

Mémoire présenté le :
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Adnane AKBARALY

Titre : Construction d'une loi de mortalité prospective sur un produit de retraite par une méthode exogène

Confidentialité : NON (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 : CNP Assurances

Nom :

Signature :

Directeur de mémoire en entreprise :

Nom : Nathalie BIAGINI

Signature :

Membres présents du jury de l'ISFA

.....

.....

.....

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

Ce mémoire a pour but de construire une table de mortalité prospective sur un produit de retraite aux caractéristiques particulières. Il s'agit d'un produit de retraite en points s'inscrivant dans un régime L441 au sein du groupe CNP Assurances.

Le portefeuille est caractérisé par une population jeune. La clôture du régime pendant la période d'observation pose un problème de données censurées. L'absence d'assurés au delà d'un certain âge est problématique dans le cadre d'une approche prospective. Ces caractéristiques fondent l'intérêt de ce travail.

Afin de réaliser la loi de mortalité, l'estimateur de Hoem est utilisé pour ses avantages. Par ailleurs, ce sont les méthodes exogènes, c'est-à-dire de positionnement par rapport à une table de référence en raison d'un faible volume de données comme la méthode du SMR, le modèle semi-paramétrique de Brass et le modèle linéaire généralisé de Poisson qui sont développés dans le corps de ce mémoire. Enfin le modèle de Denuit & Goderniaux est appliqué pour les âges élevés.

Mots-clés : *Tables de mortalité prospectives, critère de Cochran, censure et troncature, diagramme de Lexis, méthode de positionnement, SMR, modèle de BRASS, modèle de Denuit & Goderniaux*

Abstract

The purpose of this paper is to build a prospective mortality table on a pension product with specific characteristics. This is about a points-based product in a L441 plan within the CNP Assurances group.

The portfolio is characterised by a young population. The closure of the scheme during the observation period poses a problem of censored data. The absence of insureds above a certain age is problematic in a prospective approach. These characteristics are the basis for the interest of this work

In order to realise the law of mortality, the Hoem estimator is used for its advantages. In addition, exogenous methods, i.e. positioning in relation to a reference table due to a low volume of data such as the SMR method, the semi-parametric model of Brass and the generalized linear Poisson model are developed in the body of this thesis. Finally the Denuit & Goderniaux model is applied for high ages.

Keywords : *Prospective life tables, Cochran criterion, censoring and truncation, Lexis diagram, positioning method, SMR, BRASS model, Denuit & Goderniaux model*

Remerciements

J'adresse mes remerciements à l'équipe Epargne Retraite au sein de la Direction Technique et de l'Innovation chez CNP Assurances pour l'accueil, le soutien et l'accompagnement durant toute ma période d'alternance.

Je remercie également Tristan CLUZEL, manager de l'équipe Retraite pour son suivi. En effet ses questionnements et réflexions m'ont permis d'apporter de la précision dans la rédaction du mémoire.

Je remercie tout particulièrement Nathalie BIAGINI, ma tutrice en entreprise qui m'a apporté un soutien régulier et permanent tout au long de l'alternance et encore plus pour la réalisation du mémoire. Je suis extrêmement reconnaissant de tout le temps consacré à mon apprentissage pratique en entreprise, pour toutes les réunions faites sur l'avancée du mémoire, pour sa pédagogie bienveillante et pour m'avoir partagé son expertise du domaine de la retraite.

J'adresse mes remerciements à toute l'équipe pédagogique de l'ISFA pour les précieux enseignements.

Je remercie tout particulièrement Pierre Olivier GOFFARD, mon tuteur pédagogique et Frédéric PLANCHET pour leurs échanges enrichissants, leurs conseils et leurs recommandations bibliographiques qui m'ont apporté énormément de clarté dans la compréhension du sujet.

Note de synthèse

Introduction

Le but de cette étude est d'établir une loi de mortalité sur un produit de retraite inscrit dans le régime L441 du code des assurances. En effet, une analyse en interne a montré que ce portefeuille présente une surmortalité récurrente sur les trois dernières années. Le portefeuille en question présente des caractéristiques particulières qui seront prises en compte à la fois dans le choix du calibrage, des modélisations et dans l'étape de validation. Enfin, le changement de table de mortalité aura des impacts financiers qui seront mesurés à l'issue de l'étude.

Récupération et traitement des données

Le chapitre sur la récupération et le traitement de donnée aborde la manière dont ont été récupérées les bases de données. Deux systèmes de gestion ont été sollicités, un pour extraire les données sur les cotisants et l'autre pour les rentiers.

De ce fait, une série de contrôle a été établie et réalisée afin de conserver uniquement les données utiles, utilisables et cohérentes à l'étude. Les contrôles ont été réalisés sur les deux bases car celles-ci ne comportent pas toujours les mêmes variables. Enfin, la base de données des cotisants et des rentiers est concaténée pour avoir une base complète.

Enfin cette partie inclut surtout des statistiques descriptives du portefeuille, que ce soit uniquement sur les cotisants, sur les rentiers ou sur l'ensemble du portefeuille. Ces informations sont primordiales pour comprendre les choix qui seront faits dans la suite.

Caractéristiques du portefeuille

Les statistiques descriptives du portefeuille permettent de comprendre le comportement temporel du portefeuille. En effet, le régime du portefeuille à l'étude a été clôturé en 2016, les nouvelles souscriptions ne sont plus autorisées et les jeunes cotisants sortent du portefeuille et sont remboursées. Cet événement a un impact direct sur l'exposition, l'âge moyen ou encore la répartition actif/retraités.

Outre cette perturbation du régime, il faut noter que le portefeuille est constitué d'une population jeune, avec une grande majorité d'hommes et ne comportant que très peu de décès sur les premières années d'observation. Ces informations vont être utilisées pour le calibrage du domaine d'application.

Calibrage du domaine d'application

Le calibrage consiste à définir une période d'application et une tranche d'âge d'application.

En effet, toutes les années d'observations ne sont pas prises en compte dans les modélisations car certaines années apportent des informations non pertinentes dû à un volume de données trop faible, d'autres apportent des informations inscrites dans un contexte exceptionnel notamment 2020 avec l'impact COVID. Ainsi celles-ci sont à exclure dans l'optique de générer une loi de mortalité stable. La période retenue est 2012 – 2019 dans un premier temps puis 2014 – 2019 dans un second temps.

Un arbitrage a également été fait sur la tranche d'âge d'application puisque l'exposition par âge est très hétérogène avec toutefois une forte concentration de celle-ci sur la tranche d'âge 40 – 70 ans au cours du temps. De plus il a fallu prendre en considération la fermeture du régime en 2016 qui a entraîné un changement majeur dans la distribution de l'exposition par âge.

L'absence d'observations au-delà de 75 ans comporte un risque sur les projections de mortalité future qu'il va falloir contrôler et mesurer afin de ne pas retenir les tendances incohérentes et non prudentes.

Estimation et modélisation de la mortalité

La première étape commence par l'estimation des taux bruts de mortalité. Des notions clés sont à introduire afin de comprendre les différents modèles utilisés. A commencer par les concepts mathématiques pour les calculs actuariels, suivi du diagramme de Lexis, des troncatures et censures et pour finir l'estimateur de Hoem [9] qui sera l'estimateur utilisé par les modèles.

La seconde étape consiste à modéliser les taux de mortalité à l'aide de modèles. Les méthodes proposées dans ce mémoire concernent principalement les méthodes exogènes, aussi appelées méthodes de positionnement qui consistent à capter la sous-mortalité ou surmortalité du portefeuille et à l'appliquer à une table de mortalité de référence.

Trois modèles sont ainsi testés. Le modèle du SMR [6], modèle à un paramètre. Le modèle semi-paramétrique de Brass [6], modèle à deux paramètres utilisant la fonction logit. Et le modèle linéaire généralisé de Poisson, modèle à trois paramètres qui fait intervenir l'âge comme une variable explicative du taux instantané de mortalité.

Les modèles ainsi appliqués permettent d'obtenir un ajustement de la mortalité sur la tranche d'âge d'application. Afin d'extrapoler la mortalité sur la tranche d'âge au-delà de 70 ans, la méthode de Denuit & Goderniaux [1] est utilisée. Celle-ci consiste simplement à prolonger les ajustements précédemment obtenus par une fonction quadratique respectant un certain nombre de contraintes pour assurer la cohérence de l'extrapolation.

Au global, tous les modèles testés montrent que le portefeuille à l'étude est caractérisé par une surmortalité notable par rapport aux tables de mortalité réglementaires TGH-05/TGF-05. La pertinence des modèles sera alors mesurée dans la partie suivante.

Validation et impacts financiers

Ce mémoire s'achève sur l'évaluation des ajustements de mortalité à travers plusieurs critères de validité. A savoir, l'analyse graphique de différents types de résidus pour une analyse locale et la comparaison d'indicateurs sur la qualité globale des ajustements avec l'observation comme le critère du R^2 , χ^2 , $MAPE$ ou encore le test de signes et le test des runs.

Puis, il est également important d'évaluer la qualité et la pertinence des taux de mortalité projetés dans le futur. De ce fait, plusieurs indicateurs sont utilisés comme l'entropie, l'âge médian aux décès et l'espérance de vie générationnelle partielle.

Enfin, la dernière partie consiste à mesurer l'impact du changement de table sur quelques métriques financières permettant le pilotage du régime. Les métriques concernées sont la PMT, le taux de couverture et l'equity. Pour comprendre la variation d'equity, il sera intéressant d'analyser l'évolution de la PTS, des arrérages et de la VS. Toutes ces métriques sont définies et expliquées dans le chapitre concerné.

En résumé, la surmortalité avérée impacte les métriques financières dans le même sens peu importe le modèle utilisé. L'impact sur le taux de couverture est positif alors que l'impact sur l'equity est négatif. Un juste arbitrage est à réaliser afin de choisir le meilleur modèle.

Summary note

Introduction

The purpose of this study is to establish a mortality table on a retirement product registered in the L441 plan. Indeed, an internal analysis has shown that this portfolio has a recurrent mortality excess over the last three years. The portfolio in question has specific characteristics that will be taken into account in the choice of the calibration, in the modelling stage and in the validation stage. Finally, the change of mortality table will have financial impacts that will be measured at the end of the study.

Data recovery and processing

The chapter on data retrieval and processing discusses how the databases were retrieved. Two management systems were used, one to extract data on contributors and the other for annuitants.

As a result, a series of controls was established and carried out in order to keep only useful, usable and consistent data for the study. The checks were done on both databases because they do not always contain the same variables. Finally, the contributor and annuitant databases were concatenated to obtain a complete database.

Finally, this section includes descriptive statistics of the portfolio, whether it is only on contributors, on annuitants or on the whole portfolio. This information is essential to understand the choices made thereafter.

Portfolio characteristics

The descriptive statistics of the portfolio provide insight into the temporal behavior of the portfolio. Indeed, the portfolio plan under study was closed in 2016, new subscriptions are no longer allowed and young contributors leave the portfolio and are reimbursed. This phenomenon has a direct impact on the exposure, the average age or the asset/retiree distribution.

In addition to this perturbation of the plan, it should be noted that the portfolio is made up of a young population, with a large majority of men and with very few deaths during the first years of observation. This information will be used to calibrate the application domain.

Calibration of the application field

Calibration consists of defining an application period and an application age range.

Indeed, all the years of observations are not taken into account in the models because certain years bring irrelevant information, others bring information registered in an exceptional context. Thus these are excluded in order to generate a stable mortality table. The selected period is 2012 - 2019 at first then 2014 - 2019.

An choice was also made on the age range of application since the exposure by age is very heterogeneous with however a strong concentration on the age range 40 - 70 years over time. In addition, we had to take into account the closure of the plan in 2016 which led to a major change in the distribution of exposure by age.

The absence of observations beyond age 75 carries a risk on future mortality projections that will have to be controlled and measured in order to not retain inconsistent and non-conservative trends.

Mortality estimation and modeling

The first step begins with the estimation of crude mortality rates. Key notions have to be introduced in order to understand the different models used. Starting with the mathematical concepts for the actuarial calculations, followed by the Lexis diagram, truncations and censors and finally the Hoem estimator [9] which will be the estimator used by the models.

The second step is to model the mortality rates using models. The methods proposed in this thesis mainly concern exogenous methods, also called positioning methods, which consist in capturing the portfolio mortality and applying it to a reference mortality table.

Three models are thus tested. The SMR model [6], a one parameter model. The semi-parametric Brass model [6], a two-parameter model using the logit function. And the generalized linear Poisson model, a three-parameter model that uses age as an explanatory variable for the instantaneous mortality rate.

The models applied in this way make it possible to obtain an adjustment of the mortality on the age bracket of application. In order to extrapolate the mortality on the age bracket beyond 70 years, the method of Denuit & Goderniaux [1] is used. This one consists simply in extending the adjustments previously obtained by a quadratic function respecting a certain number of constraint to ensure the coherence of the extrapolation.

Overall, all the models tested show that the portfolio under study is characterized by a significant excess of mortality compared to TGHF-05. The relevance of the models will then be measured in the following section.

Validation and financial impacts

This paper concludes with the evaluation of the mortality fits through several validity criteria. Namely, the graphical analysis of different types of residuals for a local analysis and the comparison of indicators on the global quality of the fits with the observation such as the criterion of R^2 , χ^2 , $MAPE$ or the sign test and the run test.

Following this, it is also important to assess the quality and relevance of projected mortality rates in the future. For this, several indicators are used such as entropy, median age at death and partial generational life expectancy.

Finally, the last part consists in measuring the impact of the change of table on some financial metrics allowing the management of the plan. The metrics concerned are the TMP, the coverage rate and the equity. To understand the equity variation, it will be interesting to analyze the evolution of the STP, the arrears and the SV. All these metrics are defined and explained in the relevant chapter.

In summary, the proven excess mortality impacts the financial metrics in the same direction regardless of the model used. The impact on the coverage rate is positive while the impact on equity is negative. A fair trade-off must be made in order to choose the best model.

Table des matières

1	Contexte de l'étude	13
1.1	Fonctionnement de la retraite en France	13
1.1.1	Retraite par répartition	13
1.1.2	Retraite par capitalisation	14
1.1.3	Remise en question de la pérennité des régimes en répartition	15
1.2	Tables de mortalité	18
1.2.1	Tables de mortalité réglementaires	18
1.2.2	Tables d'expérience	18
1.3	Présentation du régime	19
1.3.1	Contexte historique du régime	19
1.3.2	Régime L.441	20
1.3.3	Caractéristiques du produit	20
2	Construction de la base de données	21
2.1	Système de gestion des cotisants	21
2.1.1	Récupération des données	21
2.1.2	Traitement des données	22
2.1.3	Statistiques générales des cotisants	24
2.2	Système de gestion des rentiers	29
2.2.1	Récupération des données	29
2.2.2	Traitement des données	31
2.2.3	Statistiques générales des rentiers	32
2.3	Base finale	37
2.3.1	Concaténation des bases	37
2.3.2	Statistiques générales	38
3	Problématique et réflexions spécifiques au portefeuille	42
3.1	Réflexions compte tenu des caractéristiques du portefeuille	42
3.2	Choix des années d'observation à retenir compte tenu de la fermeture du régime	44
3.3	Choix de la plage d'âge à retenir	45
4	Méthodes de construction et modélisations des tables de mortalités prospectives	47
4.1	Estimation des taux bruts de mortalité	47
4.1.1	Notions mathématiques actuarielles	47
4.1.2	Diagramme de Lexis	50
4.1.3	Troncature gauche et censure aléatoire droite	52
4.1.4	Estimateur de Hoem	54
4.2	Ajustement des taux bruts de mortalité	59
4.2.1	SMR (Standardized Mortality Ratio)	59

TABLE DES MATIÈRES

4.2.2	Régression logistique - variante du modèle de BRASS	62
4.2.3	Modèle linéaire généralisé de Poisson	68
4.3	Fermeture de la table	72
4.3.1	Méthode Denuit & Goderniaux	72
5	Validations et impacts financiers	77
5.1	Critères de validation	77
5.1.1	Validation de fidélité des ajustements	77
5.1.2	Validation de la régularité des ajustements	82
5.1.3	Validation des tendances d'évolution de la mortalité	86
5.2	Impacts financiers	90
5.2.1	Backtesting des différents modèles sur le nombre de décès	90
5.2.2	Impact sur les arrérages, la PTS et la VS	93
5.2.3	Impact sur le taux de couverture, la PMT et l'equity	96
	Conclusion	99
	Bibliographie	101
	Annexes	103

Chapitre 1

Contexte de l'étude

1.1 Fonctionnement de la retraite en France

En France, il existe deux systèmes de retraite. La retraite par répartition et la retraite par capitalisation. Ces deux systèmes sont abordés dans ce chapitre.

1.1.1 Retraite par répartition

Principe de fonctionnement :

La retraite par répartition est un dispositif mis en place dans un principe de solidarité entre les générations. Le dispositif consiste à prélever une partie des revenus des travailleurs en activité appelée cotisations qui seront versées immédiatement aux retraités sous forme de pension.

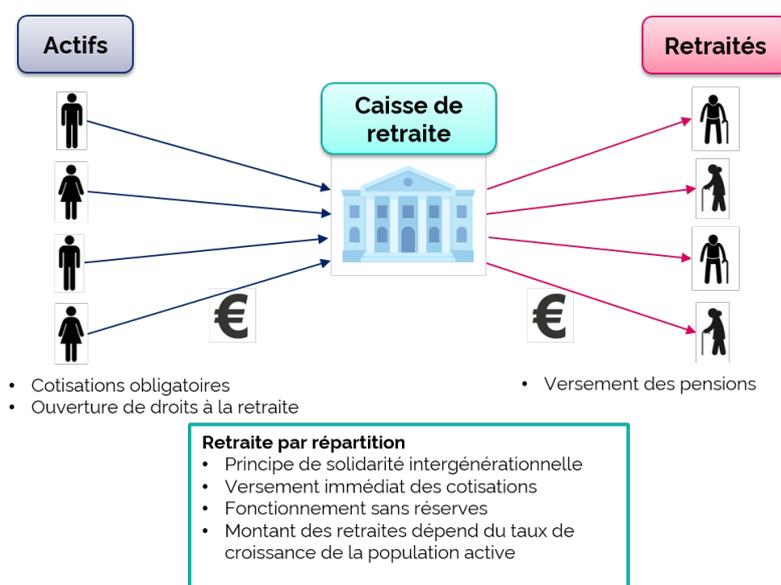


FIGURE 1.1 – Schéma simplifié de la retraite par répartition

Ce dispositif a connu le jour pour la première fois en 1910 avec les retraites ouvrières et paysannes. Puis c'est en 1945 que la retraite par répartition se démocratise en France avec la naissance de la Sécurité Sociale qui instaure le régime général des retraites qui s'applique à tous les salariés du privé (hors agriculture).

Une des motivations de ce type de régime était de pouvoir apporter un soutien financier immédiat (voir figure 1.1) aux invalides de guerre ainsi qu'aux retraités d'après-guerre. Ce système de retraite a connu un grand succès. Par la suite, de nombreux régimes par répartition ont vu le jour en fonction de la catégorie socio-professionnelle des citoyens, à savoir tous les régimes spéciaux.

L'avantage de ce type de régime est qu'il permet de verser des pensions aux retraités sans devoir faire de réserve, ce qui répond à un besoin urgent et immédiat. Cependant, pour que la retraite par répartition fonctionne de façon pérenne et stable, il est nécessaire que la proportion entre retraités et actifs ne soit pas perturbée.

Les avantages et limites de la retraite par répartition sont expliqués dans la partie suivante.

1.1.2 Retraite par capitalisation

Principe de fonctionnement :

La retraite par capitalisation consiste à cotiser pour sa propre retraite. L'assuré décide des sommes à épargner, celles-ci sont placées dans des instruments financiers, immobiliers ou autre sur le long terme. L'épargne ainsi fructifiée est versée à la retraite de l'assuré sous forme de capital ou de rente viagère.

Ce système de retraite n'est pas opposé à celui de la retraite par répartition, bien au contraire il doit être vu comme un moyen de se constituer une retraite supplémentaire qui va venir compléter la pension versée par les régimes par répartition.

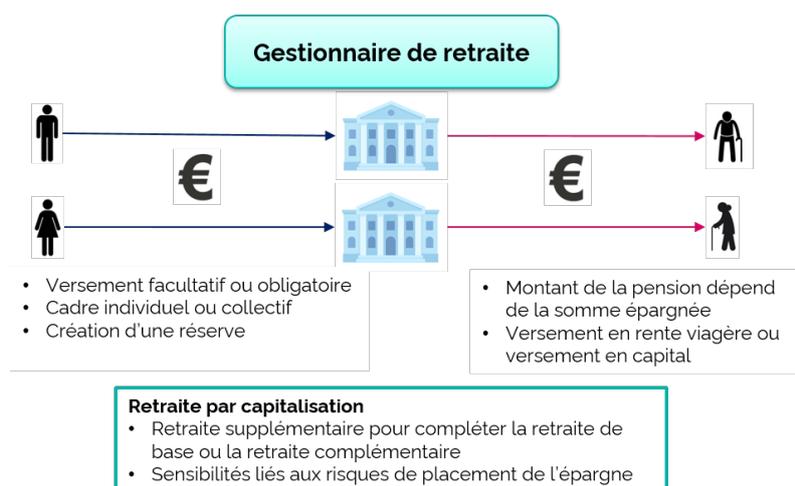


FIGURE 1.2 – Schéma simplifié de la retraite par capitalisation

La retraite par capitalisation était présente au début du XX^{ème} siècle. L'évolution du capital généré par la retraite par capitalisation est sensible aux évolutions économiques et financières. De ce fait, suite à la Grande Dépression et aux deux guerres mondiales, les épargnants ont perdu toutes leurs économies et ont perdu confiance en ce système de retraite.

Les régimes par répartition ont été mis en place après la Seconde Guerre mondiale mais commencèrent à présenter certaines limites dans leur fonctionnement à la fin du XX^{ème} siècle, ce qui encourage les individus à chercher d'autres solutions d'épargne. Ainsi la retraite par capitalisation a réapparue et s'est de nouveau popularisée.

Le cadre de la retraite par capitalisation peut être individuel, à savoir chaque individu gère son capital-épargne, ou il peut être collectif, proposé par son entreprise par exemple et géré par des organismes spécialisés.

L'épargnant est alors libre de choisir ses instruments de placement en fonction de ses objectifs à court ou à long terme et de son profil d'investisseur. De plus, il existe des produits bénéficiant d'avantages fiscaux qui ont pour but d'inciter les épargnants à se tourner vers ce mode de placement.

La retraite par capitalisation est aujourd'hui fortement encouragée car la retraite par répartition actuelle atteint ses limites. En effet, l'allongement de l'espérance de vie, le papy-boom et l'évolution démographique sont des freins à la pérennité de ce système de retraite.

1.1.3 Remise en question de la pérennité des régimes en répartition

Limites de la retraite par répartition

Le principal risque qui pèse sur le système de retraite par répartition est le risque démographique. Si l'on se replace au milieu du XX^{ème} siècle à l'après-guerre, le contexte était favorable au fonctionnement de la retraite par répartition.

En effet, l'âge de la Retraite était au départ fixé à 65 ans alors que l'espérance de vie à la naissance était autour de 63 ans pour les hommes et 70 ans pour les femmes.

De fait, la proportion de retraités à cette époque était très faible comparée à la proportion d'actifs. Arriva ensuite le « baby-boom », période durant laquelle le taux de natalité connaît une hausse inédite située historiquement entre 1945 et 1970. Cette hausse n'a pas de conséquence à cette époque sur le régime par répartition bien au contraire.

Cependant les progrès en médecine, l'amélioration des conditions de vie et les progrès technologiques entre autres ont contribué à faire grimper l'espérance de vie jusqu'à atteindre 80 ans pour les hommes et 85 ans pour les femmes en 2018.

En ce qui concerne l'âge de la retraite, celui-ci n'a pas évolué au même rythme que l'espérance de vie puisqu'il a varié entre 60 ans et 65 ans selon les gouvernements et les différentes lois à ce sujet.

Ces événements ont donc créé un déséquilibre démographique affaiblissant le système de retraite par répartition.

Pyramide des âges des cotisants et de la population française

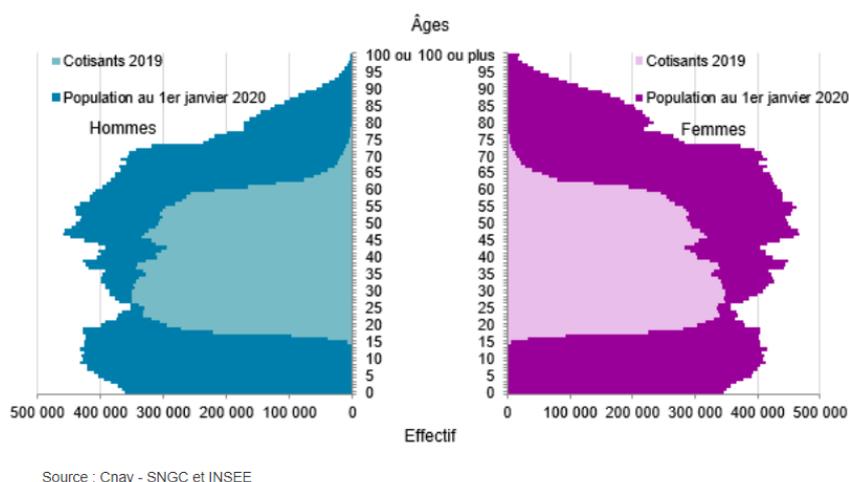
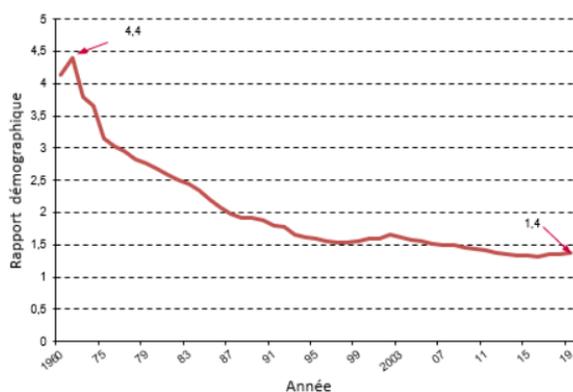


FIGURE 1.3 – Pyramide des âges de la population française en 2019

La pyramide des âges en 2020 expose qu'une grande partie de la population française a plus de 65 ans et que la part des cotisants est faible relativement à la population française que ce soit pour les hommes ou pour les femmes.

Le rapport démographique (ratio : cotisants / retraités) a sans doute diminué depuis 1945 avec l'augmentation de l'espérance de vie et la stagnation de l'âge de départ à la retraite.

Évolution du rapport démographique depuis 1960



Source : Cnav - SNGC et INSEE

FIGURE 1.4 – Evolution du ratio cotisants/retraités depuis 1960

En effet, on observe que le nombre de cotisant pour la pension d'un retraité était de 4 en 1960, puis a baissé à 2 en 1987 et est de 1,4 en 2019. Une tendance baissière du rapport démographique en fonction du temps est observable, ce qui est défavorable au système de retraite par répartition.

La retraite par répartition a été une solution efficace et rapide à l'après-guerre pour apporter un soutien financier aux démunis victimes de guerre mais elle a atteint ses limites face au déséquilibre démographique actuel. Une profonde réforme sur son fonctionnement est alors nécessaire pour que la retraite par répartition puisse fonctionner avec le contexte économique et démographique actuel.

Pistes de réflexions sur d'éventuelles solutions

Depuis les années 2010, les caisses de retraites ont de plus en plus de mal à assurer leur fonction car les cotisations ne suffisent plus à couvrir le versement des rentes. De là, certaines caisses sont devenues déficitaires et sont à la recherche de solutions pour éviter que cette situation n'empire. De nombreuses réformes et lois ont vu le jour ces dix dernières années dans le but d'améliorer la santé financière des caisses de retraite. Par ailleurs, certains régimes pensent à fusionner entre eux pour pouvoir équilibrer leur bilan financier comme la fusion AGIRC-ARRCO en 2019.

Pour assurer le bon fonctionnement et la pérennité de la retraite par répartition malgré les difficultés et contraintes dont elle est confrontée aujourd'hui, plusieurs paramètres peuvent être ajustés. Afin d'être neutre, les principaux paramètres qui permettent de piloter ce type de régime vont être cités et explicités sans pour autant privilégier une direction par rapport à une autre. Les principaux paramètres sont donc la part salariale consacrée aux cotisations des retraités, le montant des rentes et l'âge de départ à la retraite.

