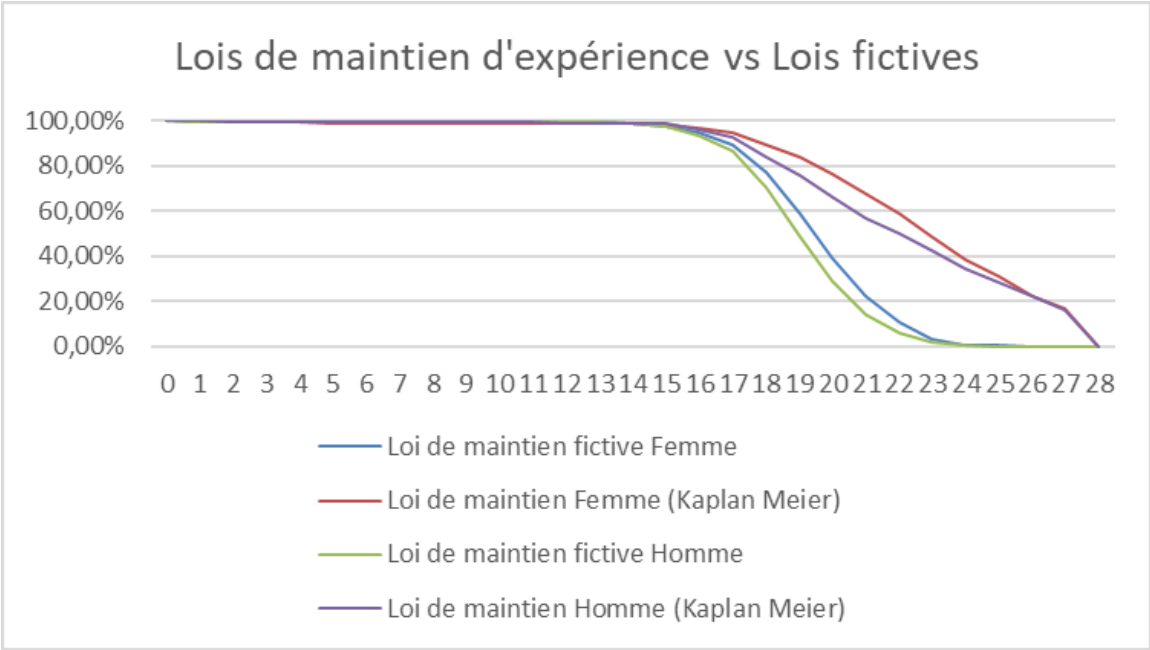


Figure 67. Lois de maintien en rente : fictive à partir des données nationales et l'estimateur de Kaplan Meier à partir des données du portefeuille d'AXA

Chapitre 4. Validation de la loi de maintien d'expérience et son impact sur le provisionnement

Dans le chapitre précédent, la loi de maintien en rente éducation a été construite à partir des données du portefeuille d'AXA France. Pour valider la table obtenue, il est important de vérifier sa capacité prédictive.

Avant de procéder au contrôle et à la validation des résultats, il faut comprendre quel est l'objectif de modélisation. S'il s'agit de la construction de la table pour le provisionnement, alors il est préférable de rester prudent et de favoriser l'approche qui minimise le risque de sous-provisionnement. Si la table d'expérience est construite dans le cadre de l'application de la Solvabilité II, notamment pour le calibrage du modèle interne, alors la loi doit être la plus réaliste possible. Le présent mémoire a pour but de construire une table de maintien pour estimer les provisions de la rente éducation, il convient alors de se montrer prudent tout en restant réaliste.

La validation de la table d'expérience s'effectue à travers deux angles : la comparaison avec la loi actuellement utilisée et son rapprochement avec les données réellement observées.

4.1 Validation des résultats

La validation du résultat peut être faite via plusieurs approches. Dans cette étude, 2 approches sont utilisées : le calcul du nombre de sorties et l'étude de l'indicateur *SMR*.

La qualité de la nouvelle loi peut aussi être mesurée à travers de sa comparaison avec la loi qui est actuellement utilisée au sein de l'équipe d'AXA pour le calcul des provisions de la garantie. Afin de rendre la comparaison possible il est nécessaire de s'assurer que les deux lois sont construites à partir de l'échantillon de données identique. L'approche qui a permis de construire l'ancienne loi de maintien en rente a été reprise et appliquée à l'échantillon de données utilisé dans le présent mémoire. Cette démarche permet d'actualiser l'ancienne loi avec de nouvelles données tout en respectant la méthodologie utilisée.

4.1.1 Nombre de sorties

Dans un premier temps il est important d'évaluer la capacité du modèle construit de prédire le nombre de sorties du portefeuille de la rente. Le nombre de rentes clôturées estimé à l'âge n du bénéficiaire de sexe i se calcule de la manière suivante :

$$N_{n,i}^{estimé} = \hat{q}_{n,i}^{KM} * E_{n,i}$$

Le nombre total de rentes clôturées est égal à la somme des clôtures pour la population féminine et masculine.

Les nombres de clôtures estimés par le nouveau et l'ancien modèles sont comparés avec le nombre de clôture réellement constaté dans la base de données d'AXA durant la période entre 01/01/2009 et 31/12/2019. La *figure 68* fait apparaître cette comparaison pour

la tranche d'âge allant de 15 à 28 ans pour la totalité de la population. L'intervalle avant 15 ans n'est pas affiché à cause du nombre très faible de sorties constaté et estimé.

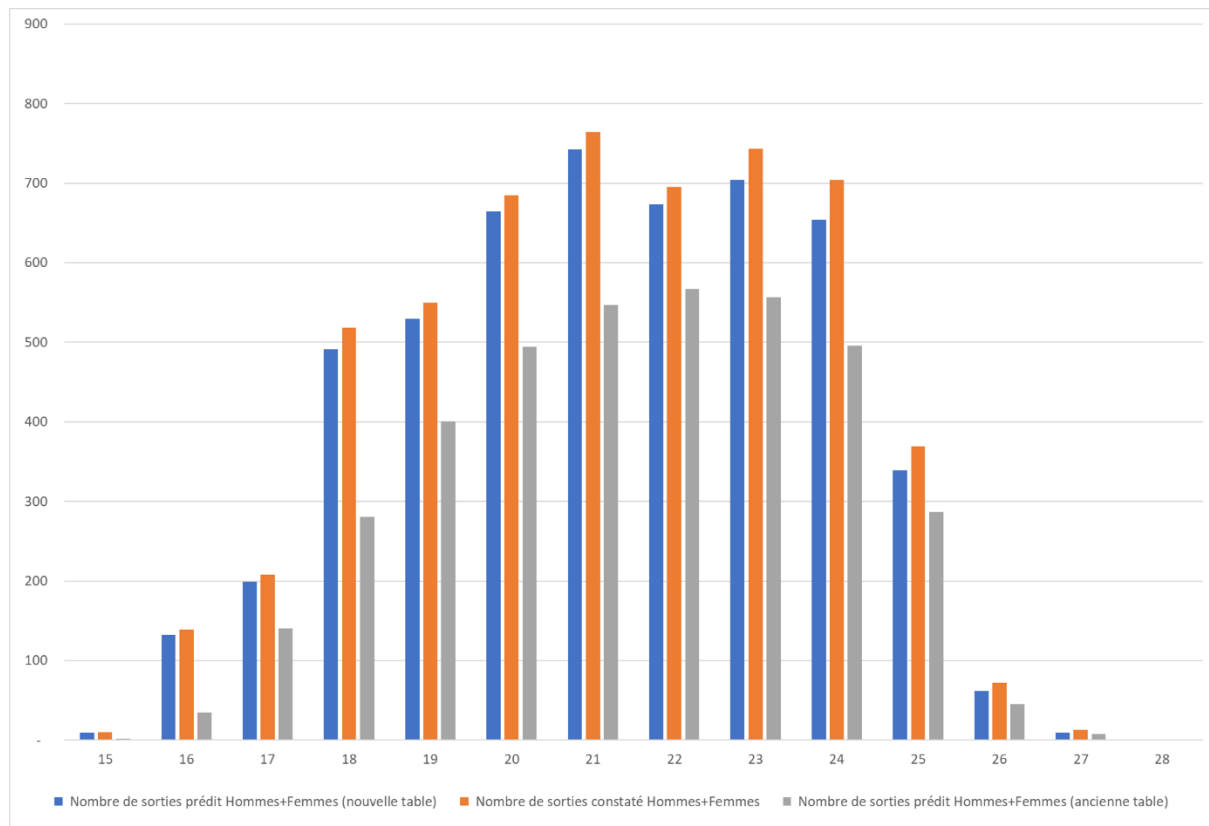


Figure 68. Nombre de sorties de la rente éducation réel et estimé par la nouvelle et l'ancienne loi de maintien par âge pour la tranche d'âge entre 15 et 28 ans

La nouvelle table de maintien par sexe, dont la construction est détaillée dans le chapitre précédent, semble prédire le nombre de sorties de manière assez précise. Cependant, pour tous les âges, le nombre de clôtures de rentes estimé par cette table reste inférieur à celui réellement constaté. Ce constat permet de conclure que la nouvelle table reste prudente tout en étant plus précise que l'ancienne. En effet, tous les âges de clôture à l'exception des 26 et 27 ans, le nombre de sorties prédit par l'ancienne table est inférieur à celui réellement constaté.

4.1.2 Calcul de SMR

Tout d'abord la notion de *SMR (Standardizes Mortality Ratio)* doit être explicitée. Il s'agit du quotient entre le nombre d'évènements réellement observé dans l'échantillon de données et le nombre d'évènements attendu pour une population de référence calculé à l'aide de la loi construite (Langmeier R., 2000, Étude de différentes méthodes d'ajustement de tables de mortalité : application aux données d'une compagnie d'assurance).

Soit :

- S_n^{obs} – nombre de sorties du portefeuille de la rente éducation réellement observé pour l'âge n

- S_n^{est} – nombre de sorties du portefeuille de la rente éducation estimé par le modèle construit pour l'âge n
- E_n – exposition au risque de l'âge n
- \hat{q}_n – taux de sortie de la rente éducation de l'âge n , estimé et lissé dans les paragraphes précédents
- $SMR_n = \frac{S_n^{obs}}{S_n^{est}} = \frac{S_n^{obs}}{E_n * \hat{q}_n}$

Le *SMR* peut être interprété de la façon suivante :

- Si le *SMR* est supérieur à 100%, alors le nombre d'évènements réellement observé est plus important que celui prédit par le modèle. Il s'agit donc d'une sous-estimation du taux de sortie du portefeuille de rente éducation.
- Si le *SMR* est égal à 100%, alors le nombre d'évènements réellement observé correspond au nombre d'évènements attendus et la loi de sortie réelle est parfaitement représentée par le modèle construit.
- Si le *SMR* est inférieur à 100%, alors le nombre d'évènements réellement observé est plus faible que celui prédit par le modèle. Il s'agit donc d'une surestimation du taux de sortie du portefeuille.

Le *SMR* est souvent calculé sur des périodes de quelques ans. Dans le cas de la loi de sortie de la rente éducation, les *SMR* cumulé pour tous les âges ainsi que ceux pour des intervalles de 5 ans seront étudiés. Pour avoir le provisionnement suffisamment prudent il est indispensable d'avoir les *SMR* supérieurs à 100%. En effet, si les taux de sortie du portefeuille sont surestimés, cela veut dire que les rentiers restent en études pendant plus longtemps que ce qui est prévu par le modèle. Le stock de provisions constitué pour la rente éducation n'est donc pas suffisant pour couvrir les engagements. L'assureur se retrouve donc dans une situation de sous-provisionnement. La situation avec les *SMR* inférieurs à 100% est une alerte et nécessite de revoir l'approche de calcul de provisions.

Pour construire les *SMR* détaillés, la base de données a été construite à partir des données d'AXA pour les 10 dernières années de survénances de rente (depuis 2009). La base de données obtenue a été retraitée selon la manière décrite dans la partie 2.2 du présent mémoire. Des expositions par âge de clôture ont été calculées pour chaque survénance pour les hommes et les femmes de l'échantillon séparément. Le *SMR* total, cumulant tous les âges de clôture ainsi que les *SMR* pour les intervalles de 5 ans ont été déterminés par la suite. Les résultats des calculs pour la loi de maintien d'expérience segmentée par sexe sont affichés dans le tableau des *figures 69 et 70*.

La *figure 69* fait apparaître les *SMR* pour la population féminine. Le *SMR* cumulé est supérieur à 100% pour toutes les survénances, ce qui montre que la table d'expérience est suffisamment prudente et ne surestime pas le nombre de sorties réellement observé.

Les SMR ventilés par intervalles de 5 ans donnent un résultat satisfaisant pour toutes les survénances : quasiment tous les SMR calculés sont supérieurs à 100%. Quelques exceptions sont observées pour certaines tranches d'âge pour les survénances 2009, 2010, 2013, 2015 et 2016. Néanmoins, les SMR trop faibles se compensent par les SMR des tranches d'âge voisines. Les ratios pour la survénance 2017 sont en revanche plus inquiétants car ils sont inférieurs à 100% pour plusieurs intervalles. Le SMR global de l'année 2017 est supérieur à 100% ce qui permet de conclure que la table reste prudente pour l'intégralité de l'échantillon.

Femmes / Loi Segmentée			2 009	2 010	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018	2 019
SMR cumulé [0 ans;28 ans]			110%	110%	109%	111%	108%	110%	108%	109%	104%	111%	112%
SMR par intervalle													
16 ans	20 ans	[16 ans;20 ans]	112%	134%	101%	113%	98%	104%	118%	104%	95%	100%	134%
17 ans	21 ans	[17 ans;21 ans]	122%	113%	109%	118%	100%	109%	108%	104%	91%	104%	123%
18 ans	22 ans	[18 ans;22 ans]	111%	107%	106%	112%	102%	110%	119%	95%	83%	107%	112%
19 ans	23 ans	[19 ans;23 ans]	105%	100%	107%	102%	102%	108%	99%	103%	84%	106%	115%
20 ans	24 ans	[20 ans;24 ans]	111%	97%	106%	100%	100%	108%	109%	106%	87%	111%	106%
21 ans	25 ans	[21 ans;25 ans]	128%	100%	112%	103%	105%	106%	123%	119%	102%	109%	100%
22 ans	26 ans	[22 ans;26 ans]	99%	104%	107%	106%	113%	109%	105%	105%	109%	113%	105%
23 ans	27 ans	[23 ans;27 ans]	102%	102%	112%	106%	110%	105%	102%	113%	120%	112%	109%
24 ans	28 ans	[24 ans;28 ans]	100%	104%	113%	114%	114%	108%	110%	108%	123%	111%	105%

Figure 69. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survénance pour les femmes

La figure 70 reprend les SMR pour la population masculine de l'échantillon. De nouveau, les SMR cumulés sont supérieurs à 100% pour toutes les survénances. Les SMR par intervalle sont majoritairement supérieurs au seuil de prudence malgré le fait que certaines tranches d'âge font apparaître les SMR par intervalle inférieur à 100%. La conclusion générale est telle que la loi segmentée permet de garder l'approche prudente pour la population masculine du portefeuille.