Le premier paramètre est la part salariale des cotisations pour la retraite. Augmenter cette part permet d'une part d'augmenter la trésorerie des caisses, ce qui permet d'avoir davantage de liquidités pour assurer le versement des pensions. D'autre part, cela a un impact direct sur les actifs puisque leur pouvoir d'achat diminue. Il y a donc un arbitrage à considérer pour le choix de ce paramètre.

Le montant des rentes est peut-être le paramètre le moins flexible car il dépend directement des sommes récupérées issues des cotisations et dépend du nombre de retraités. Le montant des cotisations peut être piloté et non pas l'évolution du nombre de retraités. En effet, plus le nombre de retraités augmente, plus la pension est faible puisque ce type de régime est basé sur l'idée de partage. L'allongement de l'espérance de vie est ici une contrainte. Néanmoins, un montant minimal de rente est garanti par l'état quelle que soit la situation financière de la caisse.

Enfin l'âge de départ à la retraite a une très forte influence sur le ratio démographique puisqu'il détermine la séparation entre ceux qui alimentent la caisse et ceux qui en bénéficient. De là, augmenter l'âge de départ à la retraite permet d'augmenter le ratio démographique, ce qui est favorable à la santé financière du régime, à l'inverse baisser l'âge de départ à la retraite fait diminuer le ratio démographique ce qui est néfaste pour le régime. C'est un paramètre à reconsidérer compte tenu de l'allongement de l'espérance de vie qui est en défaveur du bon fonctionnement de la retraite par répartition.

1.2 Tables de mortalité

L'activité assurantielle est une activité complexe faisant appel à des outils sophistiqués en fonction des secteurs d'activités. Afin de confectionner et de suivre des produits de retraite, les actuaires font usage de tables de mortalité.

Une table de mortalité est un tableau qui donne en fonction de l'âge et d'une année de naissance la probabilité de décès sur l'année d'un individu. On peut distinguer deux grandes catégories de tables de mortalité.

Les tables de mortalité réglementaires et les tables de mortalité d'expériences. Les tables de mortalité servent à tarifier et à provisionner, leurs utilisations selon la réglementation sont définies dans les articles A132-18 du Code des Assurances [8].

1.2.1 Tables de mortalité réglementaires

Au sein des tables de mortalités réglementaires, on distingue deux sous-catégories.

- Les tables TH et TF 00-02 pour les assurances en cas de décès
- Les tables de mortalité pour les contrats de rentes viagères immédiates ou différées

Les tables TH et TF 00-02 ne concernent pas l'objet de ce mémoire étant donné que le produit étudié porte sur des rentes viagères, celles-ci ne sont donc pas plus détaillées.

Parmi les tables de mortalité pour les contrats de rentes viagères immédiates ou différées, on distingue les tables de mortalité prospectives pour les rentes viagères qui sont les TGH et TGF-05, homologuées depuis le 01/08/2006. Elles ont été construites sur la période 1993-2005 sur la base de la population bénéficiaire de contrats de rentes et sur la période 1962-2000 sur la base de données de l'INSEE. Depuis le 01/01/2007, ces tables sont utilisées à des fins de tarification et de provisionnement des contrats de rentes viagères immédiates ou différées.

1.2.2 Tables d'expérience

Les tables de mortalités d'expérience sont des tables créées par les assureurs. Ceux-ci observent et jugent que la mortalité de leurs portefeuilles pour des raisons liées à leurs caractéristiques se comportent de manière significativement différente des tables de mortalités réglementaires qui concernent la population générale. Ce cadre est également régi par l'article A132-18 du Code des Assurances.

Les tables de mortalités d'expérience sont amenées à être certifiées par l'Institut des Actuaires et font l'objet d'un suivi annuel afin de s'assurer de la validité de celles-ci. Certaines tables de mortalité d'expérience n'ont pas besoin d'être certifiées et sont juste utilisées en interne.

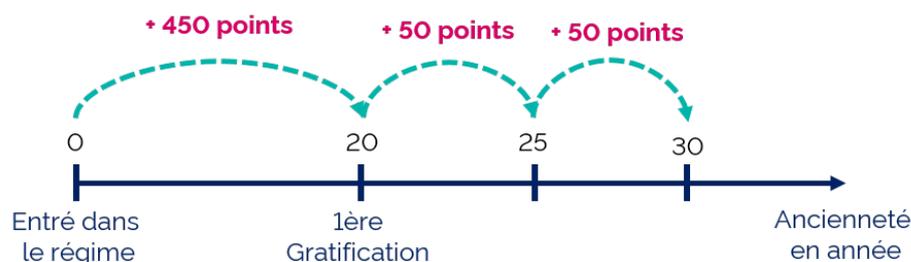
1.3 Présentation du régime

1.3.1 Contexte historique du régime

Le portefeuille sur lequel la table de mortalité est créé est un portefeuille constitué d'une population d'une profession bien particulière dont les caractéristiques seront présentées dans la suite. Pour des raisons de confidentialité, le régime sera nommé "A".

CNP Assurances a obtenu l'appel d'offre pour la création du régime A au 1er janvier 2005 pour une durée de 11 ans.

Ce régime inclut un système de bonification de points en fonction de l'ancienneté au sein du régime ayant pour objectif de fidéliser et d'attribuer une certaine reconnaissance aux assurés comme le montre la figure suivante.



Système de fidélisation et reconnaissance :

- Gratification de 450 points pour avoir atteint les 20 ans d'anciennetés
- Bonification de 50 points tous les 5 ans à partir de 20 ans d'anciennetés

FIGURE 1.5 – Schéma simplifié de la distribution de point du régime

Pour pouvoir rentrer dans le régime les conditions suivantes doivent être vérifiées : avoir entre 16 et 60 ans et satisfaire une certaine aptitude physique et médicale. Les assurés entrant dans le régime payent une cotisation faible au regard des avantages que procure le régime.

Au bout de 20 ans d'anciennetés, une première gratification de 450 points est adressée à l'assuré, puis tous les 5 ans à partir de 20 ans d'anciennetés, un bonus de 50 points est attribué à l'assuré.

En effet, ce régime a été créé à des fins de fidélisation et de reconnaissance envers les travailleurs pour leurs services.

En 2016, le contrat avec la CNP Assurances clôture car arrivant à sa fin. Le régime est mis en run-off.

Il n'y a alors plus de nouvelles souscriptions et tous les cotisants qui n'ont pas obtenu le premier seuil de points, à savoir 450 points sont remboursés et exclus du régime. Il ne reste alors depuis 2016 que des retraités et des cotisants ayant au moins 20 ans d'ancienneté.

1.3.2 Régime L.441

Le régime A est un régime en points relevant de la branche 26 du Code des Assurances (régime L441).

L'article L441 du Code des assurances définit les modalités de fonctionnement relatives aux opérations de retraite par capitalisation. Le régime L441 est un régime à caractère collectif, à cotisations définies et fonctionnant en point.

1.3.3 Caractéristiques du produit

Par ailleurs, à la création du régime, l'effectif est d'environ 185 000 personnes et depuis la fermeture du régime en 2016, l'effectif a été réduit à près de 58 000 adhérents.

Le régime A présente une particularité, celle que les hommes représentent une très large proportion du portefeuille. Dans la suite l'évolution des proportions par sexe est détaillée en fonction du temps.

Chapitre 2

Construction de la base de données

Ce chapitre est consacré à la récupération, aux traitements et à l'utilisation des données. Trois sous-parties composent ce chapitre, une consacrée aux données sur les cotisants, une autre sur les retraités et une dernière qui rassemble tout le portefeuille. Les statistiques générales sont présentées afin de percevoir une idée globale des caractéristiques du portefeuille que ce soit uniquement sur les cotisants, les rentiers ou tout l'ensemble.

2.1 Système de gestion des cotisants

2.1.1 Récupération des données

La récupération des données pour la base des cotisants est moins complexe que celle des rentiers puisque le système de gestion des cotisants contient toutes les données pour le régime A.

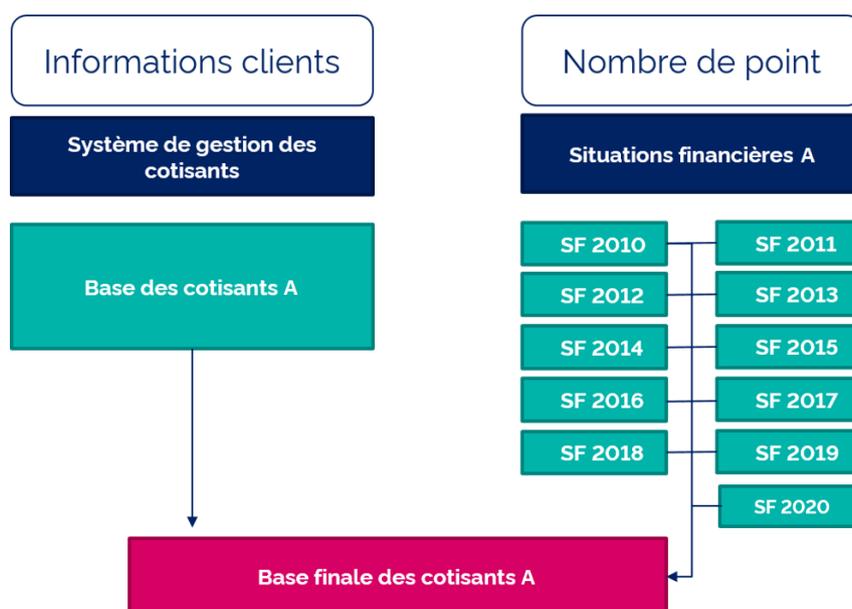


FIGURE 2.1 – Schéma descriptif des sources de données

De la même manière que pour le système de gestion des rentiers, le système de gestion des cotisants fournit une base de données à jour sur le régime A.

L'utilisation des SF (Situation Financière) sert à réaliser des vérifications d'exhaustivité et également permet si besoin de récupérer le nombre de points des assurés année après année.

La récupération de cette base de données auprès de la gestion avec toutes les variables nécessaires à l'étude a été de loin la partie la plus chronophage. Une fois que les données sont récupérées en totalité, les différents contrôles de qualité des données sont effectués, c'est à dire les contrôles d'exhaustivité, de cohérence et de format afin de pouvoir être correctement traités.

Le système de gestion fournit une liste de variable sur les contrats du régime A qui est la suivante :

- Client compte
- Id client
- Identifiant
- Sexe
- Code de sortie
- Date de naissance
- Date de souscription
- Date sortie
- Date de décès
- Points

De la même manière que dans la base des rentiers, toutes ces variables in fine ne sont pas conservées pour l'étude. Ces variables ont été demandées à la gestion afin que la réalisation des contrôles se fasse sans manquement d'information.

La base de cotisants contient 341 418 observations avant réalisation des contrôles.

2.1.2 Traitement des données

Le contrôle d'exhaustivité est effectué en premier puisque il est primordial de s'assurer de la complétude des données.

Ainsi, la base de cotisants est mise en forme pour que celle-ci soit comparée aux SF. Tous les traitements et contrôles de qualité sur les données ont été réalisés sur Python. L'importation des SF est faite, puis elles sont confrontées une par une à la base de données de travail sur laquelle un filtre est réalisé sur l'année de souscription et l'année de sortie.

De ce fait, le sous ensemble obtenu à partir de la base de cotisants peut être comparé à la SF de l'année choisie. Une comparaison est réalisée d'une part sur l'effectif total des deux bases, puis le nombre de point total afin de s'assurer qu'il n'y a pas d'écart. Dans le cas contraire, cet écart est mesuré et les contrats manquants ou les contrats en trop sont traqués dans la base des cotisants afin de les signaler à la gestion.

En ce qui concerne le nombre de point, la base de cotisants ne contient qu'une seule variable sur le nombre de point qui correspond au dernier nombre de point attribué par contrat car chaque année, le nouveau nombre de point écrase le précédent. Ainsi, l'évolution du nombre de point n'est pas vérifiable par assuré à travers les SF sauf avec la SF la plus récente. Un contrôle en point est tout de même réalisé, celui-ci consiste à vérifier que l'écart en point entre la base de cotisants et les SF est décroissant en fonction des années. Ceci assure que le nombre de point est croissant et que le nombre de point inscrit dans la base des cotisants est le dernier nombre de point enregistré.

Par ailleurs, une autre difficulté sur le contrôle des points est posée par la fermeture du régime en 2016 qui a fait qu'un grand nombre de cotisant du régime ont été remboursés et sont sortis du régime car ces derniers n'ont pas atteint le seuil d'ancienneté suffisant pour acquérir un nombre de point minimal pour rester dans le régime.

Viennent alors les contrôles de cohérence sur la qualité des données. Ces derniers sont similaires aux contrôles de cohérence établis pour la base de rentiers (DDN est la date de naissance).

Contrôle et suppression des données aberrantes :

- Contrôle du sexe (remplacer les 3 -> Mlle par des 2 -> Mme)
- Suppression de DDN > Date_Souscription
- Suppression de Date_sortie < 2005
- Suppression de DDN > Date_décès
- Suppression de Date_sortie < Date_décès
- Suppression de Date_Souscription > Date_décès
- **Suppression des actifs dont le nombre de points a toujours été nul (ne présentent pas de risque dans le portefeuille)**
- Suppression de Date_Souscription > Date_sortie
- Suppression de DDN > Date_sortie

FIGURE 2.2 – Contrôles de cohérence

Les contrôles sur la cohérence des dates sont les mêmes. Un choix a été réalisé en interne par avis d'expert qui consiste à dire que les cotisants qui n'ont pas réussi à acquérir de nombre de point après la fermeture du régime en 2016 ont été supprimés de la base puisque ces derniers ne présentent pas de risque dans le portefeuille. De plus, sont principalement concernés les jeunes cotisants qui sont faiblement concernés par le risque de mortalité. L'impact que peut avoir ce choix dans les modélisations sera discuté et évalué par la suite.

Enfin un dernier traitement identique à celui de la base de rentiers est exécuté qui consiste à compléter la date de sortie et de décès des cotisants qui n'ont en pas. On choisit de même une date arbitrairement éloignée, à savoir le 01/01/2040.

Finalement la base retraité des cotisants contient 192 775 observations.

2.1.3 Statistiques générales des cotisants

A présent quelques statistiques descriptives du portefeuille de cotisants sont présentées afin d'en déterminer les caractéristiques. Les mêmes variables que pour le portefeuille des rentiers vont être présentées, à savoir l'exposition, le nombre de décès, le nombre de nouvelle souscription, l'âge moyen, l'âge moyen de souscription et l'âge moyen de décès.

Les statistiques ont été établies sur la période d'observation la plus large possible, c'est-à-dire de 2010 à 2020.

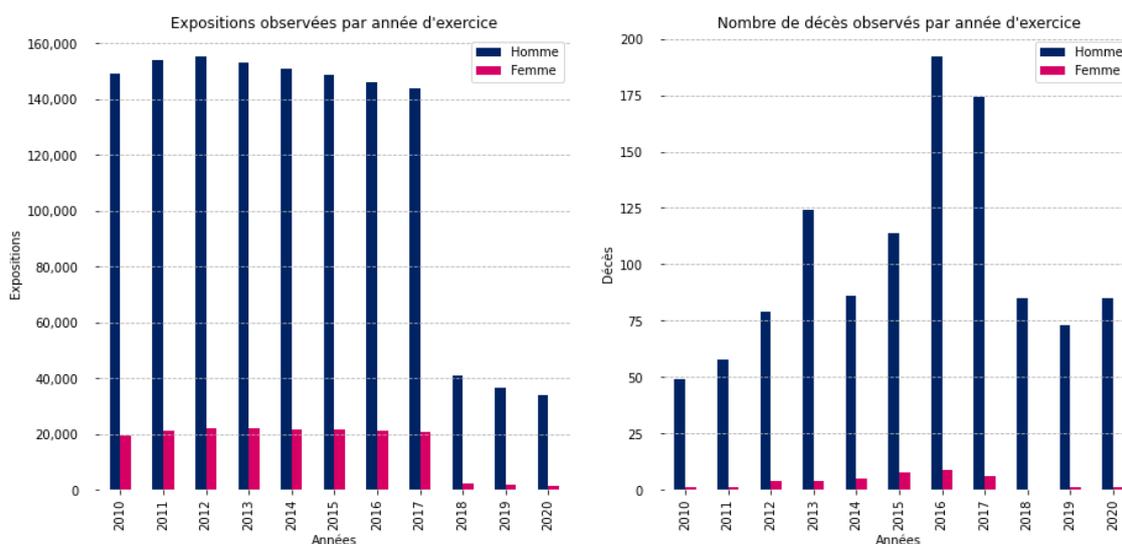


FIGURE 2.3 – Exposition et nombre de décès en nombre par année d'observation

Plutôt que de décrire ce qui est observé sur ces diagrammes en bâton, il serait plus pertinent de souligner les différences et les similarités entre le portefeuille de rentiers et celui des cotisants.

D'une part, le nombre de cotisants est bien plus élevé que celui des rentiers. Cela va donc avoir un impact considérable sur les statistiques globales du portefeuille puisque la pondération des cotisants est bien plus importante que celle des rentiers.

D'autre part, l'effectif du portefeuille est croissant de 2010 à 2012 avant de décliner jusqu'en 2020. De plus un saut entre 2017 et 2018 est observable, il est le résultat de la fermeture du régime en 2016 qui par conséquent va entraîner la sortie d'une population jeune et non négligeable du portefeuille.

De la même manière que chez les rentiers, les femmes représentent un très faible volume par rapport aux hommes chez les cotisants.

Le diagramme en bâton des décès montre que le nombre de décès est croissant malgré une diminution progressive de l'effectif, ceci étant vrai jusqu'en 2017. L'effectif entre 2017 et 2018 est divisé par environ 3.5 alors que le nombre de décès entre 2017 et 2018 est divisé par environ 2.5. Cela signifie que le nombre de décès rapporté à l'effectif, communément appelé le taux brut de mortalité est croissant en fonction du temps chez les cotisants. Cette croissance est justifiée plus tard dans l'étude.

Enfin le nombre de décès chez les femmes est très faible, peut-être trop faible pour pouvoir être utilisé dans d'autres métriques. Ce point sera discuté au moment venu.

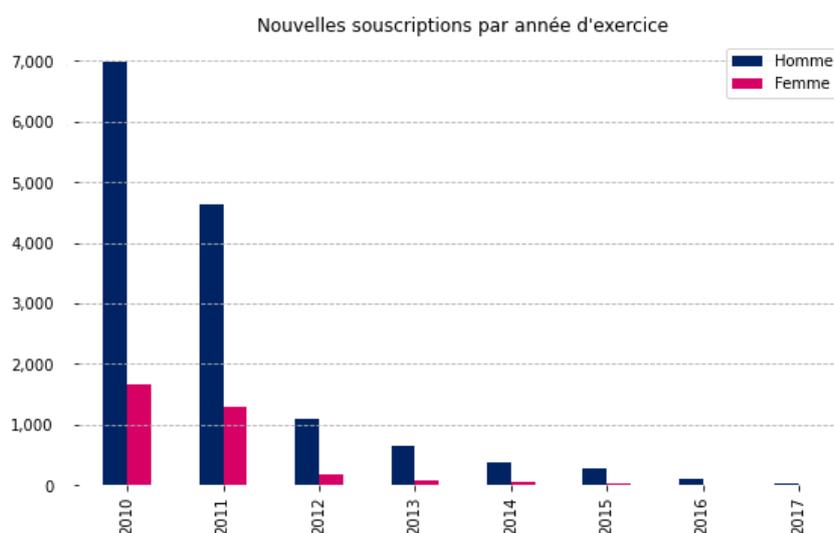


FIGURE 2.4 – Nombre de nouvelles souscriptions par année d'observation

En complément, le nombre de nouvelles souscriptions en fonction des années d'observations est déterminé. L'axe temporel s'arrête en 2017 car il n'y a pas de nouvelles souscriptions à partir de 2017, année suivant la fermeture du régime.

Le nombre de nouvelles souscriptions est décroissant, signe d'une baisse d'attractivité du régime A. On peut se demander pourquoi l'effectif du portefeuille des cotisants est décroissant alors que de nouvelles souscriptions sont comptabilisées. Il y a plusieurs raisons à cela.

D'une part, le nombre de nouvelles souscriptions contribue à augmenter l'effectif et c'est la seule variable qui joue en ce sens, d'autre part il y a également des sorties observées pour plusieurs causes possibles dont l'une étant la migration vers le système de gestion des rentiers, ce sont les jouissances.

Le nombre de nouveaux rentiers étant entre 1 500 et 2 000 individus. Ainsi depuis 2012, plus de jouissances que de nouvelles souscriptions sont observées, ce qui explique que l'effectif du portefeuille des cotisants est en déclin.

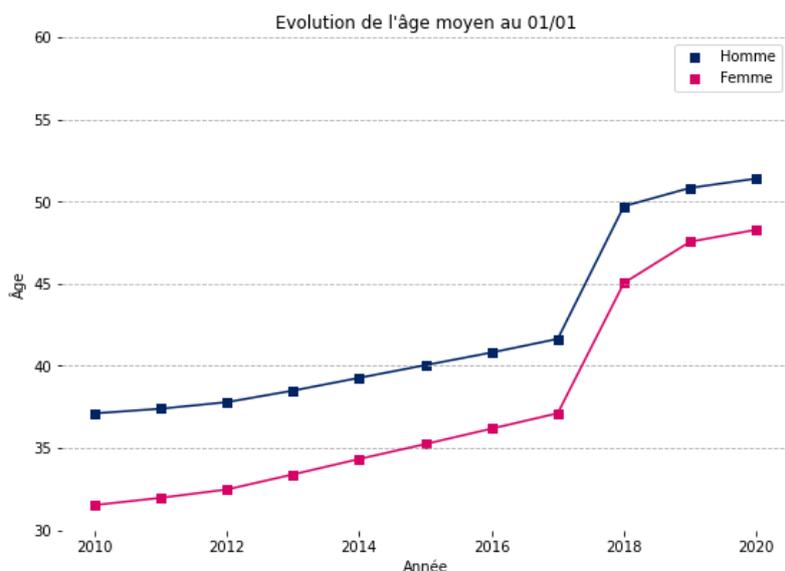


FIGURE 2.5 – Évolution de l'âge moyen en fonction du temps

L'âge des cotisants se situe globalement dans la fourchette de 18 ans à 54 ans. L'âge moyen en 2010 pour les hommes est de 37 ans et croît jusqu'à atteindre 42 ans en 2017, alors que pour les femmes l'âge moyen croît de 32 ans pour atteindre 37 ans en 2020.

Ainsi, une répartition homogène des cotisants sur toute la fourchette d'âges est possible. Ce qui sera vérifié plus tard dans un graphique plus adéquat à l'analyse.

Il démontre également que la moyenne d'âge est plutôt basse, ce qui signifie que ce portefeuille est constitué d'une population jeune, à la fois pour les hommes et pour les femmes.

En 2018, un saut sur l'âge moyen de 42 ans à 50 ans pour les hommes et de 37 ans à 45 ans pour les femmes est constaté. Ceci est dû à la fermeture du régime en 2017. Ce qui signifie que la fermeture du régime a principalement concerné les personnes les plus jeunes du régime, en effet cette conséquence est cohérente avec les caractéristiques intrinsèques du fonctionnement du régime.

Le régime a fermé en 2016 et ne connaît pas de renouvellement. C'est pourquoi, avec la sortie des jeunes cotisants qui n'ont pas atteint le seuil d'ancienneté suffisant pour rester, l'âge moyen du portefeuille depuis 2018 est d'environ 50 ans. La population sera de plus en plus vieillissante dans les années à venir jusqu'à ce que tous les cotisants soient passés en rente.

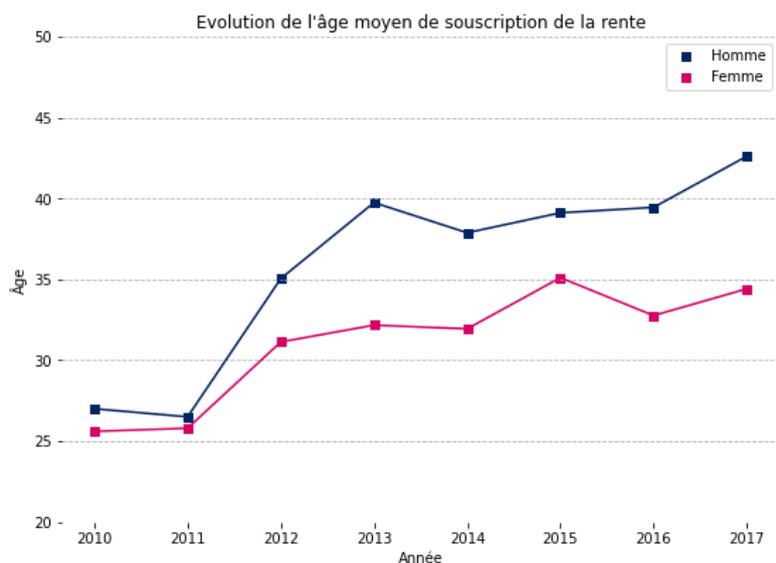


FIGURE 2.6 – Evolution de l'âge moyen de souscription en fonction du temps

L'âge moyen de souscription est en moyenne à la hausse sur la période d'observation, excepté en 2011 et en 2014 pour les hommes. L'âge moyen de souscription varie beaucoup. Il est important de mettre en parallèle ce graphique avec celui du nombre de nouvelles souscriptions par an afin d'expliquer les variations importantes de l'âge moyen de souscription.

En effet, de 2010 à 2012 une faible variation sur l'âge moyen de souscription est constatée. Pour ces années, le nombre de nouvelles souscriptions était respectivement de 8 800 et de 6 000, hommes et femmes confondus. Alors que pour les années postérieures, une forte hausse de l'âge moyen de souscription est observée mais le nombre de nouvelles souscriptions est faible, autour de 1 200 en 2012 puis diminue encore. La volatilité de l'âge moyen de souscription s'explique donc par le faible effectif de données concerné.

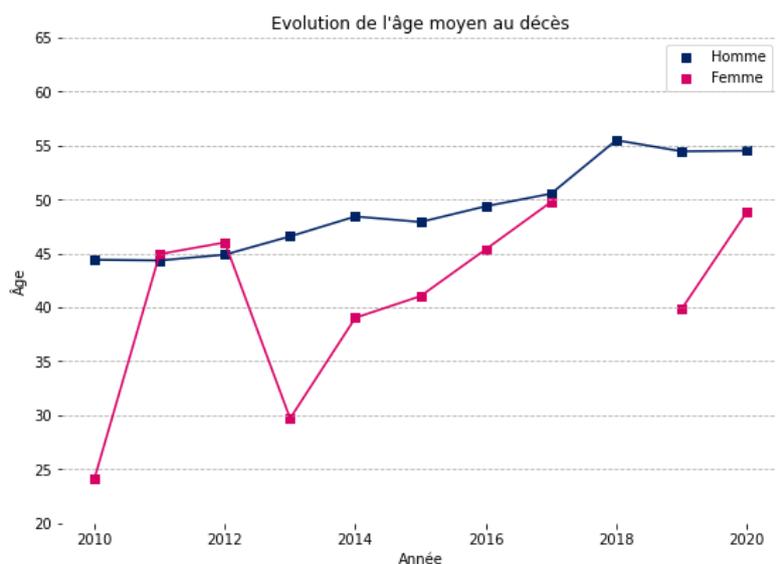


FIGURE 2.7 – Évolution de l'âge moyen aux décès en fonction du temps

Enfin, l'âge moyen aux décès chez les cotisants est présenté. Dans un premier temps, l'âge moyen aux décès chez les femmes est très volatil et même manquant même en 2018 car n'y a pas eu de décès cette année-là.

Par ailleurs, l'âge moyen aux décès est de 45 ans et augmente jusqu'à 54 ans de 2010 à 2020 chez les hommes. Cela signifie qu'il y a des décès en dessous et au-dessus de 45 ans et donc que le risque de décès touche aussi bien les jeunes cotisants que les cotisants plus âgés.

La fermeture du régime en 2016 impacte très peu l'âge moyen aux décès puisque la population restante suite à cet événement est globalement plus âgée qu'avant et donc le risque de décès n'est plus supporté par les jeunes cotisants qui étaient présents dans le régime. C'est ce qui est observé entre 2017 et 2018 où l'âge moyen aux décès connaît une hausse plus nette que pour les années antérieures.

En résumé :

Le portefeuille de cotisants a été très impacté par la fermeture du régime en 2016 contrairement au portefeuille des rentiers. L'effectif en nombre a connu une décroissance progressive puis une chute brutale depuis la fermeture du régime, en revanche le nombre de décès a connu une hausse puis une légère baisse depuis la fermeture du régime.