Hommes / Loi Segmentée			2 009	2 010	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018	2 019
SMR cumulé [0 ans;28 ans]			115%	115%	110%	114%	120%	109%	112%	114%	113%	116%	120%
SMR par intervalle													
16 ans	20 ans	[16 ans;20 ans]	109%	106%	99%	116%	106%	122%	104%	117%	121%	114%	
17 ans	21 ans	[17 ans;21 ans]	113%	114%	98%	115%	110%	101%	115%	113%	112%	107%	
18 ans	22 ans	[18 ans;22 ans]	113%	117%	93%	115%	113%	97%	108%	109%	101%	94%	
19 ans	23 ans	[19 ans;23 ans]	112%	108%	99%	110%	120%	98%	95%	112%	99%	93%	
20 ans	24 ans	[20 ans;24 ans]	125%	117%	104%	119%	115%	106%	107%	117%	96%	96%	
21 ans	25 ans	[21 ans;25 ans]	114%	118%	135%	144%	125%	119%	132%	143%	115%	121%	
22 ans	26 ans	[22 ans;26 ans]	111%	113%	119%	103%	129%	115%	109%	113%	112%	114%	
23 ans	27 ans	[23 ans;27 ans]	110%	106%	126%	102%	127%	122%	111%	114%	126%	123%	
24 ans	28 ans	[24 ans;28 ans]	116%	110%	124%	101%	126%	126%	122%	111%	137%	136%	

Figure 70. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survénance pour les hommes

Pour apprécier la qualité de la table de maintien segmentée proposée dans cette étude, il est intéressant de comparer sa performance avec celle de l'ancienne loi réactualisée avec de nouvelles données. Etant donné que l'ancienne table de maintien a été construite pour la population totale sans prise en compte du sexe de bénéficiaire, le SMR global a aussi été calculé pour le nouveau modèle : les nombres de sorties réelles et de sorties estimées pour les sous-populations féminine et masculine ont été sommés. Le résultat de calcul apparaît dans la figure 71. Les SMR pour l'ancien modèle ont directement été calculés pour l'intégralité de la population et sont affichés dans la figure 72. Les SMR cumulés sont supérieurs à 100% pour toutes les années de survénances pour les deux modèles. Il s'avère que les deux

approches sont prudentes et ne créent pas le risque de sous-provisionnement à la suite d'une mauvaise estimation du nombre de sorties. En revanche, la comparaison des deux tables montre que l'ancienne loi est plus prudente que la loi segmentée, avec des SMR plus élevés. Ce constat montre que la nouvelle loi prédit le nombre de clôtures de rente de la manière plus fidèle.

Hommes+Femmes			2 009	2 010	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018	2 019
SMR cumulé [0 ans;28 ans]			113%	112%	109%	113%	114%	109%	110%	112%	109%	113%	116%
SMR par intervalle													
16 ans	20 ans	[16 ans;20 ans]	110%	117%	100%	115%	103%	100%	120%	104%	108%	112%	123%
17 ans	21 ans	[17 ans;21 ans]	118%	114%	103%	116%	106%	104%	112%	109%	103%	106%	114%
18 ans	22 ans	[18 ans;22 ans]	112%	113%	99%	114%	108%	102%	113%	103%	93%	100%	107%
19 ans	23 ans	[19 ans;23 ans]	109%	104%	103%	106%	111%	103%	97%	108%	92%	100%	116%
20 ans	24 ans	[20 ans;24 ans]	117%	107%	105%	109%	107%	107%	108%	112%	92%	104%	114%
21 ans	25 ans	[21 ans;25 ans]	121%	108%	121%	119%	115%	112%	127%	129%	108%	114%	121%
22 ans	26 ans	[22 ans;26 ans]	104%	108%	112%	104%	120%	112%	107%	108%	111%	113%	116%
23 ans	27 ans	[23 ans;27 ans]	105%	104%	118%	104%	118%	112%	106%	113%	122%	116%	119%
24 ans	28 ans	[24 ans;28 ans]	106%	107%	117%	108%	120%	115%	115%	109%	130%	120%	113%

Figure 71. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survénance – Cumul pour l'intégralité de la population

Survénance			2 009	2 010	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018	2 019
SMR cumulé [0 ans;28 ans]			143%	136%	129%	135%	128%	129%	132%	139%	117%	130%	144%
SMR par intervalle													
16 ans	20 ans	[16 ans;20 ans]	131%	127%	108%	120%	105%	102%	126%	118%	107%	114%	134%
17 ans	21 ans	[17 ans;21 ans]	135%	121%	110%	120%	109%	106%	117%	121%	104%	109%	122%
18 ans	22 ans	[18 ans;22 ans]	124%	117%	103%	115%	109%	103%	112%	108%	92%	101%	108%
19 ans	23 ans	[19 ans;23 ans]	119%	107%	106%	105%	111%	103%	96%	112%	91%	100%	116%
20 ans	24 ans	[20 ans;24 ans]	127%	108%	110%	111%	109%	110%	109%	113%	93%	105%	115%
21 ans	25 ans	[21 ans;25 ans]	142%	139%	139%	135%	133%	140%	135%	137%	114%	133%	143%
22 ans	26 ans	[22 ans;26 ans]	144%	144%	143%	138%	143%	144%	141%	147%	126%	140%	160%
23 ans	27 ans	[23 ans;27 ans]	159%	151%	163%	148%	148%	154%	151%	166%	145%	151%	177%
24 ans	28 ans	[24 ans;28 ans]	181%	175%	179%	173%	160%	172%	180%	177%	159%	168%	184%

Figure 72. Les indicateurs SMR de l'ancienne loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans, par survénance pour l'intégralité de la population

4.2 Impact du changement de la loi de sortie sur les provisions

L'objectif de la construction de la nouvelle loi de sortie des bénéficiaires de la rente éducation du portefeuille d'AXA est d'analyser le niveau de provisionnement pour cette garantie.

4.2.1 Calcul des provisions

Dans un premier temps, il convient de présenter la méthode de calcul des provisions pour la rente éducation utilisée au sein de la branche Santé et Prévoyance Collectives d'AXA France.

Les provisions techniques sont par définition égales à la valeur actuelle probable des rentes à verser par l'assureur aux bénéficiaires de la rente éducation.

Tout d'abord quelques notations utiles pour la suite sont à définir :

- T^{esc} – taux d'escompte en vigueur (base annuelle)
- Nb – fractionnement de l'année (par exemple $Nb=4$ signifie que la rente est versée à chaque trimestre)
- t – terme de la rente :

$$t = \begin{cases} 1, & \text{si rente à terme échue} \\ 0, & \text{si rente à terme à échoir} \end{cases}$$

- $T^a = 3\%$ – taux d’arrérage qui correspond aux frais exprimés en pourcentage des prestations versées
- L_n – nombre de bénéficiaires de rente éducation d’âge n

Le taux d’escompte qui apparaît dans la formule ci-dessus est par définition un taux d’intérêt des emprunts à très court terme. Il est préconisé de retenir pour le travail de provisionnement le taux d’escompte réglementaire tel qu’il est défini par l’article A 132-1 du code des assurances et fixé à 60% du TME (taux moyen d’emprunt d’Etat) en vigueur à la date d’effet de chaque élément de constitution de la rente.

Soit un rentier d’âge n . La probabilité que cette rente soit encore en cours j années plus tard s’écrit comme suit :

$$p_{n,j} = \frac{L_{n+j}}{L_n}$$

Le calcul de cette probabilité s’effectue à partir de la table de maintien en rente dont la construction fait objet du présent mémoire. Comme évoqué dans les paragraphes précédents, le provisionnement actuel est réalisé à partir de la table de mortalité TPG93 jusqu’à l’âge de 16 ans et à partir de la table de maintien d’expérience actuelle jusqu’à l’âge de 28 ans.

Pour la suite de l’étude deux types de rente qui existent dans le portefeuille d’AXA sont définis : rente constante et rente par paliers.

Rente constante

Il s’agit de la rente dont le montant ne dépend pas de l’âge du rentier. Elle est servie jusqu’au dernier jour de la période des 26 ans du bénéficiaire (ou 28 ans dans certains cas). Pour rappel, il est possible d’avoir les rentes éducation viagères mais ces rentes ne sont pas étudiées dans le cadre de ce mémoire.

La valeur actuelle probable d’une rente temporaire constante du montant de 1€ versée au rentier de l’âge entier n et de l’âge terme de la rente k se calcule comme suit :

$$V_{RE}(n, k) = \sum_{j=1}^{k-n+t-1} p_{n,j} * \frac{1}{(1 + T^{esc})^{j+t-1}}$$

On introduit aussi la formule de calcul du correctif du fractionnement d’une rente temporaire qui se calcule comme suit :

$$F(n, k) = (-1)^t * \frac{1 - Nb}{2 * Nb} * (1 - E_{n,k-n})$$

Avec :

$E_{n,k-n}$ – engagement de verser 1€ de rente au bénéficiaire d'âge entier n si sa rente est toujours en cours ($k-n$) années plus tard

Selon les conditions du contrat souscrit, la dernière échéance peut être versée le dernier jour de la période précédant ou suivant la sortie du bénéficiaire. Dans ce cas-là, un prorata de fin est appliqué pour calculer la provision. Il se calcule de deux manières en fonction de cas de figure :

1. Si le versement se fait le dernier jour de la période suivant la date de sortie du rentier :

$$P_{fin}(n, k) = \frac{1}{Nb} * \frac{1}{(1 + T^{esc})^{\frac{1}{Nb} + \frac{1}{2}}} * [1 - T^{esc} * \left(\sum_{j=1}^{k-n} \tilde{p}_{n,j} * \frac{1}{(1 + T^{esc})^j} \right) - \tilde{p}_{n,k-n} * \frac{1}{(1 + T^{esc})^{k-n}}$$

Avec $\tilde{p}_{n,k-n}$ déterminé à l'aide de la table de mortalité TV88-90.

2. Si le versement se fait le dernier jour de la période précédant la date de sortie du rentier :

$$P_{fin}(n, k) = 0$$

La provision ultime calculée pour une rente temporaire constante du montant de 1€ se calcule comme suit (pour le cas de la rente servie jusqu'au dernier jour de la dernière période de validité de rente) :

$$Prov_{RE}(n, k) = (V_{RE}(n, k) + F(n, k) + P_{fin}(n, k)) * (1 + T^a)$$

Dans le cas où l'âge \ddot{n} du bénéficiaire n'est pas un entier au moment de calcul, les provisions sont une interpolation des âges entiers voisins :

$$Prov_{RE}(\ddot{n}, k) = (1 - \alpha) * Prov_{RE}(n, k) + \alpha * Prov_{RE}((n + 1), k)$$

Avec

$$n = Entier(\ddot{n})$$

$\alpha = \ddot{n} - n$, soit la fraction de l'année entre l'âge exact du rentier et l'âge entier précédent.

Rente par paliers

Il s'agit de la rente éducation dont le montant dépend de l'âge du bénéficiaire. Le capital versé est donc fonction de l'âge du rentier. Le calcul du montant de la provision ultime dans ce cas-là peut toujours être ramené à la somme des provisions calculée à l'aide de la méthode de calcul pour la rente constante.

Soit une rente de 3 paliers qui garantit un taux pour chaque catégorie d'âge :

- $r_1\%$ du salaire de base de l'assuré pour l'intervalle d'âge $[0, n_1]$ du rentier
- $r_2\%$ du salaire de base de l'assuré pour l'intervalle d'âge $(n_1, n_2]$ du rentier
- $r_3\%$ du salaire de base de l'assuré pour l'intervalle d'âge $(n_2, n_3]$ du rentier

La provision ultime pour la rente éducation par paliers pour un rentier de l'âge n et de salaire de base de 1€ se calcule comme suit :

$$\begin{aligned}
 Prov_{RE,paliers}(n, k) &= r_1 * 1_{n \leq n_1} Prov_{RE}(n, n_1) + r_2 \\
 &* 1_{n_1 < n \leq n_2} \left(Prov_{RE}(max(n_1, n), n_2) * (\tilde{p}_{n, n_1 - n} * \frac{1}{(1 + T^{esc})^{n_1 - n}})^{1_{n \leq n_1}} \right) \\
 &+ r_3 * 1_{n \leq k} \left(Prov_{RE}(max(n_2, n), k) * (\tilde{p}_{n, n_2 - n} * \frac{1}{(1 + T^{esc})^{n_2 - n}})^{1_{n \leq n_2}} \right)
 \end{aligned}$$

4.2.2 Comparaison des résultats

Pour analyser l'impact que le changement de la loi aurait sur le montant des provisions, ce dernier est calculé à l'aide de l'ancienne et de la nouvelle table d'expérience. La comparaison des montants ainsi obtenus permettra de répondre à la question si la nouvelle loi permet d'augmenter ou de baisser le stock de provisions pour la garantie rente éducation.

Tout d'abord, une extraction des rentes en cours au 31/12/2020 est faite pour constituer la base de données de contrôle. Le calcul des engagements est réalisé pour les 10 dernières survenances, année 2020 comprise. Sachant que l'objectif est de comparer les montants de provisions calculés pour les rentes déjà connues, et non pas d'étudier le nombre d'entrées et de sorties du portefeuille, les rentes déclarées tardivement ne posent pas de problème. L'année 2020 peut donc être incluse dans l'échantillon de test.

Le calcul des engagements est réalisé uniquement pour les rentes temporaires. De plus, la base de données de contrôle est nettoyée des rentes erronées : celles dont les bénéficiaires ont plus de 28 ans au moment de l'extraction. En effet, selon les conditions contractuelles, une rente éducation temporaire prend obligatoirement son terme au 28^e anniversaire du bénéficiaire. De ce fait, les provisions calculées pour les rentiers âgés de plus de 28 ans doivent être nulles.

La *figure 73* affiche les rapports entre les montants des provisions calculés pour la garantie rente éducation avec la nouvelle loi segmentée par sexe et l'ancienne loi actualisée avec les données récentes. Les ratios sont affichés par année de survenance. La tendance est quasiment stable sur toutes les années : si la nouvelle table d'expérience était mise en place, le stock de provisions devrait être diminué d'environ 6%. Globalement la provision devrait être revue à la baisse plus importante pour la population masculine que pour celle féminine. Cette observation semble être cohérente par rapport au constat sur le maintien en étude plus long pour les femmes par rapport aux hommes. Le fait de segmenter la population par sexe

du bénéficiaire permet de ne pas pénaliser les hommes de l'échantillon et de proposer le montant de provision plus faible par rapport au modèle construit pour l'intégralité de rentiers.

Provision nouvelle / Provision ancienne (en %)	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018	2 019	2 020	TOTAL
Ratio Femme	92,2%	94,2%	93,1%	95,0%	93,9%	94,4%	93,4%	93,9%	94,2%	93,5%	93,9%
Ratio Homme	95,2%	94,1%	93,2%	93,8%	92,9%	93,0%	93,2%	93,9%	94,0%	93,9%	93,6%

Figure 73. Rapport des montants des engagements de la rente éducation calculés par la nouvelle et l'ancienne tables par sexe du bénéficiaire

4.2.3 Limites de l'approche et axes d'amélioration

La construction de la table d'expérience est un processus assez complexe incluant plusieurs étapes dont la mise en place est associée à la validation de nombreuses hypothèses. Elle doit donc être accompagnée par les réflexions sur les forces et les faiblesses de la méthodologie appliquée. Cette partie du chapitre a pour objectif de mettre en évidence les limites de l'approche utilisée afin de mieux évaluer la qualité du modèle obtenu.

La construction de la loi d'expérience réalisée dans le présent mémoire a été effectuée en présence d'un échantillon de la taille réduite, ce qui fait apparaître des risques de nature différente.