Par ailleurs, l'âge moyen du portefeuille des cotisants informe que le portefeuille est très jeune et réparti de manière homogène. Le renouvellement du portefeuille s'est arrêté progressivement. Enfin le risque de décès touche aussi bien les jeunes cotisants que les cotisants plus âgés dans ce portefeuille.

2.2 Système de gestion des rentiers

2.2.1 Récupération des données

Les données utilisées dans cette étude proviennent de différentes sources.

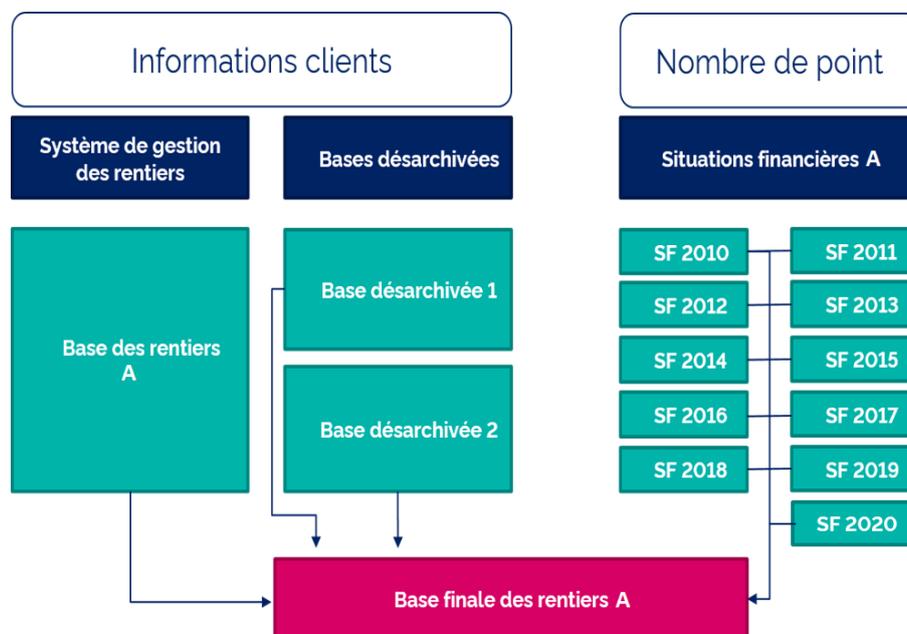


FIGURE 2.8 – Schéma descriptif des sources de données

En effet, une partie des données est issue du système de gestion consacré aux rentiers, cette base est mise à jour chaque année et fournit donc les informations les plus récentes sur les assurés.

Par ailleurs, une partie des données provient des bases désarchivées, elles contiennent les contrats clôturés avant 2016. Elles vont donc servir à enrichir la base des rentiers afin d'avoir l'ensemble des têtes assurées de 2010 à 2020. Ces deux bases sont donc fusionnées afin de simplifier l'enrichissement de la base principale par les données désarchivées.

Enfin, l'utilisation des SF produites chaque année est une photographie de l'état du régime à chaque instant annuel t . Les SF permettent de réaliser différents contrôles afin de vérifier l'exactitude des données mais également de repérer d'éventuelles informations manquantes ou erronées qui seront remontées à la gestion afin d'être corrigées.

Le système de gestion fournit une liste de variables sur les contrats du régime A qui est la suivante :

- Client compte
- Id client
- Identifiant
- Sexe
- Code de sortie
- Date de naissance
- Date de souscription
- Date de jouissance
- Date de décès
- Code intervenant
- Code état
- Code rente
- Date sortie

Concernant le périmètre d'étude, ces variables ne seront pas toutes conservées puisque seules quelques variables sont étudiées comme la mortalité des individus, l'exposition, l'âge, le sexe. Toutefois, ces variables sont essentielles afin de réaliser différents types de contrôles sur la cohérence des données. Ce sujet sera traité dans la partie suivante.

Remarque : lors de l'importation des données, il faut être extrêmement vigilant sur l'importation de la date au bon format. Il peut arriver que les dates soient au format *yyyy/mm/dd* et que l'importation des dates se fasse au format *yyyy/dd/mm*.

Afin d'enrichir la base de données des rentiers issue du système de gestion, une clé primaire est à déterminer afin de pouvoir identifier les contrats doublons dans la base des rentiers et dans la base désarchivée (fusion des deux bases désarchivées). La clé primaire choisie est Id client. Celle-ci permet de concaténer les deux bases en supprimant les doublons, comme présenté dans le schéma ci-dessous.

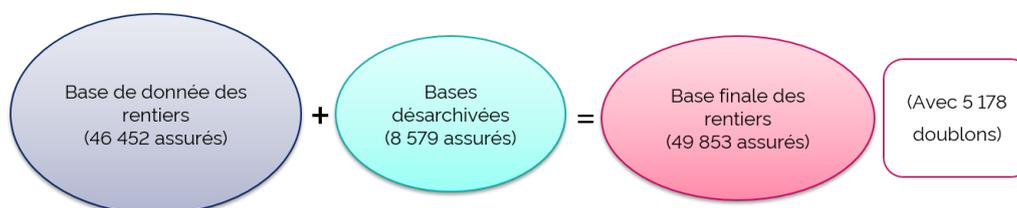


FIGURE 2.9 – Fusion des bases de données

2.2.2 Traitement des données

Le premier contrôle réalisé porte sur l'exhaustivité des données, pour cela l'utilisation des SF de 2010 à 2019 est requise.

Avant toute chose, de quoi est constituée une SF et comment permet-elle de vérifier l'exhaustivité des données ?

Une SF est une base de données contenant tous les assurés présents dans le régime au 31/12/N. Celle-ci sépare les rentiers et les cotisants. Ainsi en comparant deux SF d'années consécutives, le nombre d'assurés sortant du régime pour cause de décès ou autre peut en être déduit ainsi que le nombre de nouveaux rentiers arrivant depuis la base des cotisants.

Par ailleurs, le nombre de points acquis par chaque assuré l'année N est inscrit dans la SF, ce qui permet de suivre l'évolution de point en comparant les SF entre elles. Toutefois, une fois que l'assuré est devenu rentier son nombre de points n'évolue plus et reste identique durant toute sa retraite.

Une des difficultés rencontrées ici est le format des données et la donnée disponible dans les SF. En effet entre 2010 et 2019, plusieurs intervenants ont été sollicités afin de construire les SF. De ce fait, tantôt l'Id client est donné afin de reconnaître les assurés, tantôt c'est l'Identifiant du compte qui est donné. Ceci a engendré des difficultés d'ordre technique mais n'a en rien empêché la bonne exécution du contrôle d'exhaustivité qui est validé par toutes les SF.

Une fois que la base de rentiers est jugée complète, la réalisation des contrôles de cohérence, de qualité et d'exhaustivité des données est nécessaire. Les contrôles suivants sont donc réalisés :

Contrôle et suppression des données aberrantes :

- Contrôle du sexe (remplacer les 3 -> Mlle par des 2 -> Mme)
 - Il reste cependant la modalité « X » dans sexe provenant de la base désarchivée
- Suppression de DDN > Date_Souscription
- Suppression de Date_sortie < 2005
- Suppression de DDN > Date_décès
- Suppression de Date_sortie < Date_décès
- Suppression de Date_Souscription > Date_décès
- Suppression de Date_Souscription > Date_sortie
- Conservation du Code_intervenant = 3
 - La Modalité « X » dans Sexe a été supprimé par la même occasion
- Suppression de DDN > Date_sortie

FIGURE 2.10 – Contrôles de cohérence

Pour certains contrats de la base désarchivée, la variable sexe a été perdu et donc a été remplacée par un "X". Cependant, la totalité de ces contrats a un code intervenant égal à 32. Ces contrats ne sont pas conservés dans l'étude, ce qui résout le problème de non-connaissance du sexe pour les contrats dont le sexe est "X".

D'une part, il existe des contrats pouvant faire appel à un réservataire, ceux-ci sortent du domaine d'étude. Pour se débarrasser de cette contrainte, un filtre est appliqué sur le code intervenant pour uniquement garder la première tête, d'où la conservation du Code intervenant = 3.

D'autre part, l'incohérence de certaines informations comme une date de souscription plus grande qu'une date de sortie tient son origine des erreurs de saisie de la donnée. Les informations incohérentes ont été remontées à la gestion qui a pu corriger une partie d'entre elles, les données qui n'ont pas pu être corrigées ont tout simplement été écartées de l'étude pour des raisons de qualité de données.

Enfin, pour éviter les erreurs dues aux données manquantes lors de la configuration des modèles, une technique est utilisée, celle de remplir artificiellement les dates de sortie et les dates de décès des assurés en vie et encore présents dans le portefeuille. Le choix d'une date arbitrairement éloignée de la période d'observation est faite, à savoir le 01/01/2040.

La base finale des rentiers ne contient plus que 32 541 lignes.

2.2.3 Statistiques générales des rentiers

Pour avoir une idée générale des caractéristiques du portefeuille des rentiers, quelques statistiques descriptives sont présentées. L'analyse de ces résultats permet d'orienter les choix de modèles, les choix de paramètres mais aussi de comprendre quelles sont les limites et les faiblesses du portefeuille dans la projection de la mortalité future.

Les statistiques ont été réalisées sur l'ensemble du portefeuille des rentiers sans limite d'âge et sur la période de 2010 à 2020.

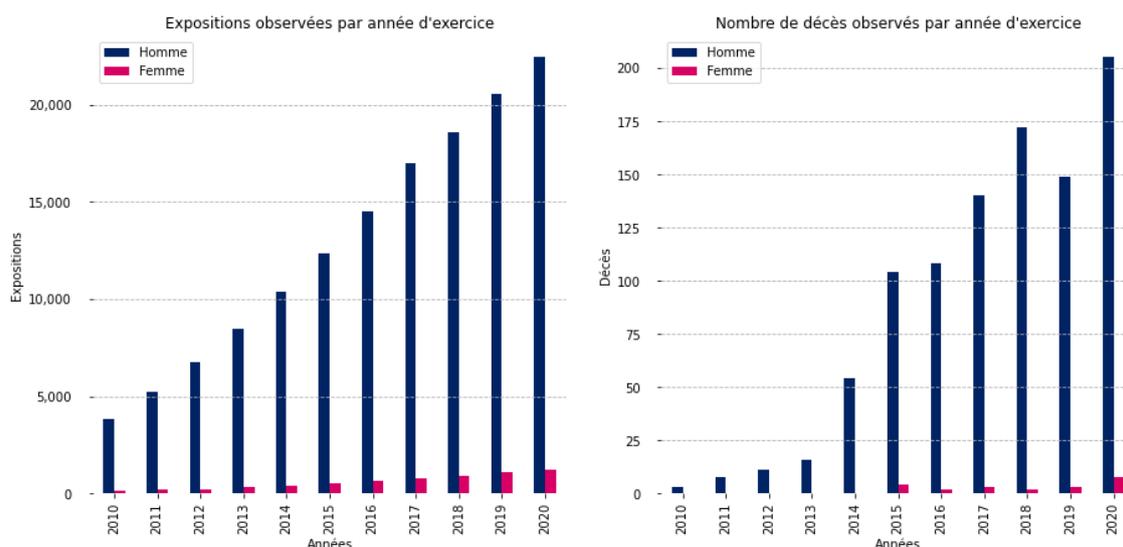


FIGURE 2.11 – Exposition et décès en nombre par année d'observation

CHAPITRE 2. CONSTRUCTION DE LA BASE DE DONNÉES

Entre 2010 et 2020, l'effectif du portefeuille croît de manière constante en commençant par une exposition de 5 000 en 2010 et atteint un effectif d'environ 23 000 en 2020. Cela signifie qu'il y a plus de nouveaux entrants en rente que de sortants, ainsi le portefeuille est constamment renouvelé entre 2010 et 2020.

Le portefeuille de rentiers est majoritairement constitué d'hommes et très peu de femmes. Ce constat est dû au fait que la population assurée concerne un corps de métier spécifique qui attire généralement les individus de sexe masculin.

Le graphique comptabilisant le nombre de décès par année d'observation présente quant à lui des irrégularités qu'il faut pouvoir expliquer tout en tenant compte de l'exposition associée. Dans un premier temps, le nombre de décès connaît en moyenne une tendance à la hausse entre 2010 et 2020. Cette allure à la hausse est justifiée par le fait que l'effectif est à la hausse également donc plus de personnes sont soumises au risque de décès.

Cependant, en 2014 une forte hausse de mortalité est observable, en effet le nombre de décès double par rapport à 2013 et ceci se produit en 2015 également. La cause de cette irrégularité ne provient pas de la hausse de l'effectif pour les deux années concernées puisque l'exposition croît de façon stable, par ailleurs il n'y a rien à signaler sur le fonctionnement du régime puisqu'il n'y a eu aucun changement majeur sur les années concernées. En revanche, le portefeuille connaît un vieillissement régulier ce qui accroît le risque de décès global du portefeuille, cette piste est à vérifier avec l'évolution de l'âge moyen du portefeuille.

Entre 2016 et 2020, le nombre de décès augmente de façon plus ou moins stable avec en 2019 un nombre de décès moindre qu'en 2018 malgré une exposition plus importante. Cette baisse est compensée dès l'année suivante puisqu'une hausse de 25% est constatée entre 2019 et 2020.

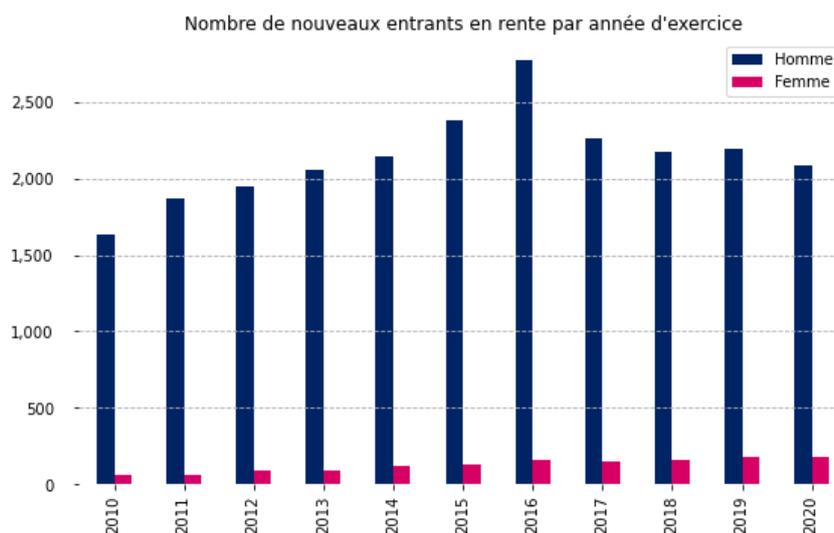


FIGURE 2.12 – Nombre de jouissances par année d'observation

En complément, le nombre de nouveaux entrants en rente par an est illustré à travers le graphique précédent. Cet histogramme permet de confirmer la régularité du renouvellement du portefeuille de rentiers, d'autant plus que le nombre de décès est négligeable par rapport au nombre de nouveaux entrants en rente.

A partir de 2016 le nombre de nouveaux entrants décroît et se stabilise autour de 2 000. Cela peut être dû à la fermeture du régime en 2016 qui a pour conséquence le non-renouvellement des cotisants et donc une population proche de l'âge pivot en déclin.

Cependant l'impact provoqué par la fermeture du régime n'est pas immédiat et devrait se confirmer sur le long terme lorsque l'âge moyen du portefeuille dans son ensemble se rapprochera de l'âge pivot.

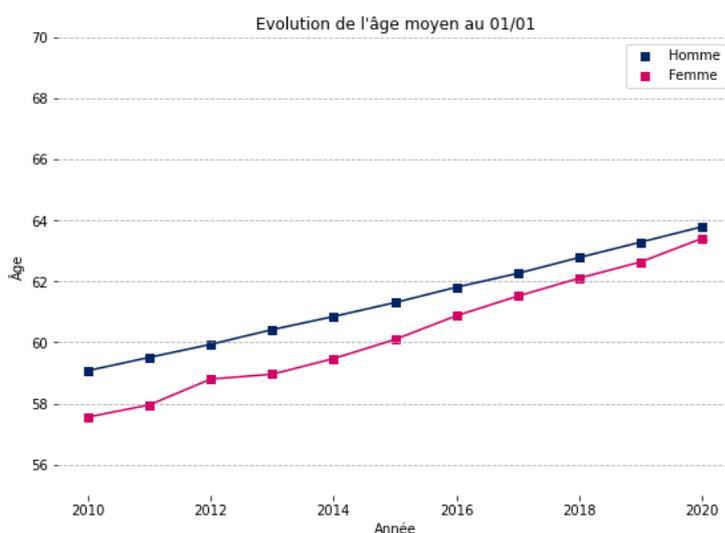


FIGURE 2.13 – Évolution de l'âge moyen en fonction du temps

L'âge pivot est défini comme l'âge minimum de départ à la retraite pour ce régime.

L'âge pivot dans du régime A est de 55 ans. En 2010, l'âge moyen est de 59 ans pour les hommes et 57,5 ans pour les femmes. Cela signifie qu'il y a une forte concentration des rentiers autour de 59 ans.

Par ailleurs, la variance autour de l'âge moyen est faible en 2010. Toutefois l'âge moyen des rentiers croît en fonction des années, de plus chaque année une vague de jeunes rentiers (55 ans) entre dans le portefeuille et pour autant l'âge moyen continue de croître. Ce qui signifie que la variance augmente et que la répartition des rentiers s'étale sur une tranche d'âge de plus en plus large.

Finalement en 2020, l'âge moyen du portefeuille de rentiers est de 64 ans pour les hommes et 63 ans pour les femmes.

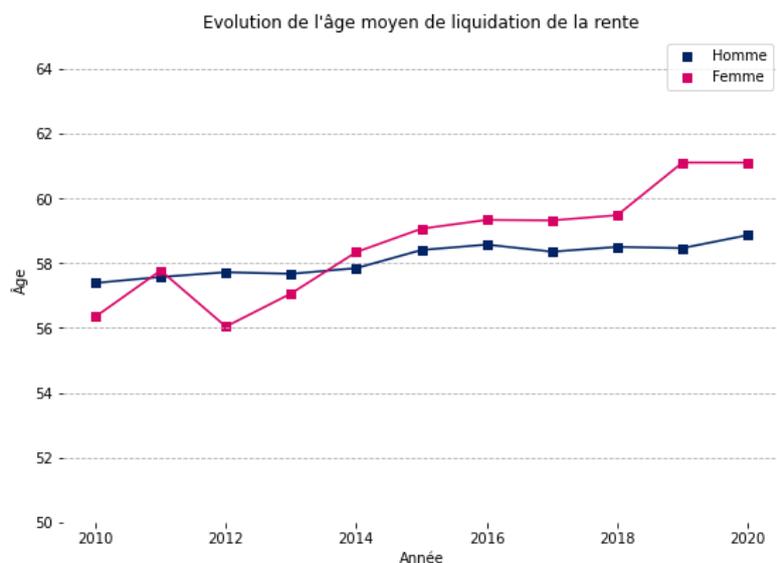


FIGURE 2.14 – Evolution de l'âge moyen de liquidation en fonction du temps

L'âge moyen de liquidation ajoute une précision supplémentaire sur les nouveaux entrants en rente. Cette information doit être analysée car sans celle-ci le constat aurait été que les nouveaux rentiers arrivent dans le portefeuille à l'âge de 55 ans comme cité précédemment. Or en moyenne l'âge de liquidation en rente est plutôt autour de 58 ans pour les hommes et entre 56 et 61 ans pour les femmes.

Il semble donc qu'en moyenne les cotisants du régime décident de partir à la retraite non pas à l'âge minimal autorisé par le régime mais plutôt quelques années après. La volatilité de l'âge de liquidation chez les femmes est bien différente de celle des hommes, ceci s'explique du fait que le volume de femme, c'est à dire l'exposition (moins d'une centaine) du portefeuille est faible par rapport à celui des hommes (plusieurs milliers, voir dizaines de milliers).

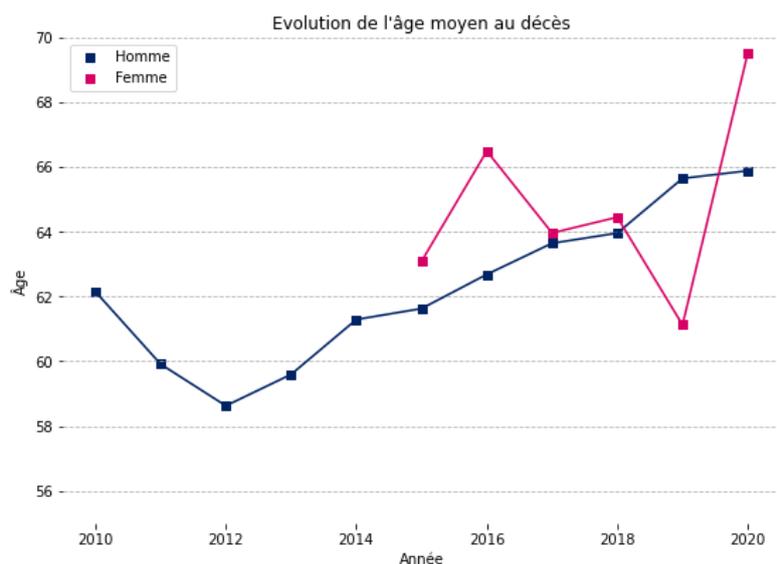


FIGURE 2.15 – Evolution de l'âge moyen aux décès en fonction du temps

CHAPITRE 2. CONSTRUCTION DE LA BASE DE DONNÉES

Une autre statistique importante qu'il ne faut pas oublier de considérer est l'âge moyen aux décès chez les hommes et les femmes en fonction des années. En effet le nuage de point pour les femmes ne commence qu'en 2015, puisque d'après le graphique présenté précédemment qui comptabilise le nombre de décès annuel par sexe, aucun décès chez les femmes avant 2015 n'est comptabilisé.

D'autre part, une forte volatilité sur l'âge moyen aux décès chez les femmes est constatée, cela est dû au faible nombre de décès (moins d'une dizaine pour chaque année d'observation) qui fait que la sensibilité à l'âge moyen aux décès est très forte et peut varier énormément au moindre décès supplémentaire.

En ce qui concerne l'âge moyen aux décès pour les hommes, une décroissance de 2010 à 2012 puis une croissance stable de celle-ci jusqu'en 2020 est observable. Ce phénomène peut être expliqué par la croissance de l'âge moyen du portefeuille.

En effet, le portefeuille est constitué d'une population vieillissante et donc de plus en plus assujéti au risque de décès. C'est pourquoi non seulement de plus en plus de décès sont observés et surtout des décès survenant à des âges de plus en plus élevés.

En résumé :

Le portefeuille de rentiers du régime A connaît une croissance stable en effectif car le nombre de nouvelles jouissances est stable (autour de 2000) malgré l'évolution croissante du nombre de décès car celui-ci est négligeable (de l'ordre de la centaine) devant le nombre de nouvelles jouissances. Le portefeuille est majoritairement constitué d'homme. Par ailleurs, le nombre de décès connaît une hausse irrégulière avec une volatilité plus ou moins importante entre 2010 et 2020.

L'âge moyen du portefeuille connaît une croissance régulière de 59 à 64 ans, ce qui signifie que les rentiers sont de plus en plus âgés. L'âge moyen de liquidation est stable est vaut environ 58 ans. Enfin, l'âge moyen aux décès passe de 62 à 66 ans en passant par un minimum à 59 ans en 2012.

2.3 Base finale

2.3.1 Concaténation des bases

Après avoir récupéré et traité les données depuis les différentes sources de données, le but est de créer une seule et unique base de données qui constituera le principal fichier de travail.

De ce fait avant de concaténer la base des rentiers et la base des cotisants, des contrôles doivent être effectués afin de ne pas retrouver des informations redondantes. Il s'agit principalement des cotisants proches de l'âge pivot de retraite qui au cours des années d'observations ont migré du côté retraite.

On s'attend à ce qu'une grosse partie des retraités se retrouve dans la base des cotisants puisque qu'en 2010, le volume des retraités était très faible et que celui-ci n'a fait qu'augmenter jusqu'en 2020.

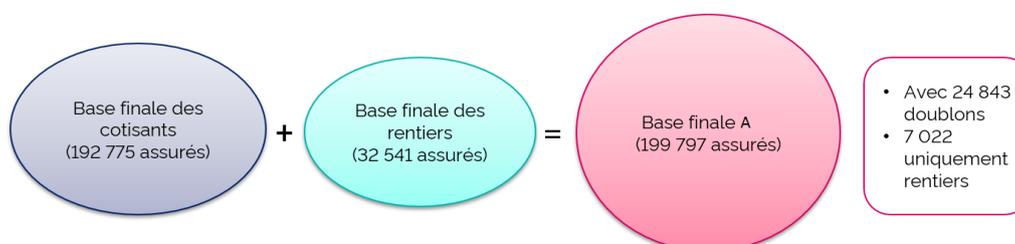


FIGURE 2.16 – Fusion des bases rentiers et cotisants

La base d'étude finale du régime A contient donc 199 797 observations.

Par ailleurs lors de la concaténation, des contrôles ont dû être effectués car les informations sont parfois différentes pour les mêmes assurés qui proviennent à la fois de la base des cotisants et de la base des rentiers.

Contrôle et modification des observations communes:

- Pas les mêmes Date_souscription
 - On conserve les Date_souscription de Loic
- Pas le même nombre de points
 - On conserve le nombre de points de Loic
- Pas les mêmes Date_deces
 - On conserve les Date_deces de Laure (données sans erreur de de gestion)
- Pas les mêmes Date_sortie
 - On conserve la Date_sortie de Laure car celle de Loic correspond à la Date_jouissance

FIGURE 2.17 – Contrôles de cohérence

La base Loic est la base des cotisants et la base Laure est la base des retraités.

Les statistiques générales de la base des rentiers et de la base des cotisants ont été présentées dans les parties précédentes. A présent, une présentation des statistiques générales du portefeuille entier va être faite.

2.3.2 Statistiques générales

Dans cette partie les mêmes statistiques générales sont présentées, à savoir l'exposition et le nombre de décès par année, puis l'évolution de l'âge moyen du portefeuille, l'âge moyen aux décès et également la répartition par sexe en fonction des années.

Les statistiques observées dans cette partie sont en principe intermédiaires en fonction de la pondération des bases initiales à celles obtenues dans les parties précédentes.

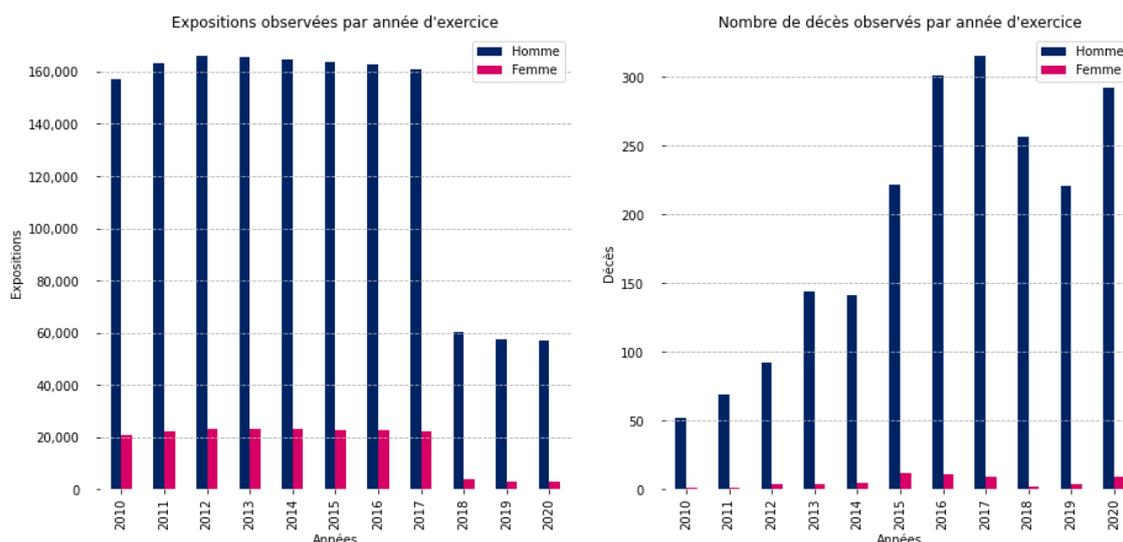


FIGURE 2.18 – Exposition et nombre de décès en nombre par année d'observation

L'effectif du portefeuille tourne autour de 180 000 assurés, hommes et femmes confondus entre 2010 et 2017, puis l'effectif diminue et passe à 60 000 en 2018 conséquence de la fermeture du régime comme expliqué dans la partie précédente.