Le risque opérationnel est lié à l'extraction des données et à leur qualité. En effet, lors de l'étape de la construction de l'échantillon pour la modélisation de la loi, plusieurs incohérences et anomalies de données peuvent être détectées. Naturellement, plus la taille de l'entreprise est importante, plus la chance de rencontrer des imperfections dans ses bases techniques est élevée. Cela est dû à l'augmentation de la complexité des outils de gestion utilisés. Malgré la mise en place de l'étape de traitement de données, le risque de présence de données erronées subsiste et peut créer les biais d'estimation dans la table construite.

Le risque systématique d'estimation est lié à l'estimation des paramètres des lois utilisées en présence des fluctuations de l'échantillon de données.

Le risque d'avis d'expert est lié au mauvais jugement qui peut apparaître dans le cas de l'utilisation de l'information externe. En effet, la construction de la loi de maintien nécessite de faire des choix lors de différentes étapes : en commençant par la décision de supprimer ou de garder certaines observations dans la base de données et en finissant par la validation de la table obtenue pour l'usage souhaité. Un mauvais choix fait à une étape peut entraîner une vision biaisée de la durée de la rente et créer une sous- ou surestimation des engagements. De ce fait, il est important de tester plusieurs approches et configurations afin de les comparer et de repérer les incohérences et les alertes.

Le risque de modèle est lié au choix de l'approche retenue. De nombreux modèles existent aujourd'hui pour la modélisation des variables de durée. Le fait de bien choisir l'approche joue un rôle primordial dans la construction de la loi d'expérience (Planchet F., Kamega A., 2013, Construction de tables de mortalité prospectives sur un groupe restreint : mesure du risque d'estimation).

Après avoir mentionné des risques généraux qui peuvent être rencontrés lors du travail de la construction des tables d'expérience, il convient de revenir au sujet de la loi construite dans le présent mémoire et de citer quelques problèmes apparus lors de sa modélisation. Premièrement, il faut souligner que le travail a été fait en présence du petit nombre d'observations avec de nombreuses données censurées. La taille de l'échantillon est surtout faible pour les tranches d'âge s'approchant de l'âge-limite de la garantie. De plus la décision de construire un modèle segmenté nécessite de diminuer la taille de la base de données davantage. C'est un sacrifice qui a été fait dans le but de gagner en capacité prédictive en prenant en compte l'impact du sexe du bénéficiaire sur sa durée de rente. Les données sont donc sujettes à des fluctuations assez importantes, surtout vers la queue droite de la distribution. De plus, la profondeur historique de l'échantillon est assez importante. Même si les données nationales font supposer que le comportement de la population étudiée n'a pas connu de changements significatifs en 30 ans, le risque d'estimation reste tout de même présent et non négligeable.

Deuxièmement, la loi calculée s'arrête à l'âge de la limite contractuelle de la garantie, soit 28 ans. Malgré les conditions imposées par le contrat de rente éducation, la base technique contient plusieurs rentiers dont l'âge dépasse la limite. Le fait de forcer la clôture de la table à 28 ans crée le risque de sous-provisionnement pour ces rentiers même s'il s'agit à priori d'erreurs de saisie dans l'outil de gestion. Idéalement, toutes les observations en question devraient être soumises à un contrôle avec le soutien des gestionnaires. A l'heure actuelle, ces rentiers sont simplement écartés de l'étude.

Troisièmement, le traitement de la base de données qui a été effectué s'appuie principalement sur des jugements des experts (dont les techniciens et les gestionnaires de l'outil). Cette approche est naturellement associée à un risque d'erreur qui peut créer des biais d'estimation et fait apparaître le risque opérationnel.

Par exemple, dans la partie 3.1.3 du présent mémoire, les observations de l'échantillon ont été classées selon le motif de clôture renseigné dans l'outil de gestion. Les motifs « *Justificatif absent* » et « *Fin théorique* » ont été analysés avec le soutien d'un gestionnaire de l'outil. Les rentes clôturées pour ces motifs ont été affectées par la suite aux groupes de sorties réelles ou d'observations censurées. Si cette analyse supplémentaire n'avait pas été faite, alors toutes les rentes en question auraient été reconnues comme de vraies sorties. Afin de comprendre l'impact que le principe de détection des observations censurées a sur le modèle, il est intéressant d'effectuer un test en construisant un nouveau modèle fictif à partir des données traitées de la manière différente. Pour ceci l'estimateur de Kaplan Meier a été utilisé une nouvelle fois sur l'intégralité de la base de données, sans segmentation par sexe. En effet, pour visualiser l'impact du traitement de données sur le résultat final il est judicieux de garder l'échantillon de la taille plus importante plutôt que de prendre en compte le facteur discriminant qui provoque la hausse des fluctuations de l'échantillonnage. La nouveauté du modèle de test est liée au traitement des rentes vis-à-vis de la censure : toutes les rentes

clôturées pour les motifs « *Justificatif absent* » et « *Fin théorique* » sont considérées comme des sorties non censurées. Le graphique de la *figure 74* fait apparaître la courbe des taux bruts de sortie construite dans la partie 3.2.2 et celle obtenue à partir de l'échantillon modifié. Les courbes montrent que les taux bruts sont différents pour les deux modèles : sur l'intervalle allant de 15 à 24 ans, les courbes se croisent plusieurs fois, ce qui indique que pour certains âges, le taux de sortie est plus élevé dans le modèle retenu alors qu'il l'est pour d'autres âges dans le modèle de test. En revanche, à partir de l'âge de 24 ans, les taux du modèle retenu sont nettement plus bas par rapport à ceux du modèle construit à partir de l'échantillon modifié. Il s'avère alors que la façon de traiter les données d'origine a un impact non négligeable sur la loi construite.

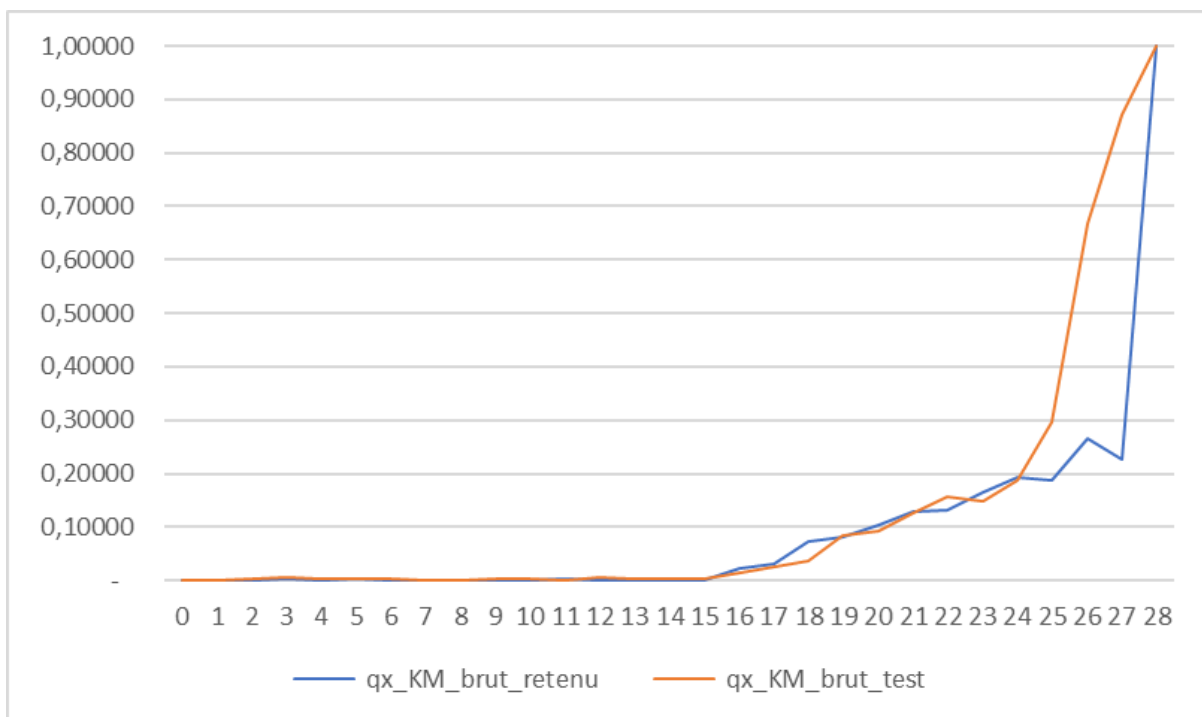


Figure 74. Taux bruts de sortie de Kaplan Meier du modèle retenu et du modèle avec des censures traitées différemment

En conclusion, il faut dire que même si la table d'expérience segmentée semble proposer une estimation assez fiable du nombre de sorties du portefeuille de la rente éducation, elle doit être considérée avec un regard critique. La comparaison de cette table avec celle construite en 2014 et actualisée avec les données récentes permet de conclure que la nouvelle loi est moins prudente que l'ancienne. En effet, le stock des provisions pour la rente éducation aurait été diminué de 6% en moyenne si le remplacement de la table avait eu lieu. En revanche, le nombre de sorties estimé avec la nouvelle loi semble être plus proche du nombre de sortie réellement constaté par rapport à l'ancienne loi.

Il est donc important de comprendre quels usages peuvent être attribués à la nouvelle table d'expérience. Si elle est construite pour les besoins de provisionnement, il est intéressant d'étudier la possibilité du changement de table. Cependant, il est important de garder une approche prudente et de s'assurer que la diminution de la taille de l'échantillon

liée à la segmentation du modèle ne crée pas une vision biaisée. Pour ceci il serait utile d'effectuer un ou plusieurs exercices de test sur l'historique en calculant les provisions aux moments différents à l'aide des 2 lois et en comparant le résultat avec les chiffres réellement constatés.

En revanche, il s'avère que le modèle segmenté est plus fidèle aux données réelles. Il pourrait donc servir pour la réalisation des travaux de la construction du modèle interne dont l'objectif est d'estimer le besoin en capital de l'assureur de manière précise ou aussi d'être exploité pour les besoins de la mise en place de la norme IFRS 17 qui exige la comptabilisation précise des contrats d'assurance.

Pour conclure, il est intéressant de citer quelques idées qui pourraient être mises en place dans le but d'améliorer la qualité de la table construite. Il serait utile de mener une étude approfondie sur les rentes dont les bénéficiaires ont un âge supérieur à la limite contractuelle et probablement prolonger la table au-delà des 28 ans.

De plus, les données disponibles au niveau national montrent clairement l'existence d'un lien entre l'origine sociale de l'enfant et la durée de ses études. Il serait donc très intéressant de vérifier l'existence de cette corrélation dans le portefeuille d'AXA. Il est aussi possible que le modèle semi-paramétrique de Cox puisse être appliqué pour la segmentation de la population par les origines sociales. Malheureusement l'échantillon de données disponibles contient beaucoup de valeurs manquantes ou mal définies pour la variable catégorie socio-professionnelle de l'assuré, ce qui rend la base de données trop petite pour la construction du modèle fiable.

Conclusion

L'objectif de ce mémoire est de modéliser la loi de maintien pour la garantie rente éducation à partir du portefeuille de la branche Santé et Prévoyance Collectives d'AXA France. La table d'expérience utilisée actuellement pour le provisionnement date de 2014. Le but de l'étude est de proposer des approches alternatives afin de s'assurer de la suffisance du provisionnement de la garantie rente d'éducation.

Une première étude portant sur les données nationales ainsi que sur les rentiers du portefeuille d'AXA montre que les bénéficiaires de la rente ont un comportement qui diffère de celui observé sur la population générale française. Ce constat fait surgir un besoin de construire une table d'expérience. L'analyse des données de l'assureur a permis de construire un échantillon de taille modeste mais suffisante pour la suite de l'étude. Plusieurs modèles ont été proposés pour la construction de la loi de maintien en rente.

Tout d'abord, la loi de sortie a été modélisée en s'appuyant sur les méthodes non paramétriques, notamment sur l'estimateur des taux bruts de Hoem et sur l'estimateur de la fonction de maintien de Kaplan Meier. Les taux obtenus ont été lissés, soumis à une validation et comparés entre eux afin de retenir la meilleure approche non paramétrique. Par la suite, une approche alternative, dite semi-paramétrique, a été utilisée dans le but d'étudier la corrélation qui existe entre la durée de la rente et certaines variables explicatives. Le modèle de Cox a été appliqué sur une population segmentée selon deux critères qui sont les suivants : « sexe du bénéficiaire », « âge du bénéficiaire au moment d'effet de la rente », puis d'un troisième qui est la combinaison de ces deux derniers. Les modèles obtenus ont été testés par rapport à la proportionnalité du hasard de base. Il a été conclu que le modèle de Cox ne peut pas être utilisé pour la construction de la loi de maintien en rente du fait de la non-validation de l'hypothèse centrale de la méthode. Cependant, le modèle de Cox a permis de mettre en évidence l'effet de la variable « sexe du bénéficiaire » sur la durée de la rente. Afin de tenir compte de ce lien, il a été décidé de construire un modèle segmenté, ou autrement dit deux modèles distincts pour les populations féminine et masculine du portefeuille. La modélisation des lois de maintien indépendantes a été faite à l'aide de l'estimateur Kaplan Meier et sans utilisation du lissage des taux bruts. Cette dernière loi permet d'obtenir le modèle optimal en termes d'équilibre entre la fidélité et la prudence, ainsi que de prendre en compte un facteur discriminant important.

Dans la dernière partie du mémoire, le modèle segmenté a été appliqué pour le calcul du nombre estimé de sorties du portefeuille et des engagements de la garantie. Les résultats obtenus ont montré que la table segmentée est moins prudente que celle utilisée actuellement, car elle propose un montant de provisions inférieur d'environ 6% par rapport au montant déterminé avec la loi actuelle. En revanche, l'évaluation de la capacité prédictive du modèle construit lors du présent mémoire a montré que ce dernier représente les données réelles de la manière plus fidèle par rapport à l'ancienne table.

Le travail muni pour la construction de la loi a rencontré quelques difficultés principalement liées aux imperfections des données disponibles. Le risque opérationnel ainsi que le risque d'avis d'expert sont bien présents dans la table d'expérience segmentée. Ce constat indique que la possibilité du remplacement de la loi d'expérience actuellement utilisée pour les travaux de provisionnement par la loi segmentée doit être étudiée avec un regard critique. Il faut notamment s'assurer que la nouvelle table ne génère en aucun cas le risque de sous-provisionnement de la garantie.

En revanche, la table de maintien segmentée semble être bien adaptée pour la projection dans le modèle interne et pour les travaux de la mise en place de la norme IFRS17 afin de mieux refléter le besoin en capital.

Par ailleurs, la qualité de la loi construite pourrait être améliorée via un travail de fiabilisation de la base de données, qui permettrait d'avoir un échantillon plus large et avec moins de fluctuations. Il est également intéressant de prolonger la table d'expérience au-delà de l'âge limite contractuel afin de lisser la queue droite de la distribution et de combler le manque de provisions pour les dossiers atypiques. Enfin, une segmentation du portefeuille selon l'appartenance des assurés aux catégories socio-professionnelles pourrait améliorer la qualité de l'estimation de la loi de maintien et de capter les subtilités qui caractérisent la population des rentiers du portefeuille d'AXA.

Bibliographie

BAILLARD E. (2011). Pilotage prévoyance individuelle. Université de Lyon. Mémoire d'actuariat.

DESPIERRES E. (2010) Construction et validation de tables d'expérience pour le provisionnement d'un portefeuille de rentes éducation. Institut de statistique de l'Université Pierre et Marie Curie. Mémoire d'actuariat.

DUPUY J.F. (2002). Modélisation conjointe de données longitudinales et de durées de vie. Université Paris V, Thèse de doctorat.