Concernant le nombre de décès, une tendance globale à la hausse est constatée en fonction des années, conséquence du vieillissement du portefeuille ce qui va d'ailleurs entraîner une exposition plus forte au risque de décès pour l'ensemble du portefeuille.

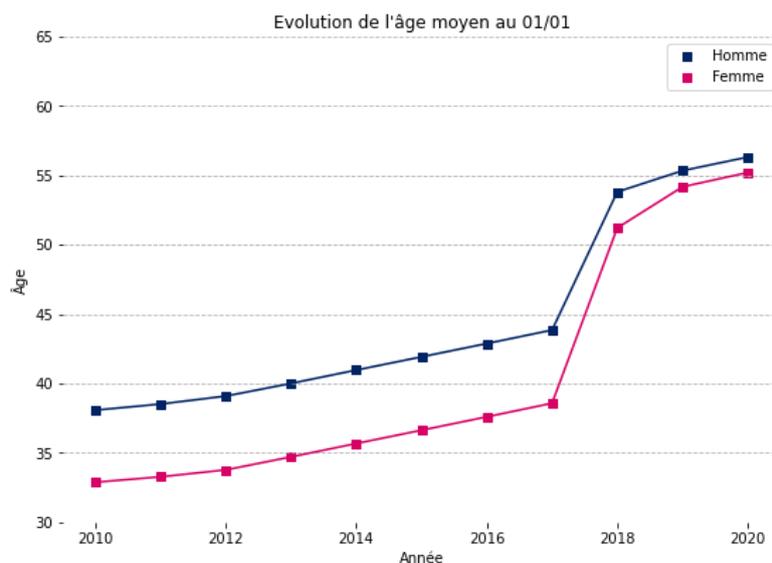


FIGURE 2.19 – Evolution de l'âge moyen en fonction du temps

Le vieillissement du portefeuille peut s'observer à partir de l'évolution de l'âge moyen du portefeuille. En effet, celui-ci croît au cours du temps et atteint en 2020 un âge moyen de 56 ans pour les hommes et 55 ans pour les femmes.

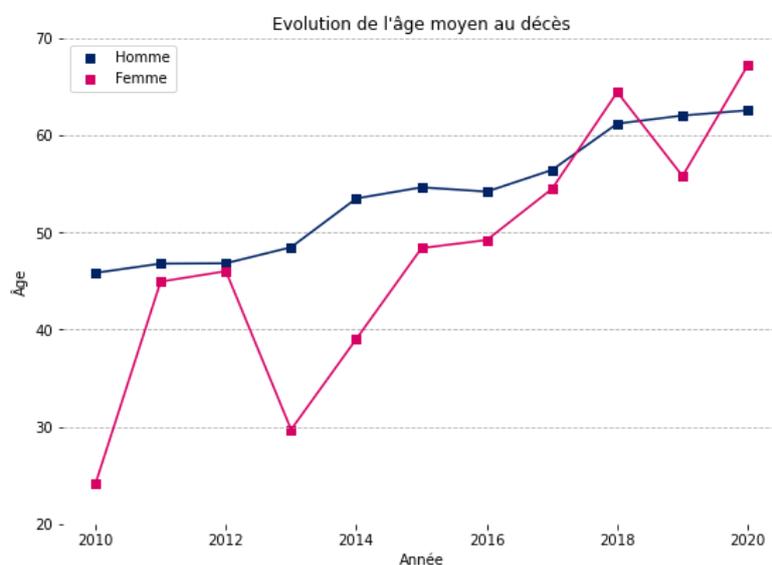


FIGURE 2.20 – Evolution de l'âge moyen aux décès en fonction du temps

En ayant fusionné les bases de rentiers et de cotisants l'âge moyen aux décès est situé entre ceux observés pour chacune des deux bases initiales. L'âge moyen aux décès est alors de 46 ans en 2010 et atteint progressivement un âge de 63 ans en 2020 pour les hommes.

Concernant l'âge moyen aux décès chez les femmes, celui-ci est très volatile pour les premières années, puis se stabilise à partir de 2015 mais reste tout de même bien plus volatile que l'âge moyen aux décès des hommes.

Pour expliquer la forte volatilité des statistiques chez les femmes, il faut s'intéresser à la proportion par sexe du portefeuille accompagnée de l'effectif en nombre des femmes du portefeuille.

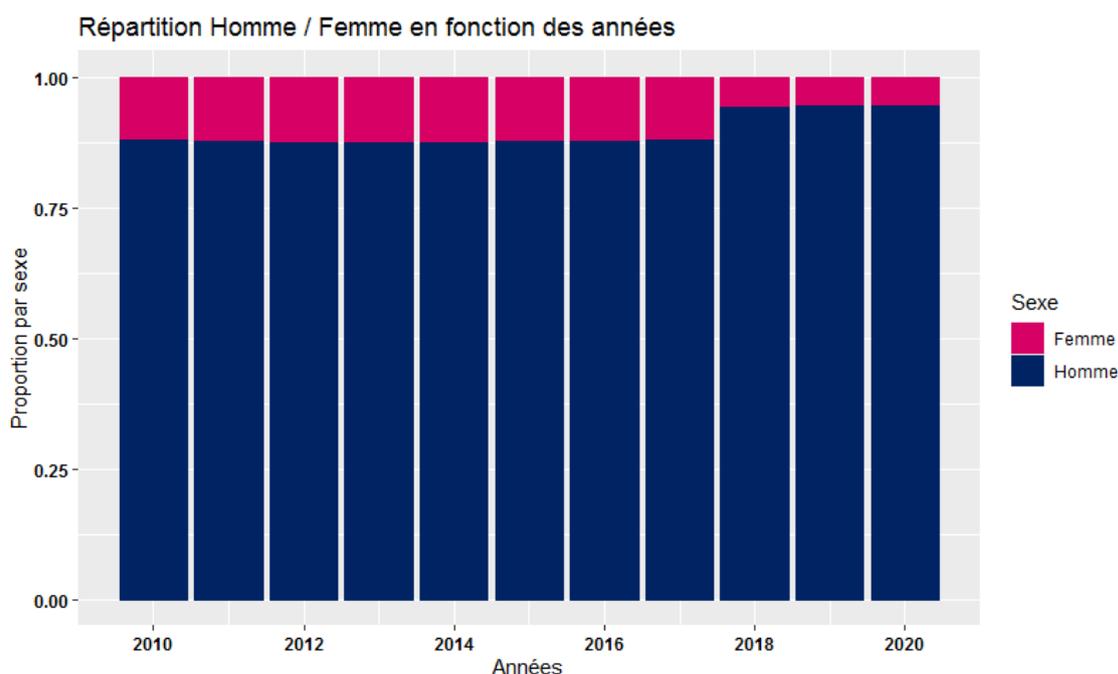


FIGURE 2.21 – Répartition Homme/Femme par année d'observation

Sans surprise, le volume de données que représente les femmes reste petit par rapport à celui des hommes, cette observation a été faite lors de l'analyse du portefeuille des rentiers et des cotisants.

Ceci est confirmé par le diagramme en bâton de la répartition par sexe du portefeuille en fonction des années. On observe en effet que la proportion de femme est toujours inférieure à 15% durant les années d'observations et que depuis la fermeture du régime cette proportion a encore baissé et représente in fine moins de 10%.

De plus d'après le diagramme en bâton des expositions en nombre par année d'observation, le volume des femmes est faible de 2010 à 2017 représentant un effectif d'environ 20 000 femmes. Puis celui-ci chute à moins de 5 000 depuis la fermeture du régime.

Enfin, le très faible nombre de décès des femmes ne fait qu'accentuer la volatilité des résultats, ce qui se remarque dans le graphique de l'âge moyen aux décès.

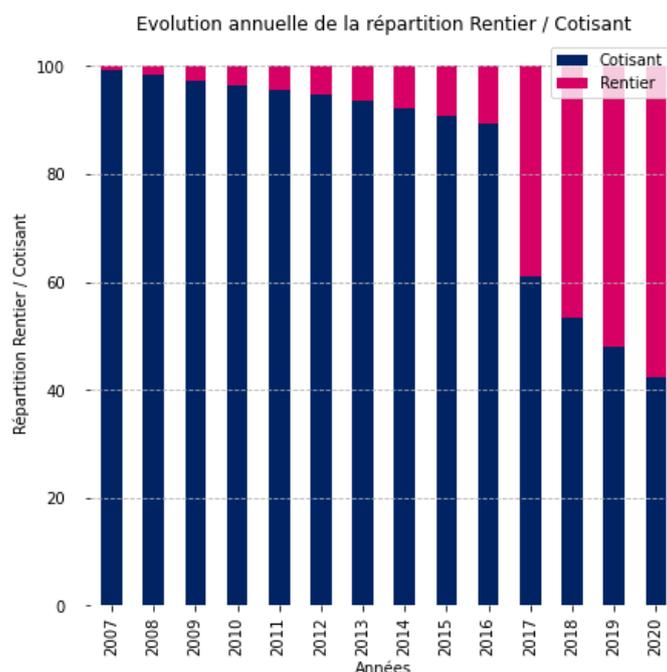


FIGURE 2.22 – Répartition Homme/Femme par année d’observation

Une autre donnée peut être analysée, il s’agit de la répartition du portefeuille entre les cotisants et les rentiers. Celle-ci s’observe avec une dimension temporelle à travers ce graphique.

La fermeture du régime se remarque en 2017 en observant que la répartition des rentiers passe de 11% à un peu moins de 40%. Ce qui est cohérent avec le fait que les jeunes cotisants soient exclus du régime cette année-là.

En résumé :

Le portefeuille complet présente des caractéristiques voisines du portefeuille des cotisants puisque celui-ci regroupe une majorité des assurés du portefeuille global.

Il faut retenir que le portefeuille global passe de 180 000 assurés à 60 000 après la fermeture du régime et que le nombre de décès augmente au cours du temps malgré la baisse brutale de l’effectif. Par ailleurs, l’âge moyen du portefeuille chez les hommes croît de 36 ans à 56 ans en 10 ans, celui des femmes croît de 33 à 55 ans.

L’âge moyen aux décès chez les hommes croît également de 46 ans à 62 ans, celui des femmes et extrêmement volatile et fluctue entre 24 ans et 68 ans.

Enfin cette volatilité chez les femmes est expliquée par le faible volume que représente les femmes au sein du portefeuille, ce qui amène de l’imprécision dans les données, cependant la part des femmes au sein du portefeuille est faible, elle représente 15% au départ et passe à 8% après la fermeture du régime.

Chapitre 3

Problématique et réflexions spécifiques au portefeuille

Après avoir analysé les caractéristiques du portefeuille, différentes problématiques surgissent quant à la calibration du domaine d'application. Ce chapitre permet donc d'une part de formuler les problématiques spécifiques au portefeuille et d'autre part de tenter d'y apporter des éléments de réponse afin de pouvoir définir le domaine de calibrage de la façon la plus cohérente possible tout en prenant en compte les risques associés.

3.1 Réflexions compte tenu des caractéristiques du portefeuille

Le but de cette partie est de prendre conscience que la qualité des données représente un enjeu crucial et déterminant sur la qualité des estimations et des modèles.

Pour établir des estimations de taux bruts de mortalité, il est nécessaire d'avoir un portefeuille d'assurés avec dans l'idéal une exposition conséquente, ceci à tout âge.

L'intérêt d'avoir une quantité de données volumineuse est de pouvoir observer les sinistres, à savoir les décès en ce qui concerne cette étude. En général, la volonté d'avoir un volume important de données n'est pas forcément nécessaire pour des âges élevés puisque le risque de décès est lui-même élevé ce qui conduit à observer de fait un nombre de décès suffisant pour avoir des estimations cohérentes.

En revanche, le besoin d'un gros volume de données pour des âges jeunes est absolument crucial car le risque de décès est faible ce qui conduit naturellement à observer beaucoup moins de décès.

Le problème se trouve également pour des âges très élevés où cette fois la probabilité de décès est très forte mais le volume des individus à ces âges est en général très faible voir absent pour tout portefeuille. Ceci conduit parfois à des résultats erronés.

De ce fait si la quantité de données n'est pas suffisante pour certains âges, cela peut conduire à des estimations aberrantes, très éloignées de la réalité. Il faut donc être extrêmement vigilant sur le calibrage choisi.

CHAPITRE 3. PROBLÉMATIQUE ET RÉFLEXIONS SPÉCIFIQUES AU PORTEFEUILLE

Ainsi quels sont les risques qu'il faut prendre en considération compte tenu des caractéristiques du portefeuille ?

Une des caractéristiques du portefeuille qu'il faut intégrer dans les réflexions est la jeunesse du portefeuille. L'âge moyen est en dessous de 40 ans les premières années d'observations.

De plus le nombre de décès durant les premières années est très faible, cela coïncide avec ce qui a été énoncé précédemment.

En revanche plus les années passent et plus l'âge moyen augmente et plus le nombre de décès augmente malgré une exposition qui stagne, voir qui diminue. Une telle observation peut inciter à considérer la valeur de l'information obtenue de manière différente en accordant plus ou moins d'importance à celles-ci.

Ainsi plus l'âge moyen du portefeuille est élevé et plus l'information doit être significative et donc avoir une masse importante. Cependant cette loi doit s'appliquer toutes choses étant égales par ailleurs, à savoir principalement l'exposition puisque celle-ci accorde du crédit donné à l'information.

Toutefois, de 2010 à 2017 le poids de l'information est croissant puisque l'effectif est constant mais à partir de 2017, l'effectif total est divisé par 3 mais l'âge moyen du portefeuille fait un saut de 10 ans. Donc la baisse de l'effectif fait diminuer le poids de l'information des années concernées mais la hausse de l'âge moyen fait augmenter le poids de cette information ce qui fait contrebalancer la baisse du poids de l'information. Il y a ici une loi à construire afin de pondérer le poids de l'information en fonction de son importance.

Le choix de construire cette loi sur la pondération de l'importance sera retenu dans la suite de l'étude si elle semble pertinente.

En outre d'autres paramètres sont à considérer et qui méritent réflexions. Il s'agit entre autre de la période d'observation et de la tranche d'âge à retenir pour le bon calibrage des estimations. Ces deux paramètres sont étudiés dans les parties suivantes.

3.2 Choix des années d'observation à retenir compte tenu de la fermeture du régime

Le choix des années d'observation est un choix déterminant dans les modélisations. En effet, c'est un paramètre qu'il ne faut surtout pas négliger puisqu'il va influencer les résultats. L'usage préconise de choisir la période d'observation la plus large possible à condition d'avoir des données justes et vérifiées. Les données vérifiées de 2010 à 2020 sont mises à disposition.

Pour des raisons de délais de traitement par les systèmes de gestion, le choix qui est fait est d'exclure l'année 2020 de la période d'observation car il se pourrait que les données ne soient pas complètes ou même influencées par le COVID19. La période d'observation devient 2010 à 2019.

Dans la partie précédente, un histogramme illustrant l'exposition et le nombre de décès observé par année d'observation a été présenté.

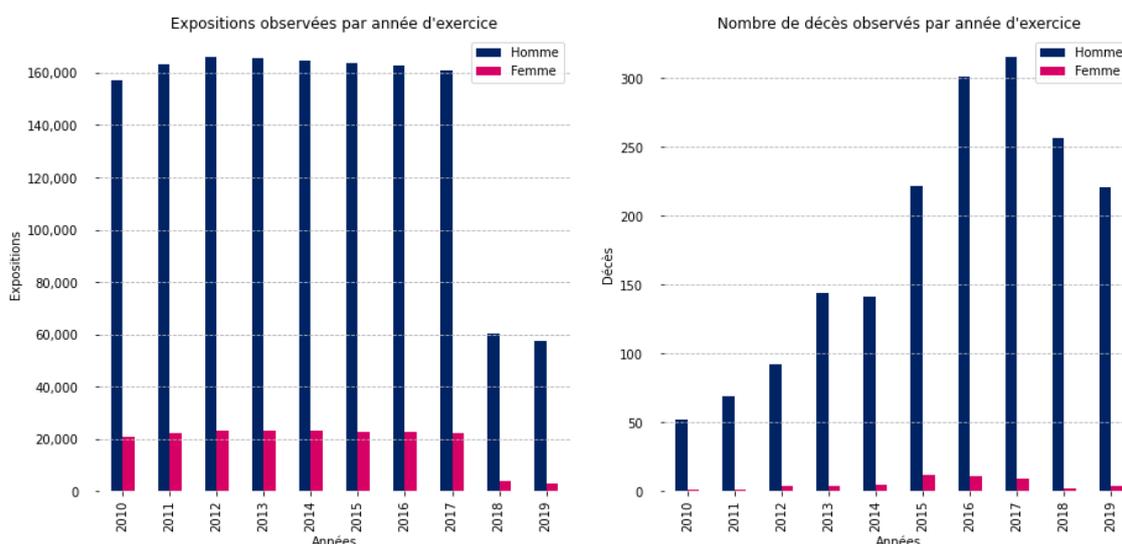


FIGURE 3.1 – Exposition et décès du portefeuille par année d'observation

Les années 2010, 2011 et 2012 présentent peu de décès relativement aux restes des années d'observation. Une question peut se poser, celle de la pertinence des informations apportées par ces années d'observation, d'autant plus qu'en 2010 le portefeuille est encore relativement jeune, donc peu assujéti au risque de mortalité.

Après réflexion, le choix pris est d'exclure les années 2010 et 2011 pour les raisons évoqués au-dessus, cependant l'année 2012 est conservée afin de tout de même avoir 8 années pour la période d'observation.

Finalement, les modèles seront calibrés sur la période d'observation 2012 – 2019.

3.3 Choix de la plage d'âge à retenir

La répartition des âges n'est pas homogène dans le portefeuille d'après la partie précédente. En effet, la tranche d'âge au-delà de 70 ans est quasi nulle en effectif pour les années d'observation de 2012 à 2019 compte tenues de la jeunesse du portefeuille.

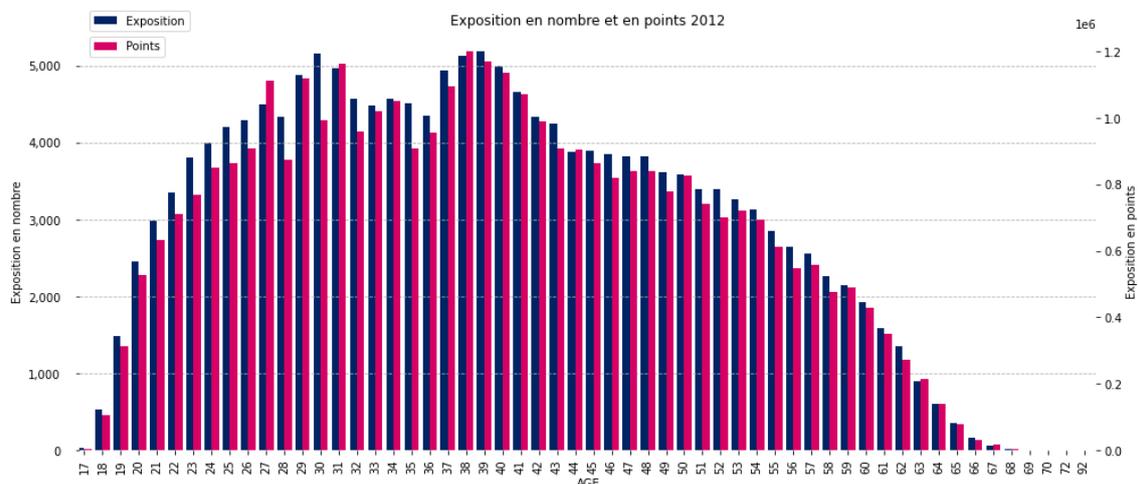


FIGURE 3.2 – Exposition par âge en 2012

D'autre part, une concentration soutenue des effectifs de population est relevée pour les âges allant de 20 à 55 ans pour l'année d'observation 2012. Cette tendance est relativement identique entre 2012 et 2017. Le portefeuille est donc fortement représenté par une population jeune dont la mortalité est très faible pour la tranche d'âge considérée.

Malgré une concentration relativement forte du portefeuille pour la tranche d'âge 20 – 55 ans, le volume de données n'est pas suffisant pour considérer les résultats comme significatifs pour les âges les plus faibles.

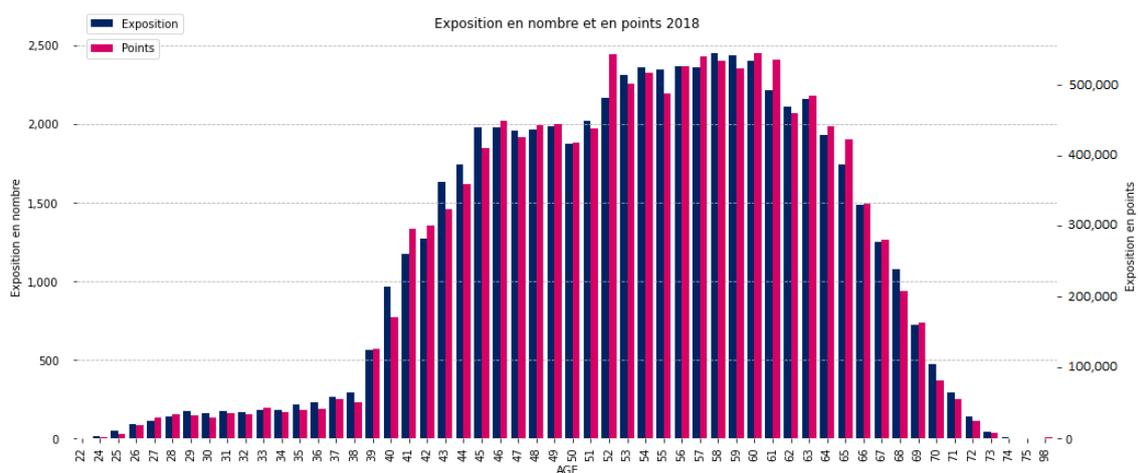


FIGURE 3.3 – Exposition par âge en 2018

A partir de 2018, année suivant la fermeture du régime, le profil de l'effectif par âge et le profil

CHAPITRE 3. PROBLÉMATIQUE ET RÉFLEXIONS SPÉCIFIQUES AU PORTEFEUILLE

des décès par âge change radicalement. En effet, suite aux remboursements et donc à la sortie des assurés les plus jeunes du régime, l'âge moyen devient plus élevé et la population se concentre sur une tranche d'âge plus élevée, à savoir 40 – 70 ans.

Etant donné que le régime ne se renouvelle plus depuis 2017, le crédit accordé aux statistiques générales du portefeuille des années 2018 et 2019 doit être plus élevé pour la prédiction de la mortalité future ainsi que dans la construction des modèles.

Nombre de décès théorique	> 1 (variante)	> 3 (variante)	> 5 (standard)
2010-2020	45-63	53-61	55-60
2010-2019	44-63	53-61	55-60
2011-2019	44-65	53-62	55-61
2012-2019	44-66	53-64	55-62

FIGURE 3.4 – Critère de Cochran

Le critère de Cochran [2] a été testé sur la base de données mais n'aboutit pas à des résultats exploitables. Celui-ci consiste à vérifier que le nombre de décès estimé est supérieur à 5 et que le nombre de décès réel est non nul sur toute la tranche d'âge.

Les variantes du critère de Cochran ont été utilisées en modifiant le seuil initialement et arbitrairement fixé à 5. Ces seuils testés sont plus faibles, c'est-à-dire moins exigeant afin d'aboutir à des résultats plus exploitables. Cependant baisser ce seuil permet d'élargir la plage d'âge à retenir pour les modélisations mais celui-ci fait baisser la significativité des SMR calculés sur la tranche d'âge obtenue.

Ainsi pour l'application des modèles, la tranche d'âge 40 – 70 ans sera retenue.

Chapitre 4

Méthodes de construction et modélisations des tables de mortalités prospectives

Le domaine d'application a été fixé dans le précédent chapitre. Celui-ci sera utilisé dans le présent chapitre qui traite la partie estimation des taux bruts de mortalité, modélisation et ajustement de ces derniers et application.

Pour ce faire, différentes notions sont alors à introduire afin de suivre au mieux le cheminement de ce chapitre. Les notions sont à la fois mathématiques, conceptionnelles et pratiques.

4.1 Estimation des taux bruts de mortalité

4.1.1 Notions mathématiques actuarielles

Définition 1 : *Fonction de répartition de T*

Soit T une variable aléatoire représentant la durée de vie d'un individu.

$$\forall t \in \mathbb{R}_+ : F(t) = \mathbb{P}[T \leq t]$$

Définition 2 : *Fonction de répartition de T_x*

Soit T_x la durée de vie d'un individu d'âge x .

$$\begin{aligned} \forall t \in \mathbb{R}_+ : F_x(t) &= \mathbb{P}[T_x \leq t] \\ &= \mathbb{P}[T \leq x+t \mid T > x] \\ &= \frac{F(x+t) - F(x)}{1 - F(x)} \end{aligned}$$

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Les définitions précédentes sont nécessaires afin d'établir les probabilités qui sont utilisées dans ce mémoire, à savoir les probabilités de décès, de survie, l'espérance de vie, la fonction de survie et la fonction de hasard.

Définition 3 : *Probabilité de décès d'un individu d'âge x au plus tard à l'âge $x + t$*

Soit x fixé dans \mathbb{N} :

$$\forall t \in \mathbb{R}_+ : {}_tq_x = \mathbb{P}[T_x \leq t] = F_x(t)$$

Définition 4 : *Probabilité de survie d'un individu d'âge x à l'âge $x + t$*

Soit x fixé dans \mathbb{N} :

$$\forall t \in \mathbb{R}_+ : {}_tp_x = \mathbb{P}[T_x > t] = 1 - {}_tq_x = 1 - F_x(t)$$

Définition 5 : *Espérance de vie à l'âge x*

Soit x fixé dans \mathbb{N} :

$$\begin{aligned} \ddot{e} = \mathbb{E}[T_x] &= \int_0^{\infty} t dF_x(t) \\ &= \int_0^{\infty} [1 - F_x(t)] dt \\ &= \int_0^{\infty} {}_tp_x dt \end{aligned}$$

Définition 6 : *Fonction de survie*

La fonction de survie est par définition le complément de la fonction de répartition.

$$\forall t \in \mathbb{R}_+ : S(t) = 1 - F(t) = \mathbb{P}(T > t)$$

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Définition 7 : Fonction de survie conditionnelle

Soit x fixé dans \mathbb{N} :

$$\begin{aligned} \forall t \in \mathbb{R}_+ : S_x(t) &= \mathbb{P}[T_x > t] \\ &= \mathbb{P}[T > x+t \mid T > x] \\ &= \frac{\mathbb{P}[T > t+x]}{\mathbb{P}[T > x]} \\ &= \frac{S(x+t)}{S(x)} \end{aligned}$$

Définition 8 : Fonction de hasard

La fonction de hasard se définit par l'expression suivante :

$$\forall t \in \mathbb{R}_+ : h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t))$$

La fonction de survie peut être exprimée en fonction de la fonction de hasard en manipulant l'expression précédente pour obtenir :

$$\forall t \in \mathbb{R}_+ : S(t) = \mathbf{exp} \left(- \int_0^t h(s) ds \right)$$

Notations supplémentaires :

- l_x : Nombre de survivants à l'âge x
- d_x : Nombre de décès à l'âge x
- p_x : Probabilité de survivre à l'âge x en ayant l'âge x
- q_x : Probabilité de décéder à l'âge x en ayant l'âge x
- N_x : Nombre d'individus à l'âge x
- E_x : Exposition à l'âge x

- $E_x(t)$: Exposition à l'âge x l'année calendaire t
- $d_x(t)$: Nombre de décès à l'âge x constaté l'année calendaire t
- $N_x(t)$: Nombre d'individu à l'âge x l'année calendaire t

- $\tilde{q}_x(t)$: Probabilité de décès ajustée à l'âge x l'année calendaire t
- $\hat{q}_x(t)$: Probabilité de décès estimée à l'âge x l'année calendaire t
- $q_x^{ref}(t)$: Probabilité de décès à l'âge x l'année calendaire t issue de la table de référence

4.1.2 Diagramme de Lexis

Le diagramme de Lexis [9] est un graphique où apparaît en abscisse l'axe du temps calendaire et en ordonnée l'axe de l'âge. Chaque segment représente la trajectoire de vie d'un individu. Ce diagramme permet de visualiser schématiquement les événements comme la naissance ou le décès qui surviennent durant la vie d'individu.

Par exemple, le segment compris entre les points de coordonnées $(t_0, 0)$ et $(t_0 + x, x)$ représente la trajectoire de vie des individus nés l'année t_0 et décédés à l'âge x l'année $t_0 + x$. Ces segments sont alors des droites affines de pente unitaire. L'une des extrémités du segment est la naissance de l'individu, l'autre extrémité du segment est le décès de l'individu qui est appelé couramment « point mortuaire ».

En pratique, les actuaires sont amenés à travailler en temps discret et ont deux représentations du diagramme de Lexis qui selon la représentation choisie aboutit à des résultats différents. En effet, le choix de la décomposition par période ou par cohorte (ou générationnelles) mène à des définitions différentes de la probabilité de décès.

Dans cette partie, les notations seront les suivantes :

- t_0 : Date de naissance de l'individu
- X : Durée de vie de l'individu

Définition 9 : Probabilité de décès selon une décomposition par période

Cette décomposition regroupe les individus atteignant l'âge x sur l'année t

$$q_p(x, t) = P(X \leq x + 1, X + t_0 \leq t + 1 | X > x, X + t_0 > t)$$

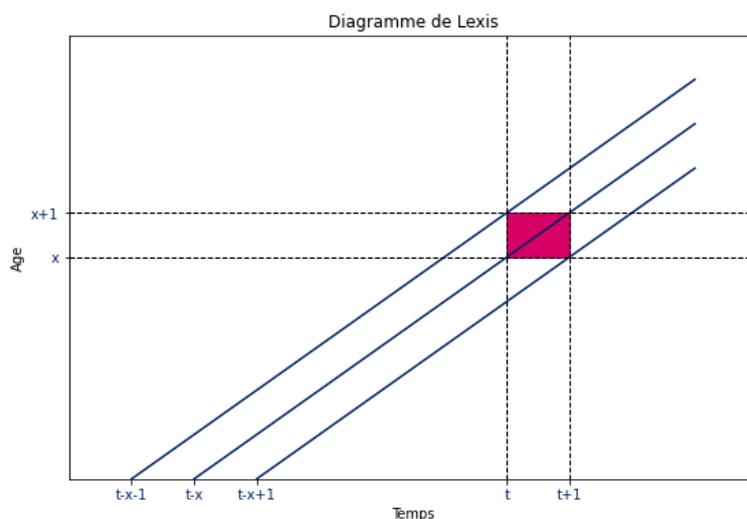


FIGURE 4.1 – Diagramme de Lexis par période

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Dans la décomposition par période, l'intérêt porte uniquement sur le carré rose du diagramme qui concerne tous les individus qui décèdent durant l'année t à l'âge x . Deux générations sont alors concernées, ce sont les individus qui sont nés les années $t - x - 1$ et $t - x$.

Cette décomposition ne permet pas de cibler les individus d'une même génération, c'est pour cela qu'il existe une seconde décomposition qui est la décomposition par cohorte.

Définition 10 : Probabilité de décès selon une décomposition par cohorte (ou générationnelle)

Cette décomposition regroupe les individus né l'année $t - x$.

$$q_c(x, t) = P(X \leq x + 1 | X > x, t_0 \in]t - x, t - x + 1])$$

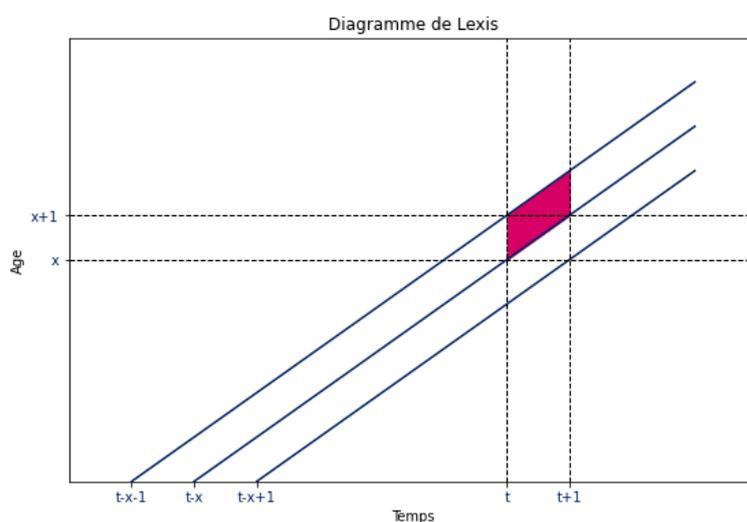


FIGURE 4.2 – Diagramme de Lexis par cohorte

Le diagramme de Lexis est le même que celui présenté précédemment, la différence va être dans son utilisation. La décomposition par cohorte comptabilise les décès durant l'année t parmi la génération $t - x - 1$, représenté non plus par un carré mais par un parallélogramme rose dans le diagramme de Lexis.

Une seule génération est concernée, en revanche les individus concernés peuvent à la fois avoir l'âge x ou l'âge $x + 1$, ce qui n'était pas le cas dans la décomposition par période.

De là en fonction de la décomposition choisie, par période ou par cohorte, l'exposition utilisée et la manière de compter les décès ne sont pas les mêmes, ce qui impacte donc les taux de mortalités calculés.

4.1.3 Troncature gauche et censure aléatoire droite

Les notions de troncature et de censure [3] sont indispensables dans le cadre de la création d'une table de mortalité car les observations doivent être considérées avec précaution et selon une certaine logique.

Définissons les termes de censure et de troncature afin de comprendre leur intérêt dans l'étude.

Censure

La donnée d'intérêt est notée X , C la censure et T l'observation. Dans le cas d'une censure à droite par exemple, l'observation T vaut :

$$T = \begin{cases} X & \text{si } X < C \\ C & \text{si } X > C \end{cases}$$

Dans le cas d'une censure à gauche, il s'agit de l'inverse :

$$T = \begin{cases} X & \text{si } X > C \\ C & \text{si } X < C \end{cases}$$

La censure C peut être fixe, dans ce cas c'est une censure de type I. La censure C peut aussi être aléatoire, dans ce cas c'est une censure de type III

Troncature

De même la donnée d'intérêt est noté x , Z est le seuil de troncature et T l'observation. Dans le cas d'une troncature à gauche par exemple, l'observation du phénomène est défini par :

$$T = \begin{cases} X & \text{si } X > Z \\ 0 & \text{si } X < Z \end{cases}$$

Dans le cas d'une troncature à droite, on a :

$$T = \begin{cases} 0 & \text{si } X < Z \\ X & \text{si } X > Z \end{cases}$$

La variable représentant la troncature peut être fixe ou aléatoire comme la censure.

4.1.4 Estimateur de Hoem

Principe de la méthode

L'estimateur de Hoem [9] est l'estimateur qui sera retenu pour le calcul de l'exposition. C'est un estimateur paramétrique, il servira à contourner la contrainte de censure à droite.

En effet, l'estimateur de Hoem permet de ne considérer les individus à risque que sur la durée de leur présence dans le portefeuille, restreinte à la période d'observation.

Dans le cadre des modélisations, l'hypothèse faite sur les décès, noté $(D_{x,i})_i$ est que les décès sont modélisés par une loi de Bernoulli toutes indépendantes et de paramètre ${}_tq_x$ sur une plage d'âge $[x, x + 1[$ et où t_i est la durée d'exposition de l'individu i .

$$D_{x,i} \sim \mathcal{B}(t_i, q_x)$$

Une hypothèse très couramment utilisée est que la force de mortalité est constante sur un âge donné, ce qui conduit à dire que la probabilité de décès est constante sur le segment $[x, x + 1[$. Le premier résultat est obtenu par un développement limité (DL) au premier ordre appliqué à la fonction exponentielle.

$$\forall t \in [0, 1[: \mu_t^x \approx q_x \Rightarrow {}_tq_x = t_i \times q_x$$

De là le nombre de décès moyen est calculé via l'espérance :

$$\mathbb{E}\left(\sum_i D_{x,i}\right) = \sum_i \mathbb{E}(D_{x,i}) = q_x \sum_i t_i$$

Ainsi la variable d'intérêt est à isoler, à savoir la probabilité de décès :

$$q_x = \frac{\mathbb{E}(\sum_i D_{x,i})}{\sum_i t_i}$$

Pour obtenir un estimateur de q_x on remplace l'espérance des décès par son estimateur, ce qui conduit à l'expression finale suivante :

$$\hat{q}_x = \frac{\sum_i d_{x,i}}{\sum_i t_i}$$

Finalement l'estimateur des probabilités de décès a une expression intuitive, celui-ci est construit en comptabilisant le nombre de décès observé des assurés du portefeuille divisé par l'exposition de ces derniers.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Application de la méthode

L'exposition par âge est calculée dans un premier temps, par année et par sexe appliquée au portefeuille. Les graphiques suivants sont obtenus :

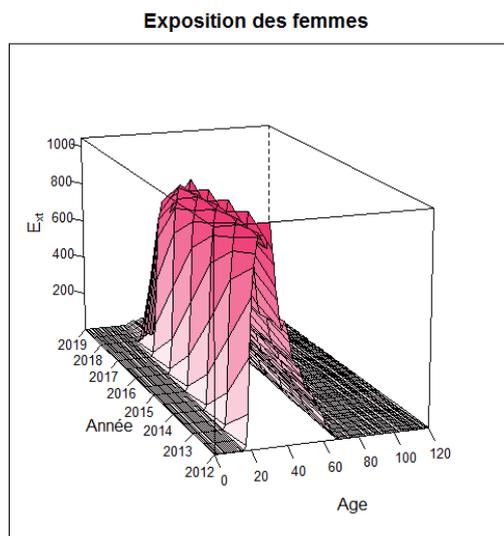


FIGURE 4.4 – Exposition en nombre par âge et par année - Femme

Ce graphique permet de constater grossièrement qu'il y a une concentration des femmes entre 20 et 60 ans de 2012 à 2017. Il faut garder en tête que le régime ferme en 2016, c'est pour cela qu'une déformation du graphique est observée pour les années 2018 et 2019. De plus l'exposition des femmes par âge et par année ne dépasse pas 1000 en effectif, ce qui illustre bien le faible volume de donnée sur la gente féminine du portefeuille.

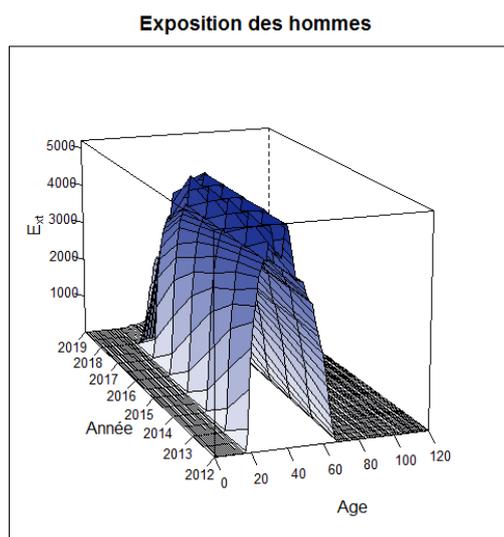


FIGURE 4.5 – Exposition en nombre par âge et par année - Homme

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Le profil d'exposition des hommes est similaire à celui des femmes. Cependant, l'échelle en ordonnée n'est pas la même. L'exposition par âge et par année atteint les 4500 en effectif, ce qui est largement supérieur à l'effectif des femmes. Par ailleurs, une déformation est constatée à partir des années 2018 dû à la fermeture du régime, ce qui est en adéquation avec l'analyse des données faite dans le chapitre 2 (figure 2.18).

Dans un second temps, le nombre de décès est comptabilisé par âge et par année par sexe, ceci est illustré dans les graphiques suivant :

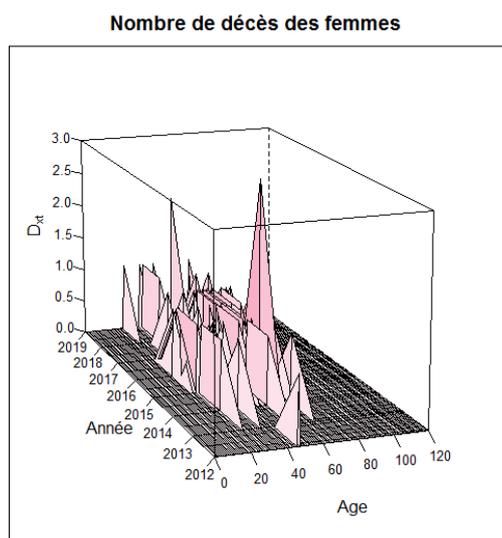


FIGURE 4.6 – Décès en nombre par âge et par année - Femme

Le graphique illustrant le nombre de décès par âge et par année pour les femmes présente de fortes irrégularités. Pas plus de 3 décès n'est comptabilisé par âge et par année, ce qui est extrêmement faible voir clairement insuffisant pour modéliser une table de mortalité. Par ailleurs le nombre de décès est nul pour beaucoup d'âge, ce qui sujet à des approximations très grossières dans les modèles. Encore une fois, le volume de donnée est insuffisant pour créer une table de mortalité pour les femmes.

Chez les hommes le nombre de décès est plus important que les femmes, ce qui est plus rassurant quant à l'ajustement des modèles. Le nombre de décès est croissant par âge pour presque toutes les années. De plus une tendance à la hausse du nombre de décès est observée par année pour tous les âges et particulièrement pour les âges élevés du portefeuille, c'est-à-dire autour de 60 – 70 ans.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

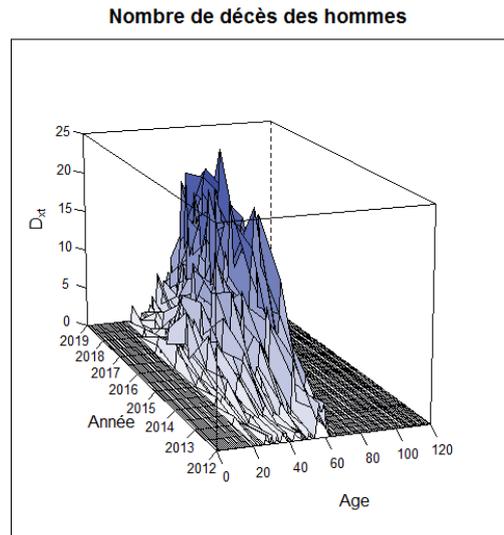


FIGURE 4.7 – Décès en nombre par âge et par année - Homme

A partir de l'exposition et du nombre de décès, les probabilités de mortalité brutes peuvent être calculées.

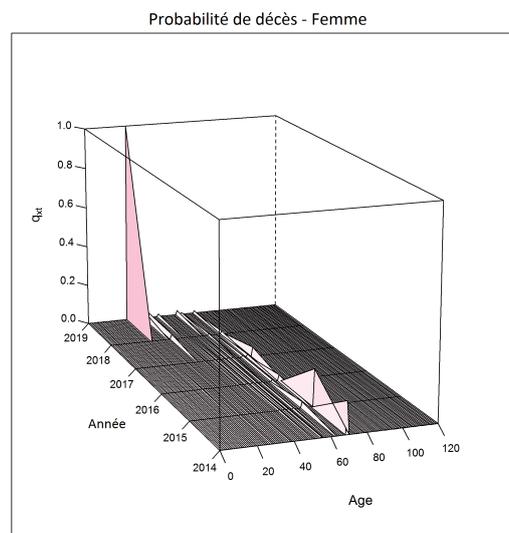


FIGURE 4.8 – Probabilité de décès par âge et par année - Femme

Sans surprise, le graphique des probabilités de décès manque de cohérence et présente de grossières erreurs. En effet, une probabilité de décès égale à 1 est constatée pour un âge autour de 20 ans en 2019, ce qui est absurde. Les données pour les femmes sont donc clairement inexploitable.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

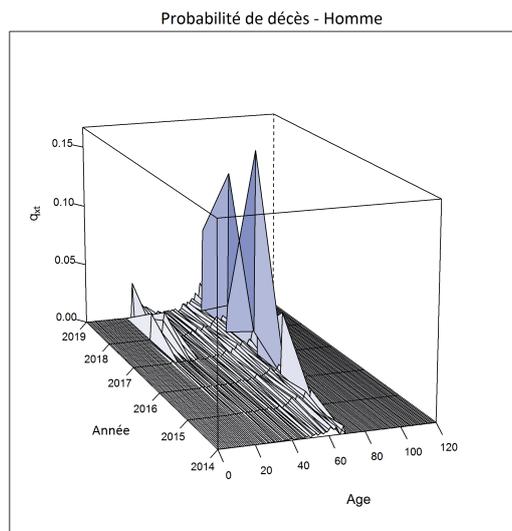


FIGURE 4.9 – Probabilité de décès par âge et par année - Homme

En revanche pour les hommes, à première vue les probabilités de décès sont plus cohérentes et moins grossières. En effet, pas de valeur de probabilité de décès absurde sur la plage d'observation.

De plus, les probabilités de décès ont l'air croissantes en fonction de l'âge, ce qui est rassurant pour la suite de l'étude. Néanmoins on remarque que les probabilités de décès sont croissantes par année pour un même âge donné, ceci est contraire à l'hypothèse d'amélioration de l'espérance de vie en fonction du temps. Peut-être qu'il s'agit d'une particularité du portefeuille ou peut-être que c'est une observation non significative et qui n'est pas vérifiée analytiquement.

4.2 Ajustement des taux bruts de mortalité

4.2.1 SMR (Standardized Mortality Ratio)

Principe de la méthode

La première méthode exogène qui est testée est celle du SMR [6]. Il s'agit de la méthode la plus simple à appliquer puisqu'elle consiste à calculer un coefficient, le SMR qui n'est rien d'autre que le rapport entre le nombre de décès observé et le nombre de décès théorique.

Ce coefficient peut être calculé soit par âge, soit par année. Il permet de constater le niveau de sur-mortalité ou de sous-mortalité du portefeuille par rapport à la mortalité dite de référence provenant d'une table réglementaire à condition d'avoir un volume de donnée suffisant pour que le SMR soit significatif.

$$SMR = \frac{\sum_{(x^*, t^*)} D_{x^*}(t^*)}{\sum_{(x^*, t^*)} E_{x^*}(t^*) \times q_{x^*}^{ref}(t^*)}$$

- Si $SMR < 1$: La table de référence surestime la mortalité du portefeuille
- Si $SMR > 1$: La table de référence sous-estime la mortalité du portefeuille

Afin de contourner la contrainte de suffisance de données, le calcul du SMR moyen sur plusieurs années peut être réalisé pour augmenter le volume de donnée. Le SMR est donc calculé sur la période d'observation envisagée.

Une fois que le SMR est obtenu pour les hommes et les femmes, celui-ci est appliqué de manière multiplicative aux probabilités de décès de la table de mortalité de référence choisie, d'où le terme de positionnement. Dans cette étude, la table de référence qui est utilisée est la TGH-05 et la TGF-05.

$$\tilde{q}_x(t) = q_x^{ref}(t) \times SMR$$

Application de la méthode

SMR par année :

La méthode du SMR est appliquée par année pour les valeurs de x et de t suivantes (tranche d'âge et période d'observation retenue pour l'étude) :

$$x^* \in [40, 70] \text{ et } t^* \in [2012, 2019]$$

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

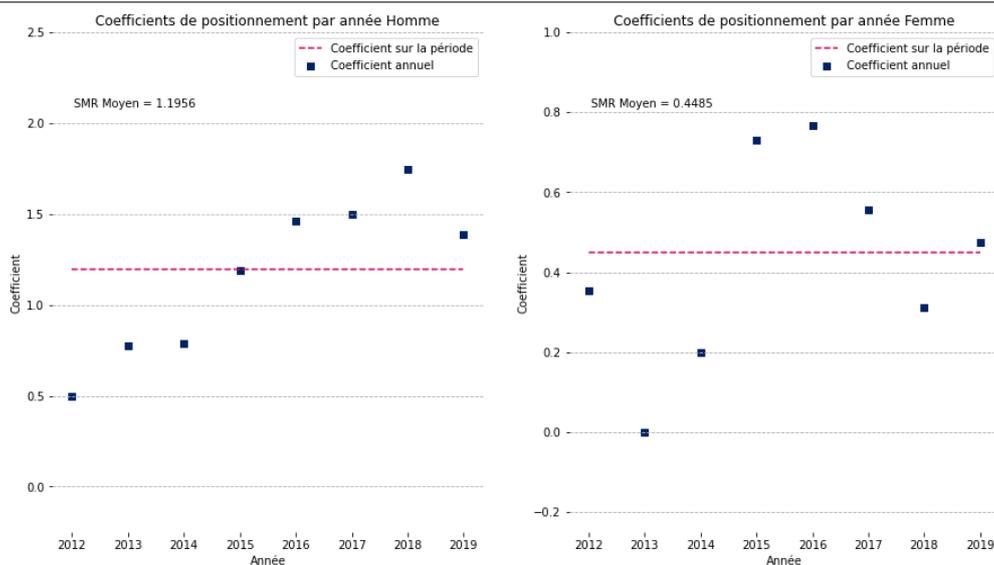


FIGURE 4.10 – SMR par année de 2012 à 2019

Ainsi les paramètres suivants sont obtenus :

- $SMR_{homme} = 1,196$
- $SMR_{femme} = 0,449$

Le SMR_{femme} n'est pas significatif car le volume de donnée est très faible et donc clairement insuffisant.

Le SMR_{homme} moyen est supérieur à 1. Ce qui signifie que la table de mortalité de référence ayant servi à calculer le SMR, à savoir le TGH-05 et le TGF-05 sous-estime la mortalité du portefeuille.

En revanche, le SMR_{homme} est en dessous de 1 de 2012 à 2014 inclus, ce n'est qu'à partir de 2015 que le SMR_{homme} passe au-dessus de 1. La surmortalité moyenne observée s'explique donc par les années d'exercice les plus récentes, information qu'il va falloir prendre en compte pour la projection de la mortalité des tables.

Par ailleurs une tendance à la hausse du SMR_{homme} est globalement observable de 2012 à 2018.

SMR par âge :

La surmortalité du portefeuille est observée, pour avoir une observation plus fine sur la tranche d'âge 40 – 70 ans, le graphique suivant est affiché. Celui-ci donne le SMR moyen par âge pour le calibrage suivant :

$$x^* \in [40, 70] \text{ et } t^* \in [2012, 2019]$$

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

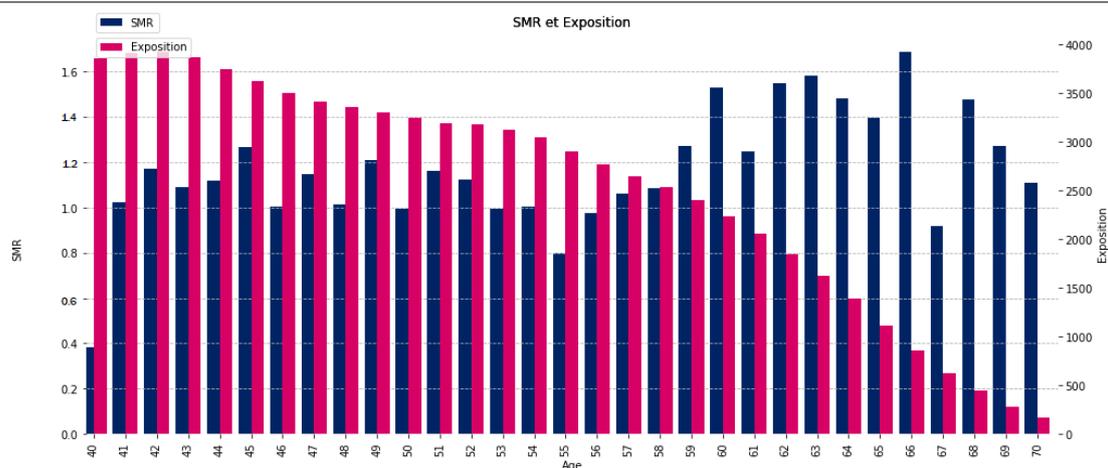


FIGURE 4.11 – SMR par âge de 2012 à 2019

Dans un premier temps, l'exposition décroissante en fonction de l'âge est avérée, ceci s'explique par la jeunesse du portefeuille. Ce qui n'est pas représentatif de l'état actuel du portefeuille depuis sa fermeture en 2016.

Par ailleurs, la surmortalité à presque tous les âges est constatée avec une accentuation plus élevée pour les âges au-dessus de l'âge de départ à la retraite (pour rappel : 55 ans pour ce régime).

A titre d'observation, le SMR par âge est calculé pour les années 2018 et 2019.

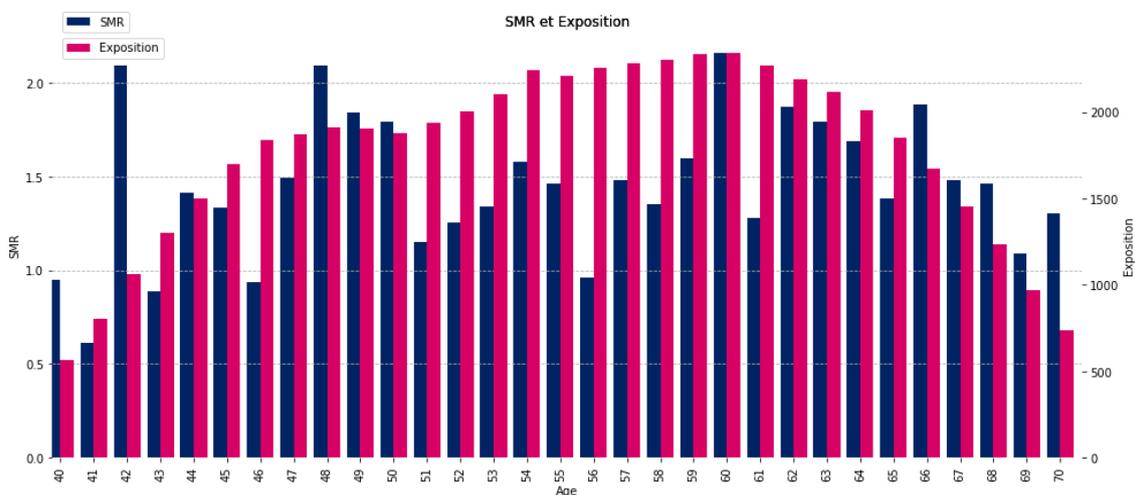


FIGURE 4.12 – SMR par âge de 2018 à 2019

L'exposition n'est pas répartie de la même manière que pour la période d'observation globale et c'est plutôt vers cette répartition de l'exposition que le portefeuille va évoluer. De plus, la surmortalité est bien présente sur la tranche d'âge 40 – 70 ans et il n'y a que très peu d'âge pour lesquels la sous-mortalité est vérifiée.

4.2.2 Régression logistique - variante du modèle de BRASS

Principe de la méthode

La deuxième méthode exogène est la régression logistique, il s'agit d'une variante semi-paramétrique du modèle de Brass (1971) [6]. Celle-ci requiert l'utilisation de la fonction « logit » qui présente des caractéristiques bien spécifiques, présentées ci-contre justifiant l'intérêt de son utilisation.

$$\text{logit}(q_x(t)) = \ln \left(\frac{q_x(t)}{1 - q_x(t)} \right)$$

avec $q_x(t) \in [0, 1 [$ et donc $\text{logit}(q_x(t)) \in] -\infty; +\infty [$

L'intérêt de l'utilisation de la fonction logit est multiple :

- D'une part, la fonction logit permet de réaliser une transformation non contrainte pour ainsi déformer les $q_x(t)$ plus aisément
- D'autre part, la fonction logit varie à peu près linéairement avec x , ce qui permet d'établir une relation linéaire entre les logit des $q_x^{ref}(t)$ et les logit des $\tilde{q}_x(t)$.

Le principe de la méthode est de positionner les logit des taux de mortalité du portefeuille par rapport aux logit des taux de mortalité d'une table de référence externe selon le modèle suivant :

$$\text{logit}(\tilde{q}_x(t)) = \alpha + \beta \times \text{logit}(q_x^{ref}(t)) + \varepsilon_{x,t}$$

- $\varepsilon_{x,t}$: est un terme représentant une erreur gaussienne *iid*
- $q_x^{ref}(t)$: est la probabilité de décès d'une table de référence externe
- α : est un indicateur de mortalité propre au portefeuille entier indépendant de l'âge
- β : est un paramètre qui mesure la contribution du vieillissement à la mortalité

Le modèle présenté est un modèle à deux paramètres par opposition au modèle de positionnement du SMR qui ne nécessite qu'un seul paramètre. Il permet ainsi d'avoir un ajustement plus précis en fonction de l'âge et de la dimension temporelle.