Direction de l'évaluation, de la prospective et de la performance (DEPP), Système d'information et des études statistiques (SIES), Ministère en charge de l'agriculture et de la santé, INSEE (2018). Formations et emploi, édition 2018.

Direction de l'évaluation, de la prospective et de la performance (DEPP), Système d'information et des études statistiques (SIES), Ministère en charge de l'agriculture et de la santé, INSEE (2018). L'école et ses sortant, édition 2018.

GEFFRAY S. (2009) Analyse des durées de vie avec le logiciel R. Université de Marseille, p. 1-8.

GUIBERT Q., PLANCHET F. (2014). Construction de lois d'expérience en présence d'évènements concurrents : Application à l'estimation des lois d'incidence d'un contrat dépendance. Bulletin Français d'Actuariat, vol. 14, n°27.

KAMEGA A., PLANCHET F. (2010). Mesure du risque d'estimation associé à une table d'expérience. Laboratoire de sciences actuarielle et financière.

LANGMEIER R. (2000). Étude de différentes méthodes d'ajustement de tables de mortalité : application aux données d'une compagnie d'assurance. Ecole HEC, Université de Lausanne. Travail de mémoire.

PLANCHET F. (2020). Modèles de durée. Introduction – concept de base. Support de cours 2020-2021 disponible sur le site : <http://www.ressources-actuarielles.net/>

PLANCHET F. (2020). Modèles de durée. Tables de mortalité. Support de cours 2020-2021 disponible sur le site : <http://www.ressources-actuarielles.net/>

PLANCHET F. (2020). Modèles de durée. Statistique des modèles non paramétriques. Support de cours 2020-2021 disponible sur le site : <http://www.ressources-actuarielles.net/>

PLANCHET F. (2020). Modèles de durée. Présentation des méthodes de lissage et d'ajustements. Support de cours 2020-2021 disponible sur le site : <http://www.ressources-actuarielles.net/>

PLANCHET F. (2007). Tables de mortalité d'expérience pour les portefeuilles de rentiers (Tables TGH 05 et TGF 05). Notice de présentation.

PLANCHET F., KAMEGA A. (2013). Construction de tables de mortalité prospectives sur un groupe restreint : mesure du risque d'estimation. Bulletin Français d'actuariat, Vol. 13, n°25, pp.5-33.

SAMSOM E. (2008). Identification des marqueurs pronostiques chez les patients atteints d'un cancer de la tête et du cou. Université de Laval. Mémoire de maîtrise en statistique.

Site internet :

<https://www.legifrance.gouv.fr/codes/id/LEGIARTI000006793514/2007-10-01/> (le service public de la diffusion du droit, Code des assurances, Article L141-1), site consulté le 24 octobre 2020.

https://www.axa-assurancescollectives.fr/qui-sommes-nous/?cli_action=1612280302.681 (AXA santé et collectives), site consulté le 24 octobre 2020.

https://entreprise.axa.fr/content/dam/axa-fr-convergence/entreprise/prevoyance/DIPa_AXA_Adaptalia_Entreprises_Metiers_Prevoyance_Socle_Version_2020_03_06.pdf (Document d'information sur le produit d'Assurance Adaptalia Prévoyance), site consulté le 27 octobre 2020.

https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006788624/2005-07-28 (Le service public de la diffusion du droit, Code des assurances, Article A335-1), site consulté le 26 octobre 2020.

<https://www.legifrance.gouv.fr/codes/id/LEGIARTI000006735021/2007-01-11/>(Le service public de la diffusion du droit, Code de la sécurité sociale, Article A931-10-10), site consulté le 26 octobre 2020.

<https://www.education.gouv.fr/l-etat-de-l-ecole-2020-307185> (Publication annuelle sur l'état de l'école en France, édition 2020), site consulté le 30 octobre 2020.

<https://www.education.gouv.fr/reperes-et-references-statistiques-2020-1316> (Repères et références statistiques 2020, Publication annuelle de la Direction de l'évaluation, de la prospective et de la performance et de la Sous-direction des systèmes d'information et des études statistiques), site consulté le 2 novembre.

<https://www.service-public.fr/particuliers/vosdroits/F675> (Le site officiel de l'administration française, sécurité sociale d'un étudiant), site consulté le 10 novembre 2020.

<https://www.education.gouv.fr/l-etat-de-l-ecole-2020-307185> (L'état de l'école 2020, analyse annuelle du système éducatif de la France), site consulté le 11 novembre 2020.

<https://fr.statista.com/statistiques/499756/taux-scolarisation-france-par-age/>

(Statista, Business Data Platform, Statistiques populaires, Education et science), site consulté le 11 novembre 2020.

https://www.insee.fr/fr/statistiques/fichier/3526077/Formemp18f1_F1.1_ecole.pdf

(Insee Références, édition 2018 - Fiches - L'école et ses sortants, volume 1), site consulté le 12 novembre.

https://www.insee.fr/fr/statistiques/fichier/3526077/Formemp18f2_F1.2_ecole.pdf

(Insee Références, édition 2018 - Fiches - L'école et ses sortants, volume 2), site consulté le 12 novembre.

<https://www.insee.fr/fr/statistiques/4238437?sommaire=4238781> (Institut National de la statistique et des études économiques INSEE, France, portrait social, édition 2019), site consulté le 14 novembre 2020.

<https://www.education.gouv.fr/reperes-et-references-statistiques-sur-les-enseignements-la-formation-et-la-recherche-2019-3806> (Publication annuelle de la DEPP et de la SD-SIES, Repères et références statistiques sur les enseignements, la formation et la recherche 2019), site consulté le 14 novembre 2020.

<https://www.education.gouv.fr/series-chronologiques-de-donnees-statistiques-sur-le-systeme-educatif-12530> (Série chronologique de données statistiques sur le système éducatif, actualisée en mai 2020), site consulté le 15 novembre 2020.

<https://www.insee.fr/fr/statistiques/2383587> (Institut National de la statistique et des études économiques INSEE, France, portrait social, édition 2020), site consulté le 14 février 2021.

<https://www.cor-retraites.fr/sites/default/files/2019-06/doc-4628.pdf> (Extraits du rapport du Cnesco 2016 : « Inégalités sociales et migratoires. Comment l'école amplifie-t-elle les inégalités ? »), site consulté le 1 mars 2021.

https://www.legifrance.gouv.fr/conv_coll/article/KALIARTI000030483257

(Convention collective nationale du personnel des prestataires de services dans le domaine du secteur tertiaire du 13 août 1999), site consulté le 3 mars 2020.

<https://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-cause-deces/table-mortalite/> (Institut National d'études démographiques. Table de mortalité 2017-2019), site consulté le 19 septembre 2021.

https://www.insee.fr/fr/statistiques/2648139#tableau-Figure_1 (Institut National de la statistique et des études économiques INSEE, France, portrait social, édition 2020), site consulté le 19 septembre 2021.