L'estimation des paramètres α et β se fait par la méthode des moindres carrés qui consiste à minimiser la distance pondérée par l'exposition entre les probabilités de décès observées et estimées. Il s'agit de minimiser la quantité suivante :

$$\tilde{q}_x(t) = \frac{\exp \left(\tilde{\alpha} + \tilde{\beta} \times \text{logit} \left(q_x^{ref}(t) \right) \right)}{1 + \exp \left(\tilde{\alpha} + \tilde{\beta} \times \text{logit} \left(q_x^{ref}(t) \right) \right)}$$

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Il faut noter que la fonction logit est concave sur l'intervalle $\left]0, \frac{1}{2}\right[$ et convexe sur l'intervalle $\left]\frac{1}{2}, 1\right[$.

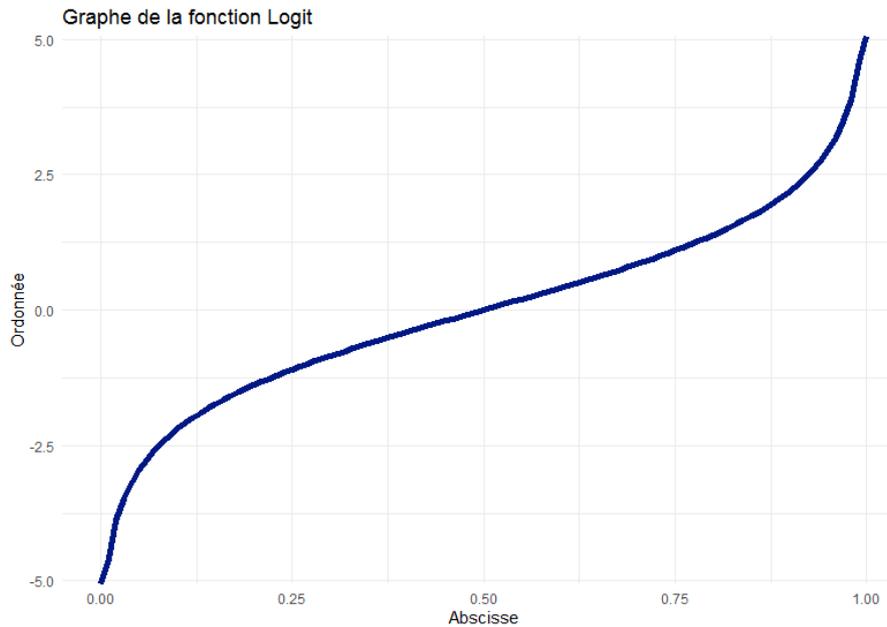


FIGURE 4.13 – Courbe de la fonction logit sur $[0, 1[$

D'après l'inégalité de Jensen, si une fonction f est convexe alors :

$$\mathbb{E}(f(x)) \geq f(\mathbb{E}(x))$$

La fonction *logit* est concave sur $\left]0, \frac{1}{2}\right[$, donc $-logit$ est convexe sur $\left]0, \frac{1}{2}\right[$, on en déduit que sur $\left]0, \frac{1}{2}\right[$:

$$\mathbb{E}(\logit(\tilde{q}_x(t))) \geq \logit(q_x(t))$$

Ainsi, d'après cette inégalité les probabilités de décès théoriques sont sous-estimées pour $q_x(t) \in \left]0, \frac{1}{2}\right[$. En revanche, les probabilités de décès théoriques sont surestimées sur l'intervalle $\left]0, \frac{1}{2}\right[$.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Application de la méthode

L'utilisation de cette méthode a été réalisée sur le logiciel R à l'aide du package *ELT*. La table de référence choisie est la TGH-05 et la TGF-05. Le modèle de régression logistique a été appliqué sur la tranche d'âge retenue et la période d'observation retenue, c'est-à-dire :

$$x^* \in [40, 70] \text{ et } t^* \in [2012, 2019]$$

Les paramètres obtenus sont les suivants :

Paramètres	Hommes	Femmes
$\tilde{\alpha}$	0.8324	-2.7716
$\tilde{\beta}$	1.084	3.3997

Le package *ELT* [5] permet de visualiser plusieurs métriques concernant le modèle choisi, il n'est pas pertinent de toutes les présenter dans le corps du mémoire. Les plus pertinentes seront présentées dans cette partie, certaines autres métriques apportant un intérêt à la compréhension de l'étude seront mises en annexe.

Les données sur les femmes ne seront pas présentées ici car les résultats ne sont pas exploitables dû à la faible volumétrie des données. Toutefois, les graphiques pour les femmes seront présentés en annexe (annexe 5.2.3, annexe 5.2.3).

Les probabilités de décès estimées et observées sont comparées, pour le moment seuls les graphiques des années 2012 et 2014 sont présentés.

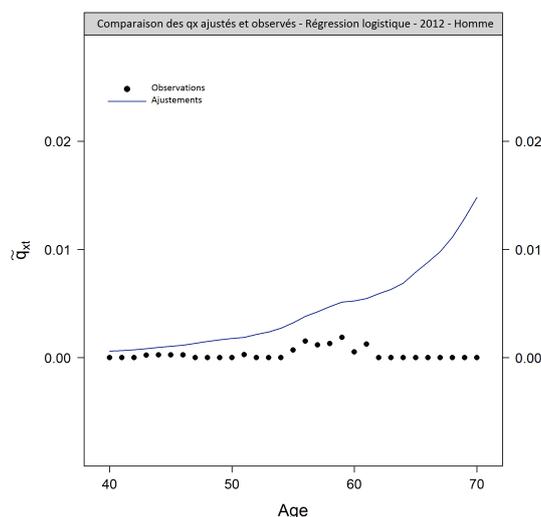


FIGURE 4.14 – Comparaison des q_x estimés et réels - 2012 - Homme

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Le graphique présente les probabilités de décès estimées par le modèle en bleu et les points en noirs sont les probabilités de décès observées compte tenu des décès réels du portefeuille en 2012.

Le modèle surestime très largement les probabilités de décès réelles à tous les âges, ceci s'explique par le nombre de décès faible observé.

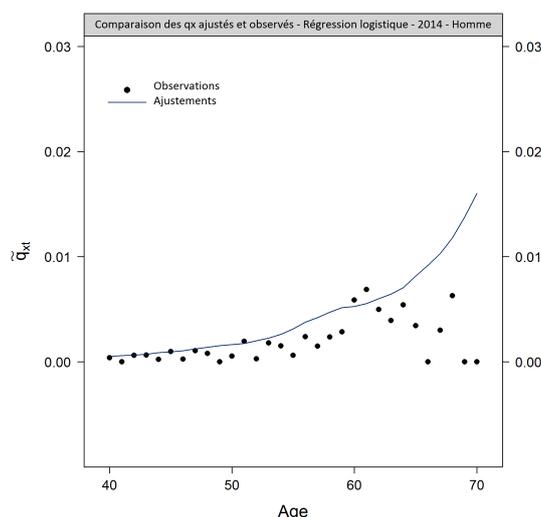


FIGURE 4.15 – Comparaison des q_x estimés et réels - 2014 - Homme

Ce graphique illustre la même variable mais pour l'année 2014. Le modèle coïncide davantage avec les observations mais reste toutefois au-dessus des observations réelles.

Pour les âges allant de 40 ans à 60 ans, l'erreur observable est faible puisque les points sont très proches de la courbe. En revanche au-delà de 60 ans, le modèle estime des probabilités de décès de plus en plus croissantes (allure convexe) mais les probabilités de décès réelles n'augmentent pas et sont de plus en plus volatiles.

L'estimation du modèle n'est pas insatisfaisante de 40 à 60 ans mais le devient au-delà de 60 ans.

Compte tenu de l'ajustement du modèle avec les observations réelles pour l'année 2012, la question se pose de savoir s'il ne faudrait pas encore réduire la période d'observation et supprimer les années 2012 et 2013, 2013 présentant des caractéristiques menant aux mêmes conclusions que 2012 (voir graphique en annexe 5.2.3, annexe 5.2.3).

Les années 2010 et 2011 ont déjà été supprimées à cause du nombre de décès insuffisant, les raisons et justifications sont données dans la section 3.2. Les années 2012 et 2013 présentent les mêmes faiblesses, elles ne sont pas représentatives des projections futures puisque le portefeuille est vieillissant et enfin elles sont trop éloignées des caractéristiques générales du portefeuille sur l'ensemble des années d'observation.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Pour la suite de l'étude, la tranche d'âge retenue et la période d'observation retenue sont les suivantes :

$$x^* \in [40, 70] \text{ et } t^* \in [2014, 2019]$$

Les paramètres obtenus sont les suivants :

Paramètres	Hommes	Femmes
$\tilde{\alpha}$	1.3154	-20.9756
$\tilde{\beta}$	1.1597	0.9551

Le logarithme des probabilités de décès estimées et observées est présenté afin de juger du bon ajustement du modèle avec les données du portefeuille.

Le package *ELT* permet d'obtenir graphiquement la comparaison des $\log(\tilde{q}_x(t))$ modélisés et des $\log(q_x(t))$ observés en fonction de l'âge pour chaque année d'observation.

Les graphiques présentés ci-dessous sont ceux de la première et la dernière année d'observation à savoir 2014 et 2019 afin de constater l'évolution de l'ajustement des $\log(\tilde{q}_x(t))$ modélisés avec les $\log(q_x(t))$ observés entre le début et la fin de la période d'observation.

Il sera également pertinent de juger du bon ajustement du modèle avec les données observées.

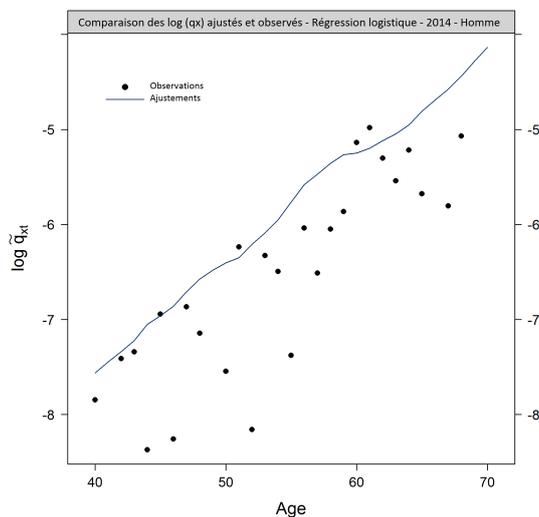


FIGURE 4.16 – Comparaison des $\log(q_x)$ estimés et réels - 2014 - Homme

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

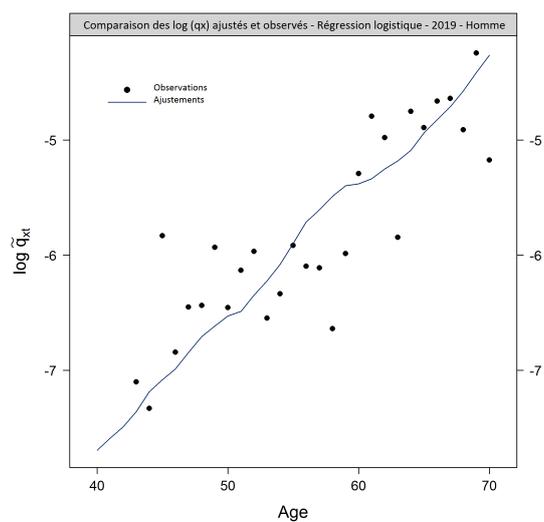


FIGURE 4.17 – Comparaison des $\log(q_x)$ estimés et réels - 2019 - Homme

La validation du modèle se fera dans un chapitre ultérieur de manière plus précise et plus poussée.

4.2.3 Modèle linéaire généralisé de Poisson

Principe de la méthode

La troisième méthode est le modèle linéaire généralisé de Poisson. Ce modèle intègre à la fois la mortalité de référence, l'âge et l'année calendaire. Cette fois-ci ce n'est plus la probabilité de décès qui est paramétrée mais directement le taux instantané de mortalité à l'âge x pour l'année t . Ainsi, ce modèle s'écrit de la façon suivante :

$$D_{x^*}(t^*) \sim \mathcal{P}(E_{x^*}(t^*)\mu_{x^*}(t^*))$$

où l'expression du taux instantané de décès est modélisée par :

$$\mu_{x^*}(t^*) = \beta_0 + \beta_1 \log(q_{x^*}^{ref}(t^*)) + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^*$$

La variable $\mu_{x^*}(t^*)$ correspond au taux instantané de mortalité qui dépend de l'âge x , de l'année calendaire t et de $\beta_0, \beta_1, \beta_2, \beta_3$ et β_4 qui sont des scalaires permettant la paramétrisation du modèle. Par ailleurs ce modèle inclut les probabilités de mortalité d'une table de référence sur laquelle les ajustements seront positionnés.

Dans l'expression du taux instantané de mortalité, le paramètre β_4 est facteur de $x^* t^*$. Ce produit permet d'inclure une interaction entre l'âge et l'année calendaire dans le modèle.

Toutefois, si le nombre d'année d'observation n'est pas suffisant, c'est-à-dire inférieur à 10 années alors la variable temporelle ne permet pas d'expliquer la variable d'intérêt, celle-ci doit être exclue du modèle. Seuls les paramètres β_0, β_1 et β_2 sont estimés et les paramètres β_3 et β_4 sont considérés comme nuls.

Une fois que les paramètres sont estimés par le modèle, le passage de probabilité instantané de décès à probabilité de décès se fait en passant à l'exponentiel, ce qui donne l'expression suivante :

$$\tilde{q}_x(t) = \exp\left(\tilde{\beta}_0 + \tilde{\beta}_1 \log(q_x^{ref}(t)) + \tilde{\beta}_2 x + \tilde{\beta}_3 t + \tilde{\beta}_4 x t\right)$$

Les probabilités de décès ajustés sont ainsi obtenues.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Application de la méthode

Le calibrage de la tranche d'âge et de la période d'observation est le même que dans les modèles précédents. L'application de la méthode se fait à l'aide du package **ELT**, il s'agit de la méthode 3 du package.

$$x^* \in [40, 70] \text{ et } t^* \in [2014, 2019]$$

Compte tenu du nombre d'année d'observation réduit (inférieur à 10), le choix est fait d'exclure la variable temporelle dans l'expression du taux instantané de mortalité. Ainsi le modèle estimé est le suivant :

$$\mu_x(t) = \beta_0 + \beta_1 \log(q_x^{ref}(t)) + \beta_2 x$$

Les paramètres sont estimés par un GLM sur R. Leurs estimations sont données par les tableaux suivants :

	Hommes			
Paramètres	Estimation	Erreur standard	z valeur	p valeur
$\tilde{\beta}_0$	-23.047	4.076	-5.655	1.56E-08
$\tilde{\beta}_1$	-0.989	0.360	-2.748	0.006
$\tilde{\beta}_2$	0.200	0.033	5.975	2.3E-9

	Femmes			
Paramètres	Estimation	Erreur standard	z valeur	p valeur
$\tilde{\beta}_0$	-6.150	17.464	-0.352	0.725
$\tilde{\beta}_1$	0.585	1.692	0.346	0.730
$\tilde{\beta}_2$	0.054	0.124	0.435	0.664

Les coefficients du tableau pour les femmes ne seront pas retenus car leurs p valeur sont largement supérieurs au seuil d'acceptation de 0.05.

En revanche les coefficients déterminés pour les hommes sont tous acceptés à en juger par la p valeur qui est pour les trois paramètres en dessous de 5%. Donc l'application portera sur uniquement sur les hommes.

Ici, seuls les graphes du log des probabilités de décès sont exposés afin de tenir une démonstration concise allant à l'essentiel.

Comme précédemment, les résultats pour deux années extrêmes sont exposés à savoir l'année 2014 et l'année 2019.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

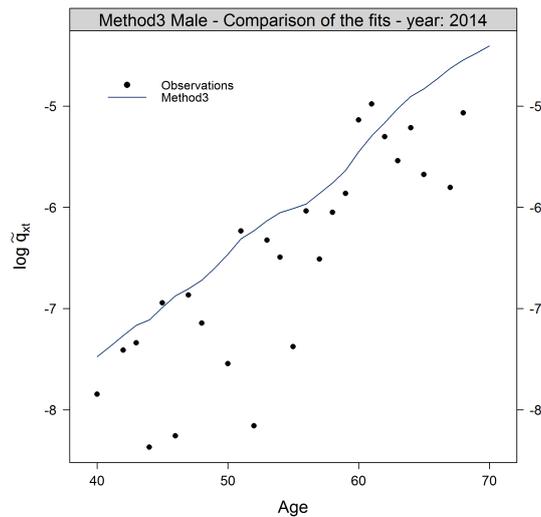


FIGURE 4.18 – Comparaison des $\log(q_x)$ estimés et réels - 2014 - Homme

Comme pour les modèles précédents, le modèle surestime les probabilités de décès observées pour l'année 2014. L'écart est d'autant plus important que l'âge est proche de 40 ans, en revanche pour les âges proches de 70 ans l'estimation de la mortalité coïncide davantage avec la mortalité observée. L'écart se lit en mesurant la distance entre une observation donnée et la courbe du modèle de façon verticale.

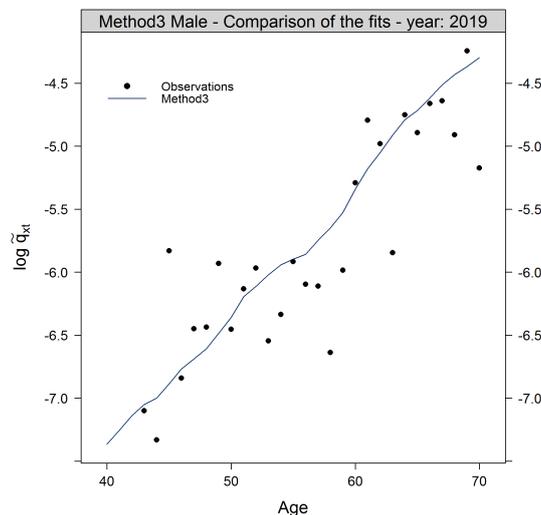


FIGURE 4.19 – Comparaison des $\log(q_x)$ estimés et réels - 2019 - Homme

Sur la dernière année d'observation, le modèle estime la mortalité réelle bien mieux que pour l'année 2014 avec une proportion de point en dessous et au-dessus de la courbe plus équilibrée.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Cette méthode de positionnement offre une modélisation basée sur un concept différent des deux précédents modèles. En effet, le paramétrage porte non pas sur la probabilité de décès mais sur le taux instantané de décès et inclut un paramètre supplémentaire par rapport au modèle semi-paramétrique de Brass. Bien que l'origine de la modélisation soit différente, l'ajustement des taux de décès réel est satisfaisant d'un point de vue graphique.

Trois modèles ont été présentés et testés sur le portefeuille à l'étude. Cependant la tranche d'âge retenue est 40 - 70 ans et donc au-delà de 70 ans les modélisations ne s'appliquent pas étant donné que les données sont absentes pour cette tranche d'âge.

Appliquer les modélisations sur la tranche d'âge au-delà de la tranche d'âge d'observation est risquée d'autant plus que sur le portefeuille à l'étude, une surmortalité est observée. Dans le contexte d'un portefeuille de rentes viagères, projeter une telle tendance en l'absence d'observation est peu prudent donc cette méthode n'est pas à réaliser.

C'est pourquoi il existe des méthodes de fermeture de table pour extrapoler la mortalité aux grands âges. L'inconvénient est que les tendances de mortalité sont également projetées. C'est donc un inconvénient qu'il faudra tenir compte lors de l'étape de validation des modèles.

4.3 Fermeture de la table

4.3.1 Méthode Denuit & Goderniaux

Principe de la méthode

Le principe de la méthode de Denuit & Goderniaux [1] est de modéliser le logarithme des probabilités de décès par une fonction quadratique. Le modèle est appliqué sur une tranche d'âge élevée car il s'agit d'une extrapolation dont la volatilité du modèle est fortement dépendante des données prises en entrée. Le modèle est le suivant :

$$\ln(\tilde{q}_x) = a + bx + cx^2 + \varepsilon_x$$

Avec a, b et c des réels et ε_x *iid* suivant une loi normale centrée de variance μ^2 .

Des contraintes de bon sens sont imposées par ce modèle, à savoir que la probabilité de décès à 130 ans est un évènement certain et que la dérivée de cette probabilité de décès à 130 ans est nulle. La tangente à la courbe au point d'abscisse 130 est donc plate, il s'agit du sommet de la parabole. Ces contraintes impliquent :

$$\begin{aligned} \ln(\tilde{q}_{130}) &= 0 \\ \Rightarrow a + 130b + 130^2c &= 0 \end{aligned}$$

Puis dans un second temps :

$$\begin{aligned} \frac{\tilde{q}'_{130}}{\tilde{q}_{130}} &= 0 \\ \Rightarrow b + 260c &= 0 \end{aligned}$$

Les contraintes mènent au système suivant :

$$\begin{cases} a + 130b + 130^2c = 0 \\ b + 260c = 0 \end{cases}$$

$$\Rightarrow \begin{cases} a = 130^2c \\ b = -260c \end{cases}$$

Le système fournit une infinité de solution car a et b dépendent de la valeur de c . Ainsi le modèle devient :

$$\ln(\tilde{q}_x) = c(130 - x)^2 + \varepsilon_x$$

Enfin, la valeur de c sera déterminée par la méthode des moindres carrés.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Application de la méthode

La méthode de Denuit & Goderniaux préconise de prendre comme dernier âge de décès 130 ans. Cet âge ultime sera conservé dans l'application du modèle, la table ainsi construite fournira des probabilités de décès jusqu'à l'âge de 130 ans.

L'application de la méthode est réalisée via le package *ELT* par la fonction *CompletionA* et *CompletionB* qui permettent d'appliquer la fermeture de table par la méthode de Denuit & Goderniaux. Un paramétrage est nécessaire et est déterminant quant au bon ajustement du modèle.

Tout d'abord, l'ajustement du modèle se fait sur une plage d'âge d'observation qui est au choix de l'utilisateur. Les paramètres *AgeRangeOptiMale* ($\hat{=} \alpha_{age}$) et *BegAgeCompMale* ($\hat{=} \beta_{age}$) sont à entrer dans la fonction.

Les paramètres optimaux α_{age} et β_{age} maximisant le coefficient de détermination R^2 sont retenus. De là les valeurs de (c_t) en fonction du temps sont également déterminées. Pour simplifier, la relation suivante est considérée :

$$(R^2)_t = f(\alpha_{age}, \beta_{age})$$

Dans un premier temps le max et le min de la série $(R^2)_t$ sont retenus pour chaque couple $(\alpha_{age}, \beta_{age})$. Puis on retient le couple pour lequel le *max* de $(R^2)_t$ est le maximum de tous les *max* de $(R^2)_t$, et pour lequel le *min* de $(R^2)_t$ est le maximum de tous les *min* de $(R^2)_t$. Ceci se traduit par les conditions suivantes :

$$(\alpha_{age}, \beta_{age}) = \begin{cases} \operatorname{argmax}_i [\max_t (R^2)_t]_i \\ \operatorname{argmax}_i [\min_t (R^2)_t]_i \end{cases}$$

Les paramètres optimaux sont donnés dans le tableau suivant :

Variables	α_{age}	β_{age}	R^2_{min}	R^2_{max}
Paramètres	67 ans	70 ans	0.9975	0.9999

Le coefficient c est donné par le graphique suivant en fonction du temps pour les hommes et pour les femmes. Cependant les résultats sur la population féminine ne seront pas retenus car elles ne sont pas significatives en raison du manque de donnée.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

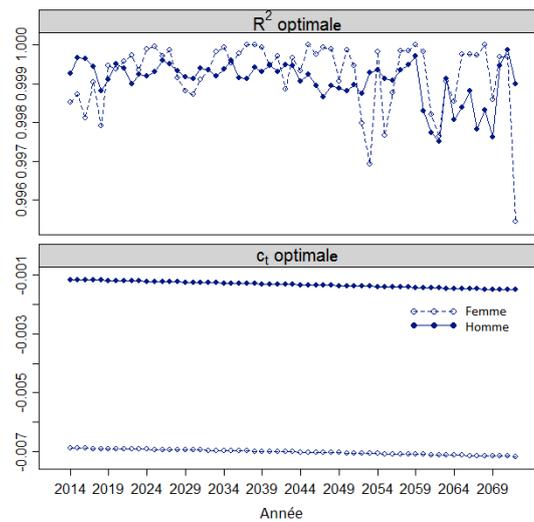


FIGURE 4.20 – Valeur de c en fonction du temps pour une erreur minimale

Le coefficient de détermination fluctue à la hausse comme à la baisse en fonction des années mais reste proche de 1. Ce qui est satisfaisant pour la significativité du coefficient c en fonction du temps.

Par ailleurs, le package *ELT* fournit la fermeture de la table par âge en fonction des années d'observation.

Pour l'année 2018 par exemple, le coefficient c_{2018} est prélevé afin de réaliser l'extrapolation de la mortalité aux âges de 70 à 130 ans. Le graphique suivant est ainsi obtenu par application du SMR :

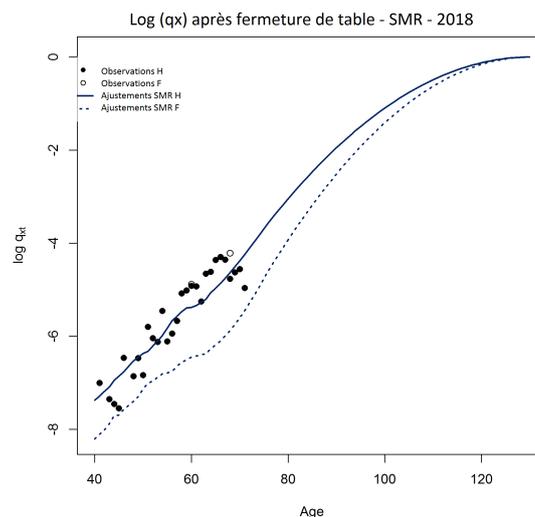


FIGURE 4.21 – Fermeture de table par la méthode de Denuit&Goderniaux - SMR - 2018

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

Ici la fermeture de table est réalisée sur une seule année. Si celle-ci s'applique à toutes les années alors les graphiques qui suivent sont obtenus. Chaque graphique correspond à un modèle d'ajustement des taux de mortalité.

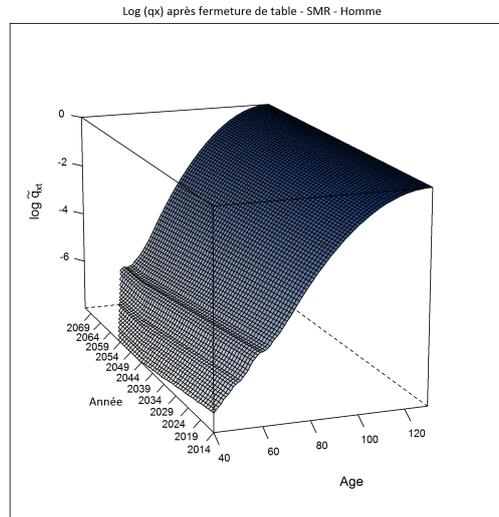


FIGURE 4.22 – Fermeture de table par la méthode de Denuit&Goderniaux - SMR

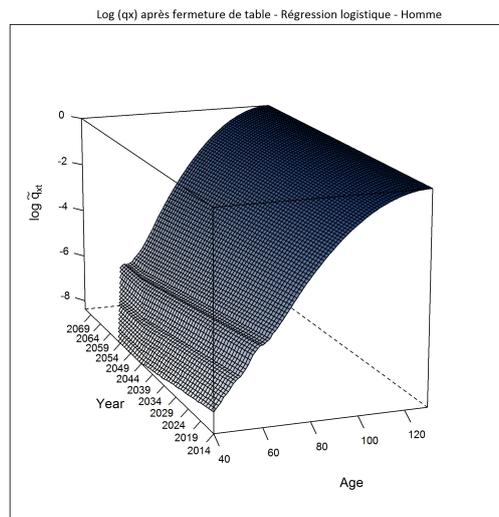


FIGURE 4.23 – Fermeture de table par la méthode de Denuit&Goderniaux - Régression logistique

Globalement la fermeture de table appliquée aux modèles du SMR et de la régression logistique sont assez similaires.

CHAPITRE 4. MÉTHODES DE CONSTRUCTION ET MODÉLISATIONS DES TABLES DE MORTALITÉS PROSPECTIVES

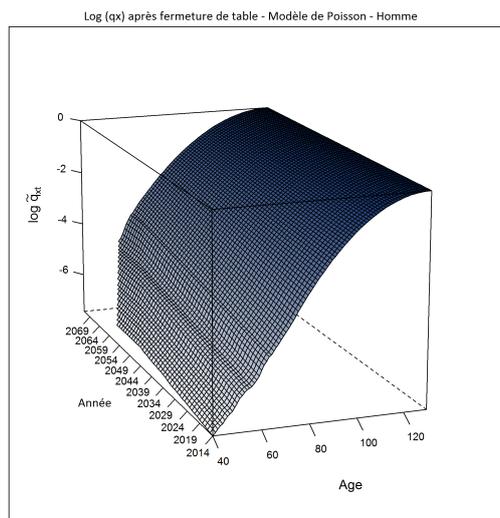


FIGURE 4.24 – Fermeture de table par la méthode de Denuit&Goderniaux - Modèle de Poisson

En revanche, l'application de la fermeture de table sur le modèle linéaire généralisé de Poisson donne des résultats bien différents des deux modèles précédents. Les probabilités de décès ont l'air d'être croissantes avec le temps, ce qui est contraire à l'hypothèse d'amélioration de l'espérance de vie. L'étape de validation donnera des indicateurs qui mesureront les écarts avec les deux autres fermetures de table.

Chapitre 5

Validations et impacts financiers

Dans la partie précédente, il a été question de présenter des modèles d'ajustement de taux brut de mortalité et de les appliquer. Les résultats des divers modèles testés donnent un ajustement des taux bruts de mortalité plus ou moins spécifiques et différents.

La comparaison graphique permet d'observer si les résultats sont absurdes ou grossièrement erronés, en revanche si les modèles sont globalement satisfaisants il est alors difficile de les comparer et de choisir le meilleur. Il existe alors des critères de validations quantitatifs permettant de juger du bon ajustement du modèle. Ces critères sont nombreux, quelques-uns d'entre eux seront présentés dans cette partie accompagnée de leurs résultats pour chaque modèle testé.

5.1 Critères de validation

5.1.1 Validation de fidélité des ajustements

Résidus de la réponse, de Pearson et de la déviance :

Principe de la méthode

La validation du modèle par le critère des résidus de la réponse se fait par analyse graphique. Il s'agit d'un critère de validation au niveau local qui compare les ajustements des modèles avec la mortalité observée.

Les résidus de la réponse [7] sont donnés par la formule suivante :

$$r_x(t) = \hat{q}_x(t) - \tilde{q}_x(t)$$

Cette expression fournit soit des graphiques dont l'abscisse est le temps soit des graphiques dont l'abscisse est l'âge ou les deux en fonction de son utilisation. L'analyse de ces graphiques permet de valider le bon ajustement des modèles en s'assurant qu'il n'y a pas de grands résidus au niveau local.

En effet, dans le cas parfait c'est-à-dire un ajustement exact avec l'observation, on aurait un graphique avec comme courbe une droite horizontale d'ordonnée nulle. Dans le cas réel on souhaite observer sur le graphique des points le plus proche de l'axe des abscisses en oscillant entre les valeurs négatives et positives de manière régulière sans saut. Ceci permet alors de confirmer qu'il n'y a pas d'anomalie dans l'ajustement du modèle.

Il existe aussi d'autres critères de validation au niveau local construit non plus sur l'écart entre l'ajustement des modèles et la mortalité observée mais plutôt sur le nombre de décès réel et le nombre de décès estimé par le modèle. On peut citer les résidus de Pearson et les résidus de la déviance.

Les résidus de Pearson [7] sont donnés par la formule suivante :

$$r_x(t) = (D_x(t) - E_x(t) \tilde{q}_x(t)) \times \sqrt{\text{Var}[E_x(t) \tilde{q}_x(t)]}$$

Les résidus de la déviance [7] sont donnés par la formule suivante :

$$r_x(t) = \text{signe}(D_x(t) - E_x(t) \tilde{q}_x(t)) \times \sqrt{\text{Deviance}_x(t)}$$

Les résidus de Pearson sont construits avec la variance alors que les résidus de la déviance sont construits avec la déviance. La spécificité de leurs expressions permet d'analyser les résidus par des approches différentes.

Une particularité est à souligner quant aux résidus de Pearson, ceux-ci doivent se trouver dans l'intervalle $[-2, 2]$. Dans ce cas, la variabilité des données est correctement modélisée.

Concernant les résidus de la déviance, le but est d'observer des résidus de signe opposé de façon successive pour valider que les données ne soient pas localement sur-lissées.

Application de la méthode

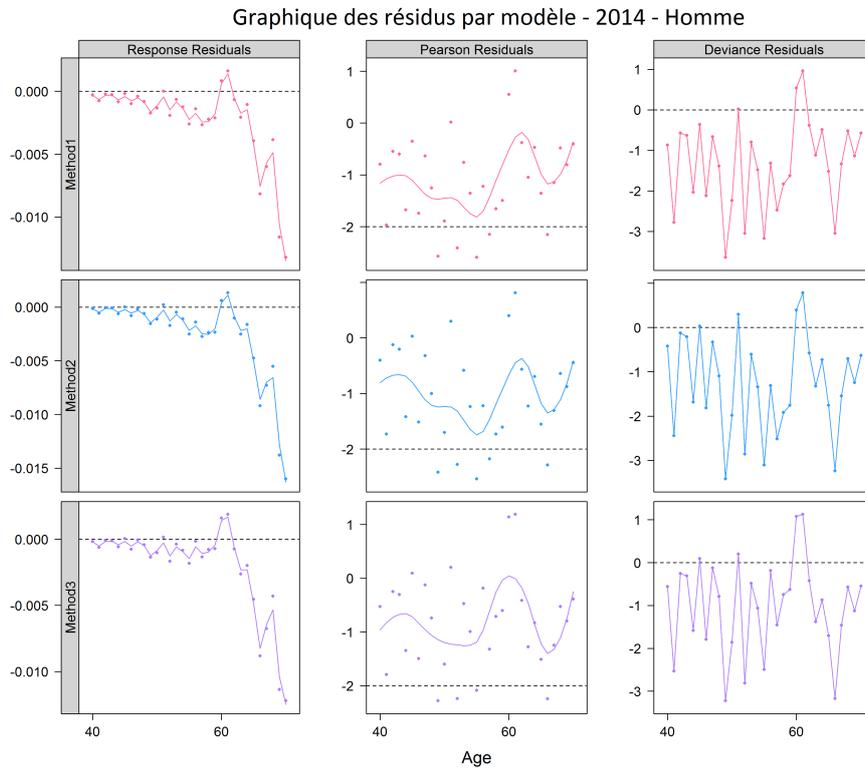


FIGURE 5.1 – Test des résidus - 2014 - Homme

Le graphique ci-contre présente les résidus de la réponse, de Pearson et de la déviance pour les modèles du SMR, de la régression logistique et du modèle linéaire généralisé de Poisson appliqués à l'année 2014.

Les résidus de la réponse sont proche de zéro entre 40 et 60 ans environ mais s'éloignent de l'origine à partir de 60 ans dans les valeurs négatives. Ce qui signifie d'après la formule des résidus de la réponse que l'ajustement des modèles surestime la mortalité observée, ceci étant vrai pour les trois modèles. Ce constat a été fait lors de l'analyse des probabilité de décès et est affirmé dans cette partie par analyse graphique.

Les résidus de Pearson sont globalement dans l'intervalle de validité avec quelques points en dehors, ce qui est plutôt satisfaisant. Toutefois une certaine volatilité est observable puisque pour des âges très proches, les résidus peuvent être très éloignés.

Enfin concernant les résidus de la déviance, l'alternance de signe des résidus n'est pas validée ici puisque les résidus sont en majorité négatifs oscillant entre -3 et 0. Une chose importante à souligner est que les résidus pour les trois modèles testés sont globalements similaires. Il est donc difficile de classer les modèles par leur qualité d'ajustement à travers le critère des résidus pour l'année 2014.

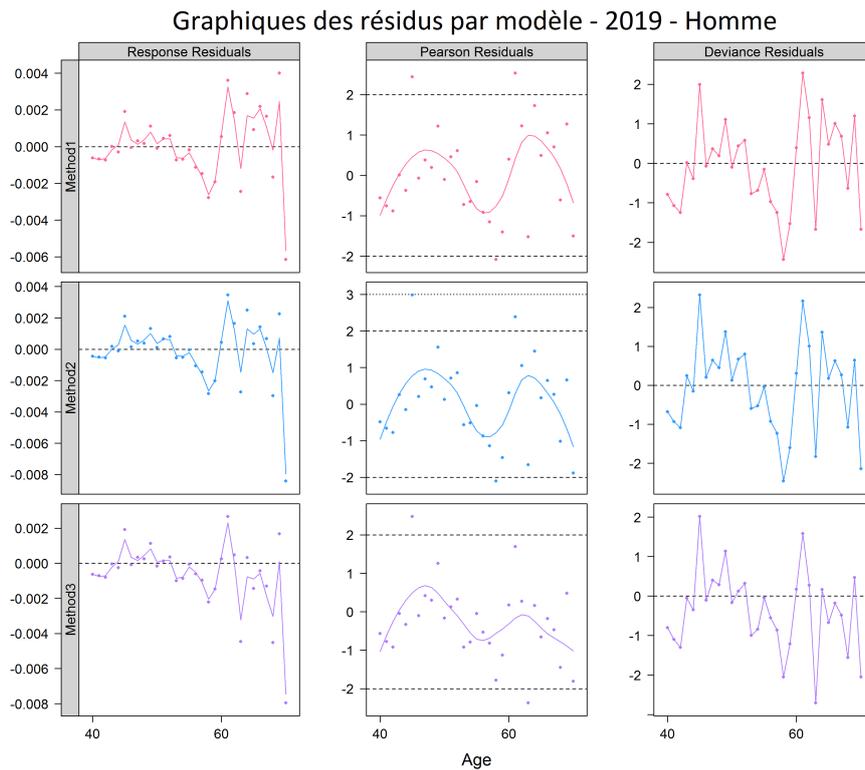


FIGURE 5.2 – Test des résidus - 2019 - Homme

Pour ne pas conclure trop vite sur les résultats des résidus, il est important d’analyser ces derniers pour différentes années d’observations. C’est pour cela que l’analyse des résidus continue pour l’année 2019 cette fois-ci.

Les résidus de la réponse offrent de bien meilleurs résultats pour l’année 2019 par rapport à l’année 2014 puisque les résidus sont proches de l’origine de 40 à environ 65 ans. Et la distance maximale avec l’origine est atteinte à 70 ans avec un écart 0,008 contre 0,015 pour l’année 2014, soit deux fois moins. De plus le signe des résidus de la réponse est bien plus éparpillé autour de l’axe des abscisses.

Pour les résidus de Pearson les résultats sont bien meilleurs que pour l’année 2014 puisque les points sont situés autour de zéro avec une nette majorité de point dans l’intervalle de validité. Cependant quelques points font exception et sortent de l’intervalle de validité tantôt au-dessus de 2 et en dessous de -2.

Quant aux résidus de la déviance, la courbe reliant les résidus oscille autour de l’axe des abscisses mais pas de façon strictement régulière. Ce n’était pas le cas pour l’année 2014, ça l’est un peu plus pour 2019 donc l’amélioration des résultats est bien constatée.

Globalement, les résultats pour les trois modèles sont meilleurs et assez proches encore une fois. Ce qui permet de vérifier que l’ajustement des modèles ne connaît pas de valeurs aberrantes ou disproportionnées. Néanmoins, ce n’est pas par une analyse graphique des résidus que le meilleur modèle pourra être déterminé. Il est donc nécessaire de réaliser des tests plus globaux donnant des résultats quantitatifs comparables.

Le critère de χ^2 :

La valeur du χ^2 [7] est un indicateur qui permet d'évaluer la qualité d'ajustement d'un modèle. Il se calcule de la façon suivante :

$$\chi^2 = \sum_{x,t} \frac{(D_x(t) - E_x(t) \tilde{q}_x(t))^2}{E_x(t) \tilde{q}_x(t)(1 - \tilde{q}_x(t))}$$

Le meilleur modèle sera celui qui aura une valeur de χ^2 la plus faible.

Le critère de *MAPE* :

Le critère *MAPE* [7] signifiant *Mean Average Percentage Error* est un indicateur en pourcentage évaluant l'exactitude de l'ajustement du modèle. Celui-ci s'obtient par l'expression suivante :

$$MAPE = \frac{\sum_{x,t} |(\hat{q}_x(t) - \tilde{q}_x(t)) / \hat{q}_x(t)|}{\sum_{x,t} D_x(t)} \times 100$$

Le meilleur modèle sera celui dont la valeur de *MAPE* sera la plus faible, c'est-à-dire le modèle qui aura le moins d'écart moyen en valeur absolue.

Le critère de R^2 :

L'adéquation entre le modèle et les observations se mesure par la valeur de R^2 [7]. C'est un indicateur qui se définit comme la part de variance expliquée par rapport à la variance totale. Le R^2 est défini comme :

$$R^2 = 1 - \left(\frac{\sum_{x,t} (\hat{q}_x(t) - \tilde{q}_x(t))^2}{\sum_{x,t} \left(\hat{q}_x(t) - \left(\frac{\sum_{x,t} \hat{q}_x(t)}{n} \right) \right)^2} \right)$$

Cet indicateur est compris entre 0 et 1. On privilégiera le modèle dont la valeur de R^2 sera la plus proche de 1

Application des critères

Ci-dessous les résultats des trois indicateurs précédemment présentés sur l'ensemble des modèles à la fois pour les hommes et pour les femmes.

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
χ^2	254,110	255,620	208,920	340,250	2,30433E+11	284,380
R^2	0,412	0,261	0,444	0,010	-0,018	0,020
MAPE	49,480	45,970	43,160	79,110	100	79,220

FIGURE 5.3 – Résultats des critères de qualité d'ajustement

Les résultats pour les femmes sont aberrants en constatant par exemple que l'indicateur de R^2 est négatif pour le modèle de régression linéaire. De plus pour les deux autres modèles, la valeur de R^2 est très faible, à savoir que le modèle explique respectivement 1% et 2% de la variance des probabilités de décès pour le modèle du SMR et du modèle linéaire généralisé de Poisson. A partir de ce constat, les autres résultats ne sont pas pris en considération.

En revanche, les résultats pour les hommes sont exploitables. Les valeurs des indicateurs sont du même ordre de grandeur pour chaque modèle, ce qui est rassurant. Il ressort que le modèle linéaire généralisé de Poisson est celui qui a les meilleurs résultats pour les trois indicateurs. En effet, ce modèle a la valeur du χ^2 la plus faible, la valeur de R^2 la plus élevée et la valeur de MAPE la plus faible.

5.1.2 Validation de la régularité des ajustements

Tests des signes

Principe de la méthode

Le test des signes est un test au niveau global à la différence des tests sur les résidus. Il porte sur la qualité du lissage, en effet ce test permet de connaître de manière quantitative s'il y a eu un sur-lissage ou un sous-lissage des données.

La mortalité ajustée par les trois modèles pour l'année 2014 montre que la mortalité réelle était surestimée, en effet les points réels étaient constamment en dessous de l'ajustement des modèles pour tous les âges et encore plus pour les âges proches de 40 ans. Le test des signes permet alors de chiffrer cette surestimation de l'ajustement sur les décès observés.

Le principe de ce test est basé sur la fréquence des changements de signes des écarts entre le taux de mortalité observé et le taux de mortalité ajusté. La bonne qualité de lissage du modèle est vérifiée lorsqu'il y a autant de signes positifs que de signe négatif. Pour valider ce test, la statistique du test des signes est évaluée par l'expression suivante :

$$\xi^{SIG} = \frac{|n_+ - n_-| - 1}{\sqrt{n}}$$

Ici n_+ et n_- comptent respectivement le nombre de signes positif et négatif avec $n = n_+ + n_-$. Pour savoir si la statistique du test des signes suit une loi Normale centrée réduite, la comparaison de cette statistique est faite avec le quantile de la distribution Normale centrée réduite de la façon suivante :

$$|\xi^{SIG}| > N_{\alpha/2}(0, 1)$$

Enfin, la quantité qui finalement permet de sélectionner le modèle respectant au mieux le test des signes est celui de la p valeur. La p valeur est calculée de la manière suivante :

$$\text{p-valeur} = \mathbb{P}[N_{1-\alpha/2}(0, 1) > |\xi^{SIG}|] = 2 \times (1 - F_{N(0,1)}(|\xi^{SIG}|))$$

Le modèle dont la p valeur sera la plus proche de 1 sera le meilleur modèle vis-à-vis de ce test.

Application de la méthode

Le test des signes est réalisé sur les trois modèles de ce mémoire. Les résultats sont présentés à la fois pour les hommes et pour les femmes.

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
n_+	80	86	82	31	31	31
n_-	106	100	104	155	155	155
ξ^{SIG}	1.8331	0.9532	1.5398	9.0188	9.0188	9.0188
Quantile	1.96	1.96	1.96	1.96	1.96	1.96
Hypothèse	H0	H0	H0	H1	H1	H1
p valeur	0.0668	0.3405	0.1236	0	0	0

FIGURE 5.4 – Résultats du test des signes

La ligne Hypothèse permet de conclure quant à la validation du test des signes. Lorsque l'hypothèse H_0 est retenue le test est validé, dans le cas contraire le test est invalidé.

Les trois tests sont donc invalidés pour les femmes alors que les trois tests sont validés pour les hommes. Reste à savoir lequel des trois modèles respecte au mieux le critère du test des signes. Il s'agit alors du modèle de la régression logistique qui a une p valeur de 0.3405, c'est la p valeur la plus grande parmi les trois modèles. Donc du point de vue de ce critère, le modèle de régression logistique est le modèle qui a le meilleur lissage parmi tous les modèles.

Test des runs

Principe de la méthode

Le test des runs est un test qui permet de déterminer si les éléments d'une séquence sont mutuellement indépendants. Ce test est un test qui repose également sur le signe des écarts entre la mortalité observée et la mortalité ajustée. Il est également construit avec le nombre de signes positif et négatif mais de façon différente car il fait intervenir le nombre de run. Un run est une série de signes identique peu importe la taille de la série, celui-ci peut être une série de signes positif ou une série de signes négatif.

La statistique du test des runs s'écrit de la manière suivante :

$$\xi^{RUN} = \frac{\text{Nombre de runs} - \mu}{\sigma}$$

Où μ et σ s'exprime de la façon suivante :

$$\mu = \frac{2 n_+ n_-}{n_+ + n_-} + 1 \quad \text{et} \quad \sigma^2 = \frac{2 n_+ n_- (2 n_+ n_- - (n_+ + n_-))}{(n_+ + n_-)^2 (n_+ + n_- - 1)}$$

Les variables intervenant dans ces quantités sont les mêmes que dans le test des signes.

Le but est le même que précédemment, l'hypothèse H_0 doit être validée pour conclure que la statistique suit bien une loi Normale centrée réduite.

La p valeur se calcule de la même manière que précédemment et le meilleur modèle est celui dont la p valeur est la plus proche de 1.

Application de la méthode

Les résultats du test des runs sont présentés en détail dans le tableau suivant pour les hommes et pour les femmes pour chaque modèle.

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
Nombre de runs	81	71	83	56	56	56
n_+	80	86	82	31	31	31
n_-	106	100	104	155	155	155
ξ^{SIG}	1.6773	3.3235	1.4465	0.8861	0.8861	0.8861
Quantile	1.96	1.96	1.96	1.96	1.96	1.96
Hypothèse	H0	H1	H0	H0	H0	H0
p valeur	0.0935	9e-04	0.148	0.3755	0.3755	0.3755

FIGURE 5.5 – Résultats du test des runs

Les résultats sont identiques pour les trois modèles chez les femmes et sont tous validés par ce test. Néanmoins, il ne faut pas prendre le résultat de ce test en considération car les données pour les femmes sont d'une part extrêmement faible et d'autre part, les paramètres de chaque modèle pour les femmes ne sont pas significatifs. Ainsi la non-robustesse des modélisations pour les femmes mène à des résultats erronés, c'est le cas ici pour les femmes.

Pour les hommes, le test des runs valide le modèle du SMR et le modèle linéaire généralisé de Poisson. Le modèle de la régression logistique est ressorti comme étant le meilleur pour le test des signes mais ressort comme le moins bon pour le test des runs.

Il faut noter que le test des signes et le test des runs portent tous les deux sur la qualité du lissage des ajustements mais doivent être considérés comme complémentaire. Ainsi la validation des deux tests est nécessaire pour conclure quant à la bonne qualité de lissage du modèle.

Ici le meilleur modèle est celui du modèle linéaire généralisé de Poisson avec une p valeur de 0.148 contre 0.0935 pour le modèle du SMR.

5.1.3 Validation des tendances d'évolution de la mortalité

Entropie

Principe de la méthode

L'entropie se définit comme une grandeur qui mesure la quantité de désordre dans un ensemble. Il s'agit de la définition communément utilisée en science physique, celle-ci s'apprête particulièrement à la thermodynamique et à la chimie entre autres.

De la même manière, ici si l'entropie mesurée est grande alors les décès sont désordonnés, dans le sens où ceux-ci interviennent à des âges éloignés par rapport à la moyenne. Si l'entropie est faible alors les décès sont concentrés, ordonnés autour de la moyenne.

L'entropie se définit comme :

$$H[\omega T_{\bar{x}}] = - \frac{\int_1^{\omega} S_{\bar{x}}(u) \ln(S_{\bar{x}}(u)) du}{\int_1^{\omega} S_{\bar{x}}(u) du}$$

La fonction de survie ici utilisée est la suivante :

$$S_{\bar{x}}(x) = \prod_{j=0}^{x-1} (1 - q_{\bar{x}+j}(t + j))$$

Application de la méthode

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
$H[50T_{50}]$	0,0119	0,0121	0,0319	0,0054	0,0000	0,0078
$H[60T_{50}]$	0,0538	0,0557	0,1099	0,0275	0,0001	0,0369
$H[70T_{50}]$	0,2417	0,2536	0,3886	0,1336	0,0038	0,1675
$H[80T_{50}]$	1,1345	1,1953	1,5052	0,6650	0,0532	0,7925

FIGURE 5.6 – Entropie à 50 ans sur différents horizons de temps

Le tableau présenté ci-contre donne les résultats de l'entropie à 50 ans sur 50, 60, 70 et 80 années pour chacun des trois modèles. Les données pour les femmes sont également affichées mais ne sont pas à considérer.

Les modèles du SMR et de la régression logistique obtiennent des résultats proches alors que le modèle linéaire généralisé de Poisson s'en éloigne. Par ailleurs, c'est le modèle linéaire généralisé de Poisson qui affiche l'entropie la plus élevée, c'est-à-dire le niveau de désordre le plus élevé.

Dans le contexte de décès à un horizon de temps lointain, l'entropie peut être assimilée à un niveau de volatilité. En ce sens, le modèle linéaire généralisé est le modèle qui présente la plus grande volatilité sur les âges de décès par rapport aux deux autres modèles. Le meilleur modèle à retenir par rapport au critère de l'entropie est celui qui présente l'entropie la plus faible, à savoir le modèle du SMR.

Age médian au décès

Principe de la méthode

L'âge médian au décès est la valeur médiane de la fonction de survie sur la variable aléatoire durée de vie d'un individu d'âge x . Celle-ci se définit comme suit :

$$S_{\bar{x}} (Med [\omega T_{\bar{x}}]) = 0.5$$

Cet indicateur donne l'âge restant à vivre tel que la probabilité de survie au-delà et en dessous de cet âge est de 50% pour un individu d'âge x sur ω année.

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
Med_{50T50}	39,19	39,02	32,40	44,49		42,05
Med_{60T50}	28,28	27,85	23,98	33,87		31,75
Med_{70T50}	17,95	17,37	15,90	23,42	45,85	21,68
Med_{80T50}	9,62	9,19	9,04	13,78	35,78	12,58

FIGURE 5.7 – Age médian au décès à 50 ans sur différents horizons de temps

Ici l'âge médian au décès est calculé sur les mêmes paramètres que pour l'entropie. Le manque de donnée pour les femmes ne permet pas de calculer l'âge médian à 50 ans sur 50 et 60 ans, ce qui explique l'absence de résultat dans le tableau.

Globalement les conclusions sont similaires aux résultats pour l'entropie. En effet, l'âge médian au décès pour le modèle du SMR et pour la régression logistique sont proches et meilleurs que le modèle linéaire généralisé de Poisson.

Donc ici encore, le meilleur modèle pour le critère de l'âge médian est le modèle du SMR car il présente l'âge médian le plus élevé et le plus réaliste des trois modèles par rapport à l'hypothèse d'amélioration de l'espérance de vie.

Espérance de vie générationnelle partielle

Principe de la méthode

L'espérance de vie générationnelle partielle se calcule en faisant intervenir la fonction de survie, elle se calcule pour un individu d'âge x sur ω années par l'expression suivante :

$$\omega e_{\bar{x}} = \int_1^{\omega} S_{\bar{x}}(u) du$$

De plus si l'hypothèse de constance des forces de mortalité pour les valeurs entières de x et t est vérifiée, alors l'expression se simplifie comme suit :

$$\omega e_{\bar{x}} = \sum_{x=1}^{\omega} \prod_{j=0}^{x-1} (1 - q_{\bar{x}+j}(t + j))$$

Application de la méthode

	Hommes			Femmes		
	SMR	Régression logistique	Modèle linéaire généralisé de Poisson	SMR	Régression logistique	Modèle linéaire généralisé de Poisson
50 ^e 50	35,90	35,79	29,88	41,06	49,99	39,02
60 ^e 50	25,74	25,35	21,80	31,31	49,52	29,27
70 ^e 50	16,22	15,73	14,27	21,26	43,86	19,63
80 ^e 50	8,69	8,33	8,00	12,29	33,85	11,21

FIGURE 5.8 – Espérance de vie générationnelle partielle à 50 ans sur différents horizons de temps

Encore une fois, l'espérance de vie est calculée avec les mêmes paramètres que les deux autres métriques qui ont été présentées précédemment. Les résultats pour les femmes ne sont pas significatifs et cohérents. Mais il en ressort que l'espérance de vie pour les femmes est toujours supérieure à celles des hommes.

Sinon pour les hommes, l'âge médian au décès donnait une indication des tendances produites par chaque modèle. Ces tendances se retrouvent à travers l'analyse de l'espérance de vie générationnelle partielle.

En effet, le modèle qui affiche la tendance de mortalité la plus forte est toujours le modèle linéaire généralisé de Poisson, celle-ci est bien en dessous de l'espérance de vie actuelle pour les hommes de 85 ans. Ce qui montre bien que les tendances de mortalité projetées pour ce modèle ne tiennent pas compte de l'amélioration de l'espérance de vie. Donc à ce stade, pour plus de prudence et en absence de donnée au-delà de 75 ans, le modèle du SMR sera celui à préconiser.

Enfin, l'étape d'analyse et de validation des tendances futures permet de constater que le modèle linéaire généralisé de Poisson projette des tendances de mortalité aberrantes et peu prudentes.

Ainsi malgré le fait que celui-ci ressort comme étant le meilleur en terme d'ajustement avec l'observation, celui-ci présente des faiblesses en termes de mortalité future projetée. De plus, dans le contexte d'un produit de retraite avec des sorties en rente les tendances de mortalité futures sont primordiales et même plus importantes que l'ajustement avec le réel même si cette dimension n'est pas à négliger.

Donc, en pratique il vaudrait mieux retenir le modèle du SMR qui est simple d'utilisation, robuste, bon en terme d'ajustement et surtout cohérent dans les tendances de mortalité projetées.

5.2 Impacts financiers

Cette partie a pour but de présenter l'influence de la nouvelle loi en fonction du modèle considéré sur des métriques financières. Une étape de Backtesting est réalisée afin de vérifier que la loi de mortalité coïncide avec les observations.

Ensuite, l'intérêt est porté sur l'évolution du taux de couverture, la PMT et l'équity qui sont les principales métriques à considérer pour l'aide à la décision sur ce régime. Ces métriques évoluent en fonction des arrrages, de la PTS et la valeur de service qui seront présentés dans la suite.

5.2.1 Backtesting des différents modèles sur le nombre de décès

L'étape de backtesting sur les décès consiste à tester si avec la nouvelle loi le nombre de décès estimé coïncide avec l'observation. Ce test est nécessairement réalisé sur les années antérieures. Les années les plus importantes pour ce régime compte tenu du contexte sont les dernières années, c'est pourquoi le backtesting est réalisé sur les trois dernières années d'observation, à savoir 2017, 2018 et 2019.