Table de figures

Figure 1. Offre Adaptalia d'AXA France : parts utilisées pour le calcul de la rente par tranche d'âge d'assuré et de bénéficiaire	10
Figure 2. Evolution du nombre d'élèves, d'étudiants et d'apprentis en France métropolitaine et DROM depuis 1980 (en milliers).....	12
Figure 3. Nombre de naissances vivantes en France métropolitaine 1990-2019 (en milliers).....	13
Figure 4. Evolution de l'espérance de scolarisation à 15 ans depuis 1985-1986 en France métropolitaine et DROM (en nombre d'années)	14
Figure 5. Niveau de diplôme le plus élevé des personnes de 25-34 ans selon le milieu social en France métropolitaine et DROM en 2019 (en %).....	15
Figure 6. Origine sociale des étudiants de nationalité française en France métropolitaine et DROM en 2019-2020 (en %)	16
Figure 7. Evolution de taux de scolarisation par âge (12 - 29 ans) pour les années 2010, 2015, 2018 en France métropolitaine et DROM	17
Figure 8. Evolution de l'espérance de scolarisation à l'âge de 2 ans selon le sexe en France métropolitaine et DROM depuis 1985	18
Figure 9. Répartition des sortants de formation initiale en fonction de leur diplôme le plus élevé en 2013,2014,2015 et 2017 en France métropolitaine et DROM.....	19
Figure 10. Loi fictive de maintien en rente éducation à partir des données d'INSEE	20
Figure 11. Rentes éducation d'AXA France ventilées par le motif de clôture.....	24
Figure 12. Rentes éducation d'AXA France ventilées par le nouveau motif de clôture.....	25
Figure 13. Statistiques descriptives pour le montant de la rente annuelle versé aux bénéficiaires	26
Figure 14. Statistiques descriptives pour l'âge du bénéficiaire au moment d'effet, de clôture, de fin théorique de la rente et au 31/12/2019	27
Figure 15. Statistiques descriptives pour l'âge du bénéficiaire au moment de clôture de la rente	28
Figure 16. Ventilation des rentiers par âge au moment d'effet de la rente	28
Figure 17. Ventilation des rentiers par âge au moment de clôture de la rente.....	29
Figure 18. Ventilation des rentiers par âge au 31/12/2019.....	30
Figure 19. Ventilation des rentiers par âge au moment de la clôture théorique de la rente	31
Figure 20. Statistiques descriptives pour la durée réelle de rente, la durée théorique de rente et la durée effective de rente au 31/12/2019.....	32
Figure 21. Histogramme pour la durée réelle de la rente.....	32
Figure 22. Histogramme pour la durée théorique de la rente.....	33
Figure 23. Histogramme pour la durée effective de la rente au 31/12/2019.....	33
Figure 24. Répartition des rentiers par sexe	34
Figure 25. Répartition des bénéficiaires de la rente éducation en fonction du motif de clôture.....	34
Figure 26. Répartition des bénéficiaires de la rente éducation en fonction du sexe et du motif de clôture	35
Figure 27. Nombre de sorties et de censures dans le portefeuille des rentiers	40
Figure 28. Exposition au risque de sortie du portefeuille calculée par 2 méthodes.....	41
Figure 29. Taux bruts de sortie de la rente éducation (estimateur de Hoem).....	44
Figure 30. Taux bruts de sortie de la rente éducation (estimateur Kaplan Meier).....	46
Figure 31. Fonction de maintien en rente (estimateur Kaplan Meier)	47
Figure 32. Taux de sortie lissés pour h=1, 5, 10 avec des poids identiques (estimateur de Hoem) pour l'intervalle de 16 à 28 ans.....	50

Figure 33. Valeurs de AIC pour le lissage des taux bruts de Hoem pour $z=2$, poids identiques et $h=1, 5, 10$	50
Figure 34. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids proportionnels à l'exposition (estimateur de Hoem) pour l'intervalle de 16 à 28 ans	51
Figure 35. Valeurs de F pour le lissage des taux bruts de Hoem pour $z=2$, poids proportionnels à l'exposition et $h=1,5,10$	51
Figure 36. Taux de sortie lissés pour $h=1,5,10$ avec des poids choisis analytiquement (estimateur de Hoem) pour l'intervalle de 16 à 28 ans	52
Figure 37. Valeurs de F pour le lissage des taux bruts de Hoem pour $z=2$, poids analytiques et $h=1, 5, 10$	52
Figure 38. Taux de sortie lissés pour $z=2$ avec des poids et le h attribués selon les 3 approches (estimateur de Hoem)	53
Figure 39. Taux de sortie bruts, lissés et retenus (estimateur de Hoem)	54
Figure 40. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids identiques (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans	55
Figure 41. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids identiques et $h=1, 5, 10$	55
Figure 42. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids proportionnels à l'exposition (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans.....	56
Figure 43. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids proportionnels à l'exposition et $h=1, 5, 10$	56
Figure 44. Taux de sortie lissés pour $h=1, 5, 10$ avec des poids choisis analytiquement (estimateur de Kaplan Meier) pour l'intervalle de 16 à 28 ans	57
Figure 45. Valeurs de F pour le lissage des taux bruts de Kaplan Meier pour $z=2$, poids analytiques et $h=1, 5, 10$	57
Figure 46. Taux de sortie lissés pour $z=2$ avec des poids et le h attribués selon les 3 approches (estimateur de Kaplan Meier)	58
Figure 47. Taux de sortie bruts, lissés et retenus (estimateur de Kaplan Meier)	59
Figure 48. Taux de sortie lissés retenus (estimateur de Hoem et de Kaplan Meier)	60
Figure 49. Fonctions de maintien en rente éducation (estimateur de Hoem et de Kaplan Meier).....	61
Figure 50. Résultats de l'estimation du modèle de Cox pour la variable « sexe de bénéficiaire » issus de R.....	66
Figure 51. Courbes Log-log des fonctions de maintien (estimateur Kaplan Meier) pour les hommes et les femmes	67
Figure 52. Taux de sortie de la rente éducation pour la population féminine selon Kaplan Meier et Cox.....	68
Figure 53. Résultats de l'estimation du modèle de Cox pour la variable « groupe d'âge au moment d'effet de la rente » issus de R	69
Figure 54. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 3 groupes d'âge au moment d'effet de la rente	70
Figure 55. Taux de sortie de la rente éducation pour le groupe d'âge 2 selon Kaplan Meier et Cox ...	71
Figure 56. Taux de sortie de la rente éducation pour le groupe d'âge 3 selon Kaplan Meier et Cox ...	71
Figure 57. Répartition des rentiers par groupe en fonction du sexe et de l'âge au moment d'effet de la rente	72

Figure 58. Résultats de l'estimation du modèle de Cox pour les variables « sexe » et « groupe d'âge au moment d'effet de la rente » issus de R	72
Figure 59. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 6 groupes d'âge au moment d'effet de la rente	73
Figure 60. Taux de sortie de la rente éducation pour les groupes combinés « sexe/âge au moment d'effet » selon Kaplan Meier et Cox	74
Figure 61. Répartition des rentiers par groupe en fonction du sexe et de l'âge au moment d'effet de la rente	74
Figure 62. Résultats de l'estimation du modèle de Cox pour les variables « sexe » et « groupe d'âge au moment d'effet de la rente » avec le nombre de catégories réduit issus de R	75
Figure 63. Courbes Log-log des fonctions de survie (estimateur Kaplan Meier) pour 5 groupes d'âge au moment d'effet de la rente	76
Figure 64. Taux de sortie de la rente éducation pour de nouveaux groupes combinés « sexe/âge au moment d'effet » selon Kaplan Meier et Cox	77
Figure 65. Taux de sortie bruts pour les hommes et les femmes séparément et taux de sortie lissés pour la totalité de l'échantillon	78
Figure 66. Lois de maintien en Rente pour les hommes et les femmes de l'échantillon.....	79
Figure 67. Lois de maintien en rente : fictive à partir des données nationales et l'estimateur de Kaplan Meier à partir des données du portefeuille d'AXA.....	80
Figure 68. Nombre de sorties de la rente éducation réel et estimé par la nouvelle et l'ancienne loi de maintien par âge pour la tranche d'âge entre 15 et 28 ans.....	82
Figure 69. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survenance pour les femmes.....	84
Figure 70. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survenance pour les hommes.....	84
Figure 71. Les indicateurs SMR de la nouvelle loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans par survenance – Cumul pour l'intégralité de la population	85
Figure 72. Les indicateurs SMR de l'ancienne loi de maintien calculés pour la totalité des âges et pour les intervalles de 5 ans, par survenance pour l'intégralité de la population	85
Figure 73. Rapport des montants des engagements de la rente éducation calculés par la nouvelle et l'ancienne tables par sexe du bénéficiaire	89
Figure 74. Taux bruts de sortie de Kaplan Meier du modèle retenu et du modèle avec des censures traitées différemment	91