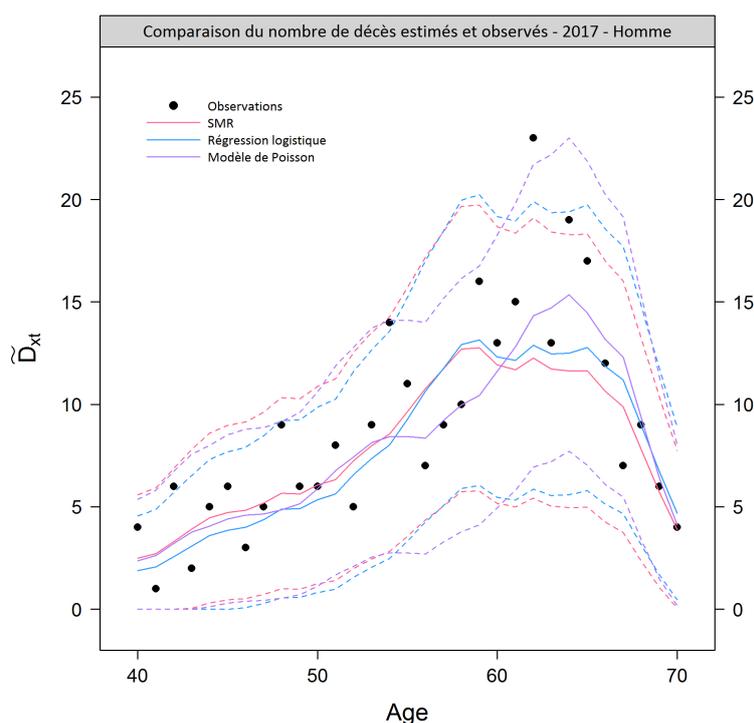


FIGURE 5.9 – Comparaison du nombre de décès estimé et observé - 2017 - Homme

Ce graphique illustre le nombre de décès estimé par chacun des modèles à chaque âge compte tenu de l'exposition du portefeuille pour l'année 2017. Les observations réelles y sont également inscrites et ces dernières sont représentées par les points noirs.

Les courbes en pointillé délimitent l'intervalle de confiance à 95% du nombre de décès estimé par les lois.

Ainsi pour l'année 2017, une observation est en dehors de l'intervalle de validité à l'âge de 62 ans. Mise à part cette observation, toutes les autres se situent dans l'intervalle de confiance des modèles.

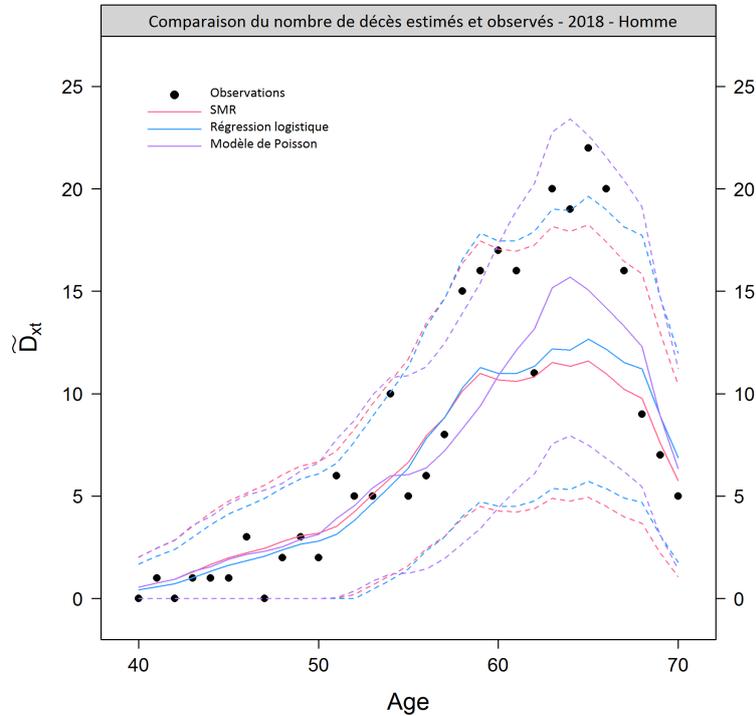


FIGURE 5.10 – Comparaison du nombre de décès estimé et observé - 2018 - Homme

Ce graphique représente le backtesting pour l'année 2018 et fonctionne de la même manière que précédemment.

Pour l'année 2018, le backtesting est plutôt satisfaisant puisqu'aucune observation ne se situe en dehors de l'intervalle de validité.

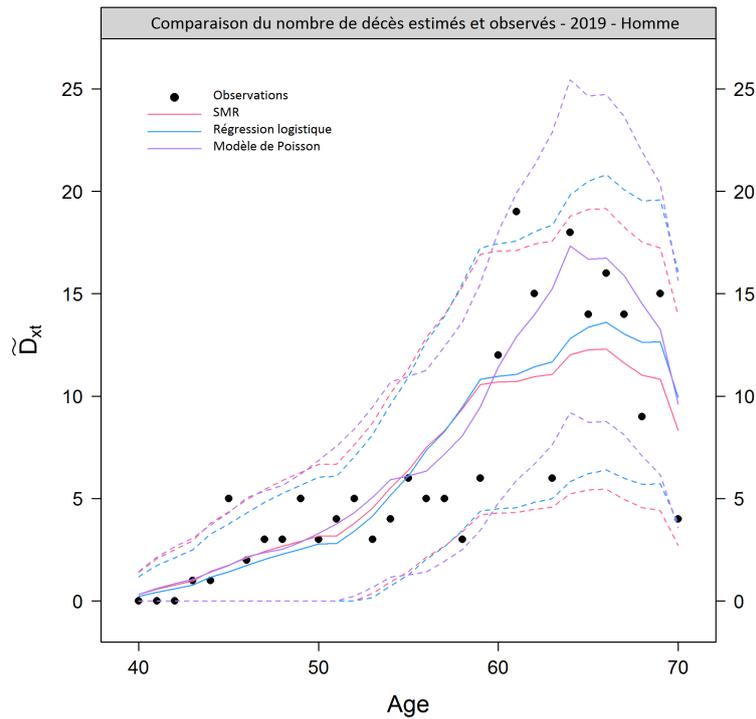


FIGURE 5.11 – Comparaison du nombre de décès estimé et observé - 2019 - Homme

Tout d’abord, de manière générale sur ces trois années de backtesting, le nombre de décès observé est inclus dans l’intervalle de confiance à 95% du nombre de décès estimé par chaque modèle. L’intervalle de confiance à 95% peut être considéré comme un intervalle de validité de notre test à la façon du test des résidus de Pearson.

Enfin pour l’année 2019, un point à l’âge de 45 ans est en dehors de l’intervalle de validité sinon les autres points sont correctement inscrits dans l’intervalle de validité des modèles.

Sur une vue d’ensemble, le modèle linéaire généralisé de Poisson surestime nettement la mortalité par rapport aux deux autres modèles sur la tranche d’âge 60 – 70 ans. C’est ce qui lui permet d’avoir des résultats plus proche de l’observation car en effet sur cette tranche d’âge la mortalité réalise un bond par rapport à la tranche d’âge antérieure.

5.2.2 Impact sur les arrérages, la PTS et la VS

Dans cette partie, quelques métriques financières intermédiaires sont présentées comme les arrérages, la provision technique spéciale (PTS) et la valeur de point également appelée valeur de service (VS).

Les données présentées ci-contre ne constituent pas les données réelles du régime car celles-ci sont confidentielles. Cependant, l'impact des nouvelles tables sur ces métriques est représentatif de l'impact réel. Il sera intéressant alors d'analyser ces évolutions par rapport aux données propres fournies par la TGHF-05.

Les assurés du régime reçoivent des points tout au long de leur vie active. Une fois à la retraite (55 ans pour ce régime), ceux-ci reçoivent des rentes qui sont calculées comme le produit de la VS avec le nombre de points acquis. Il est donc important de s'intéresser à l'évolution de la VS en fonction du temps.

La valeur de service (VS)

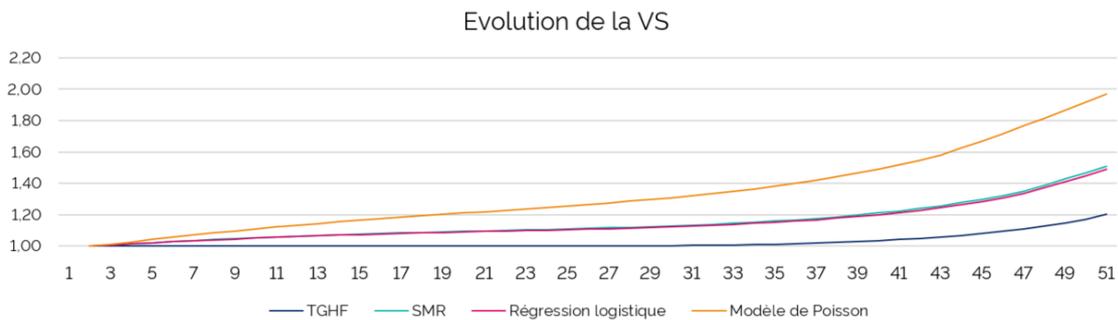


FIGURE 5.12 – Comparaison de la valeur de service

Ce graphique illustre l'évolution de la valeur de service en fonction du temps selon la table utilisée. Le calcul de la valeur du point en fonction du temps est fait par un algorithme de revalorisation du point dont le fonctionnement ne sera pas détaillé ici.

L'évolution de la VS à travers la table du SMR et de la régression logistique sont très proches et croît plus vite qu'avec la table TGHF-05. Cependant avec la table du modèle linéaire généralisé de Poisson, la croissance de la VS est encore plus marquée.

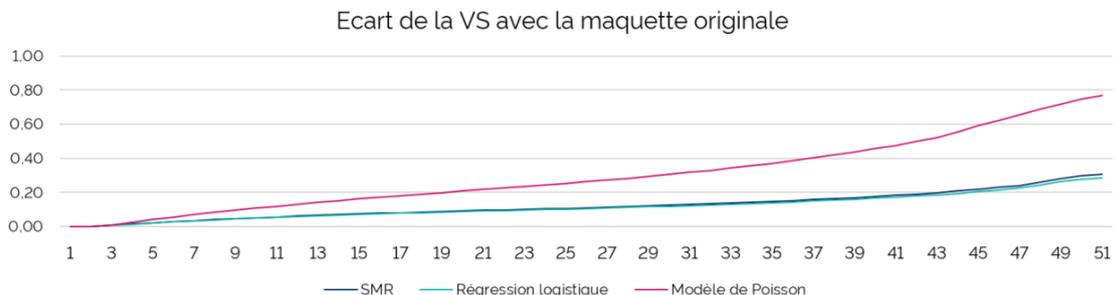


FIGURE 5.13 – Écart de VS avec la maquette originale (TGHF-05)

Afin de mieux comparer l'évolution de la VS, il vaudrait mieux observer l'écart de la VS de chaque table avec la VS de la table TGHF-05. Les premières années la VS évolue de manière similaire peu importe le modèle, en revanche à partir de 4 ans les écarts se manifestent et continuent à croître avec la maquette originale. Ainsi avec les nouvelles tables, les rentiers qui survivent le plus longtemps sont ceux qui bénéficieront des rentes les plus avantageuses.

Les arrérages

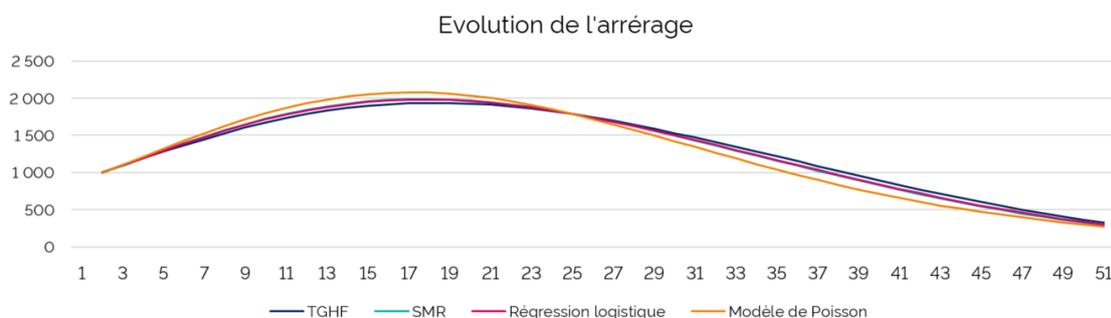


FIGURE 5.14 – Comparaison des arrérages

Le changement de table a un impact sur l'évolution des arrérages. Les arrérages sont la somme des rentes versées par le régime aux rentiers en une année. Ils sont calculés comme le nombre de points multiplié par la VS.

Ainsi les arrérages sont croissants et atteignent un pic à 17 ans environ. Il s'agit vraisemblablement du moment où le nombre de rentiers est parmi les plus élevés puisque l'âge moyen du régime en 2019 est autour de 55 ans, 17 ans plus tard la grande majorité du portefeuille sera composée de rentier. Ensuite les arrérages diminuent car le nombre de rentier commencera à baisser dû aux décès, le nombre de nouvelles jouissances sera proche de zéro et la croissance de la VS ne compensera pas la baisse de l'effectif de rentier.

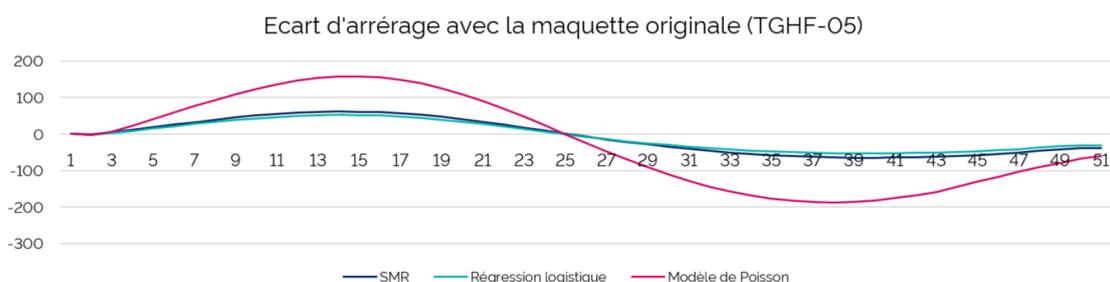


FIGURE 5.15 – Écart d'arrérages avec la maquette originale (TGHF-05)

Compte tenu de l'échelle, l'évolution des arrérages ne permet pas de distinguer correctement les différences entre les modèles c'est pourquoi le graphique présentant l'écart des arrérages est donné. Celui-ci montre que les arrérages croient plus vite pour les trois modèles et encore plus pour le modèle linéaire généralisé de Poisson par rapport aux arrérages obtenus avec la TGHF-05 sur la première partie du graphique. En revanche, à partir de 25 ans la tendance s'inversent pour tous les modèles. Encore une fois, les modèles du SMR et de la régression logistique fournissent des résultats assez proches.

La provision technique spéciale (PTS)

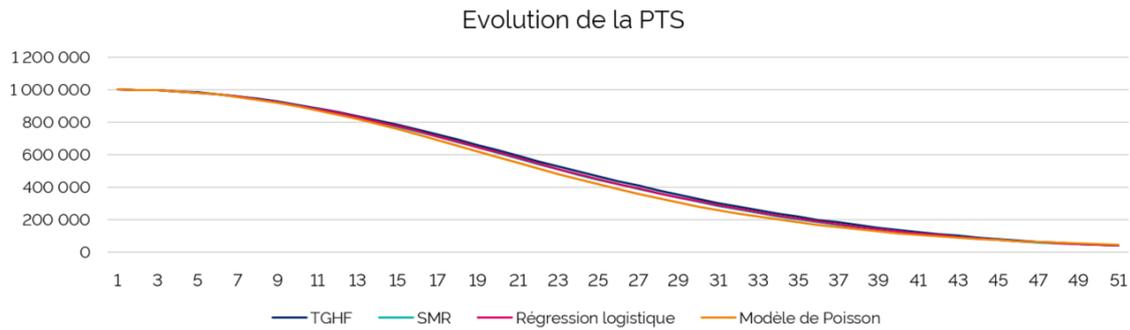


FIGURE 5.16 – Comparaison de la PTS

La PTS est constituée de l'ensemble des cotisations nettes de déchargements et de taxes. La PTS diminue avec le versement des arrérages. Comme les arrérages sont plus élevés pour les nouvelles tables, il est normal d'observer que la PTS diminue davantage qu'avec la TGHF-05.

Ce constat est avéré et est identifiable sur le graphique suivant illustrant l'écart de PTS des différents modèles avec la PTS de la maquette originale (TGHF-05).

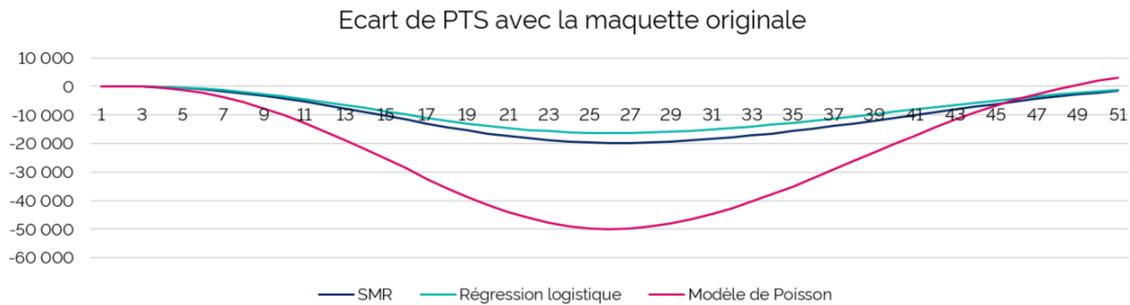


FIGURE 5.17 – Écart de PTS avec la maquette originale (TGHF-05)

Cependant à partir de 25 ans, l'écart de PTS avec la TGHF-05 diminue jusqu'à s'annuler au bout de 48 ans pour le modèle linéaire généralisé de Poisson. En revanche pour les deux autres modèles, la PTS reste toujours inférieure à la PTS de la TGHF-05.

5.2.3 Impact sur le taux de couverture, la PMT et l'equity

Les métriques financières servant à piloter le régime sont celles qui sont abordées dans cette partie. Il y a notamment la PMT, le taux de couverture et l'equity. Leurs variations dépendent des métriques qui ont été présentés précédemment.

La PMT est la provision mathématique théorique, celle-ci est égale à un instant t à la valeur actuelle des points acquis à servir sur la base de la dernière valeur de service connue. Celle-ci dépend directement de la table de mortalité utilisée.

Ensuite l'intérêt porte également sur le taux de couverture qui est le rapport entre la somme de la PTS et des plus ou moins-values sur la PMT. De ce fait, si la PMT baisse le taux de couverture s'améliore.

Enfin l'equity représente le gain de l'assureur. Celle-ci se définit comme la différence entre les prélèvements nets et les coûts. Les prélèvements nets comportent les chargements sur PTS et les coûts sont composés des coûts de gestions et des coûts de sorties pour ce régime. Cette décomposition de l'equity servira à expliquer l'évolution de celle-ci en fonction de la loi utilisée.

Hommes	Variation de PMT	Variation du taux de couverture	Variation d'equity
TGHF-05			
SMR	-7,98%	8,68%	-3,90%
Régression logistique	-8,36%	8,34%	-3,22%
Modèle linéaire généralisé de Poisson	-17,46%	19,22%	-8,85%

FIGURE 5.18 – Impacts financiers des nouvelles lois

Le tableau ci-contre résume l'impact des nouvelles lois sur la PMT, le taux de couverture et l'equity. Seuls les écarts en pourcentages sont donnés afin d'en apprécier l'évolution.

Les trois modèles fournissent des tables qui font baisser la PMT. Le modèle linéaire généralisé de Poisson étant celui dont l'impact est le plus élevé. Les modèles du SMR et de la régression logistique ont un impact similaire et deux fois moins important que le troisième modèle.

Naturellement, pour une PTS identique le taux de couverture s'améliore à la hauteur de la baisse de la PMT. Ainsi le modèle linéaire généralisé de Poisson présente le meilleur taux de couverture avec une hausse de 19%, suivi du SMR et de la régression logistique. Cet impact positif sur le taux de couverture est recherché et favorable au pilotage du régime.

Malgré la hausse du taux de couverture, l'equity est en baisse. Comme précédemment, le SMR et la régression logistique obtiennent des résultats similaires éloignés des résultats obtenus par le modèle linéaire généralisé de Poisson. L'equity dépend directement de la PTS puisque que celle-ci est constituée grâce aux chargements sur PTS, or la PTS pour les trois modèles est plus faible que la PTS de la TGHF-05. Ceci explique en partie que l'equity est plus faible pour les nouvelles lois.

En résumé, ce chapitre a permis de vérifier la validité des modèles par de multiples critères. Il en ressort que le meilleur modèle en terme d'ajustement est celui du modèle linéaire généralisé de Poisson, cependant les tendances futures données par ce modèle ne sont pas cohérentes et manquent de prudence vis-à-vis de l'incertitude que porte le portefeuille à l'étude.

Par ailleurs, l'absence de donnée pour les individus au-delà de 75 ans comporte un risque à ne pas négliger. Afin de contrôler et réduire ce risque difficile à quantifier, le modèle du SMR sera retenu car nécessite d'estimer un unique paramètre, ce qui réduit le risque d'estimation des paramètres de modèle. De plus ce modèle donne lieu à des tendances de mortalité futures cohérentes et plus prudentes que le modèle linéaire généralisé de Poisson.

En outre, les impacts financiers des trois modèles ont été mesurés afin de constater l'influence de la surmortalité sur les différentes métriques. Les trois modèles montrent une surmortalité générale du portefeuille, cette surmortalité est projetée pour les âges élevés et agit directement sur le taux de couverture et l'equity. Ainsi la surmortalité agit de façon favorable sur le taux de couverture mais de façon défavorable sur l'equity car provoque le déclin de la PTS. Les impacts sont de plus faibles amplitudes pour le modèle du SMR, c'est une raison de plus pour conserver ce modèle en pratique.

Conclusion

Dans cette présente étude, les méthodes de positionnement sont les principaux outils utilisés compte tenu des caractéristiques du portefeuille. Par ailleurs les données sur les femmes n'étant pas significatives, l'analyse qui suit ne concerne que la population masculine. Trois modèles ont donc été utilisés.

Le premier est le modèle du SMR. Celui-ci est un modèle à un paramètre dont l'estimation permet simplement d'observer si le portefeuille présente une sous-mortalité ou une surmortalité. Le second modèle est le modèle semi-paramétrique de Brass, communément appelé régression logistique. C'est un modèle à deux paramètres faisant intervenir la fonction logit. Le troisième modèle est le modèle linéaire généralisé de Poisson. Modèle à trois paramètres basé sur la modélisation du taux instantané de mortalité et faisant intervenir l'âge comme variable explicative.

L'application de ces trois modèles montre de façon unanime que le portefeuille à l'étude est caractérisé par une surmortalité sur le domaine d'application. L'analyse graphique des ajustements de la mortalité des trois modèles montre que les modèles coïncident assez bien avec l'observation. De plus, les courbes modélisées par les trois modèles sont très similaires. Une simple analyse graphique ne permet donc pas de déterminer le meilleur modèle. Il est donc nécessaire de procéder à une étape de validation quantitative pour évaluer la qualité d'ajustement des modèles.

Toutefois, un autre élément est à considérer. Il s'agit du risque d'estimation des paramètres. Celui-ci est d'autant plus important que la quantité de donnée dans la base de travail est faible. Or cette étude préconise des méthodes de positionnement car en effet le portefeuille à l'étude comporte peu de donnée en terme de volume. Ainsi le modèle linéaire généralisé de Poisson est le modèle le plus exposé à ce risque alors que le modèle du SMR est celui qui est le moins exposé à ce risque. Cette information est à prendre en compte en complément des tests de validation des modèles.

L'étape de validation commence par une validation au niveau locale. Le protocole consiste à juger du bon ajustement des modèles à travers l'analyse graphique des résidus de la réponse, de Pearson et de la déviance. Tout compte fait, l'analyse des résidus est conforme et ne présente pas d'anomalie pour les trois modèles. Une étape de validation au niveau global est alors envisagée afin de pouvoir juger et comparer les ajustements de la mortalité sur une base quantitative. Celle-ci fait intervenir plusieurs indicateurs qui testent différents aspects de la qualité d'ajustement. Il en ressort que le meilleur modèle est le modèle linéaire généralisé de Poisson, suivi du modèle de SMR et de la régression logistique.

Bien que le modèle linéaire généralisé de Poisson est le meilleur modèle en terme d'ajustement, cela ne suffit pas à le choisir en tant que nouvelle loi. Dans le contexte de la retraite, il est primordial d'évaluer les taux de mortalité projetés car ceux-ci ont une réelle influence sur le calcul des rentes et sur la revalorisation du point de service. Cette étape fait appelle à trois indicateurs : l'entropie, l'âge médian au décès et l'espérance de vie générationnelle. Les résultats des modèles de SMR et de régression logistique sont très proches, en revanche ils sont très éloignés du modèle linéaire généralisé de Poisson qui est le moins prudent des modèles. De plus, celui-ci projette des taux de mortalité croissants avec le temps, ce qui est incompatible avec l'hypothèse d'amélioration de l'espérance de vie. Le modèle linéaire généralisé de Poisson est donc le plus mauvais en terme de tendance projetée.

Outre l'aspect théorique vient l'aspect pratique en portant attention à l'évolution de certains agrégats financiers avec les nouvelles lois. Ainsi la surmortalité captée par les nouvelles lois fait baisser la PMT qui entraîne une hausse du taux de couverture, d'autre part l'*equity* est en baisse. Ce constat est vrai pour les trois lois avec des variations plus élevées pour le modèle linéaire généralisé de Poisson. Finalement, peu importe le modèle choisi le taux de couverture s'améliore malgré une baisse de l'*equity*.

Maintenant que les différents paramètres concernés par le changement de loi ont été identifiés et que leurs évolutions ont été mesurées. Le choix du modèle à retenir arrive en prenant en compte d'une part le risque d'estimation des paramètres de modèle, d'autre part le risque d'avis d'expert qui concerne le choix du domaine de calibration, puis les résultats obtenus par l'étape de validation et enfin l'évolution du taux de couverture. L'arbitrage retenu est de retenir le modèle du SMR en y ajoutant une marge de prudence afin de limiter les différents risques qui pèsent sur le portefeuille.

Par ailleurs, il a également été établi que cette étude est à réitérer ultérieurement lorsque le portefeuille sera plus mature et que la visibilité sur une tranche d'âge plus avancée sera meilleure.

Bibliographie

- [1] AKI QUASHIE et al. [2005]
Modèles d'extrapolation de la mortalité aux grands âges.

- [2] BASTIEN LAIZET [2018]
Apports de la théorie de la crédibilité à l'estimation de la mortalité.

- [3] CATHERINE HUBER-CAROL [1994]
Journal de la société statistique de Paris, tome 135.

- [4] FREDERIC PLANCHET et al. [2007]
Mesure de l'incertitude tendancielle sur la mortalité – Application à un régime de rentes en cours de service.

- [5] JULIEN TOMAS et al. [2014]
Construire une table de mortalité prospective : le package ELT

- [6] JULIEN TOMAS et al. [2014]
Méthodes de positionnement : Aspects méthodologiques

- [7] JULIEN TOMAS et al. [2014]
Critères de validation : Aspects méthodologiques

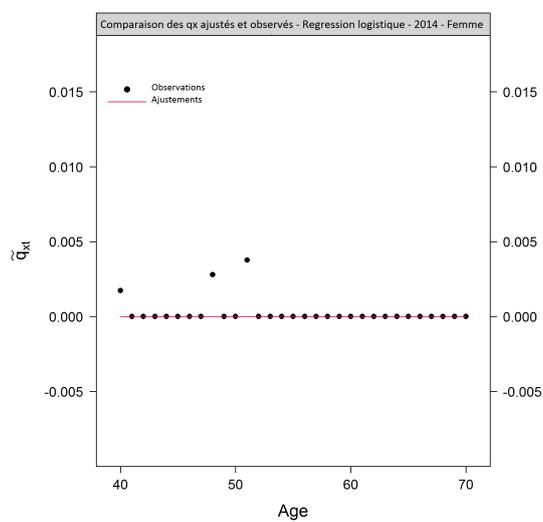
- [8] LEGIFRANCE [2017]
Article A132-18.

- [9] QUENTIN GUIBERT et al. [2017]
Utilisation des estimateurs de Kaplan-Meier par génération et de Hoem pour la construction de tables de mortalité prospectives.

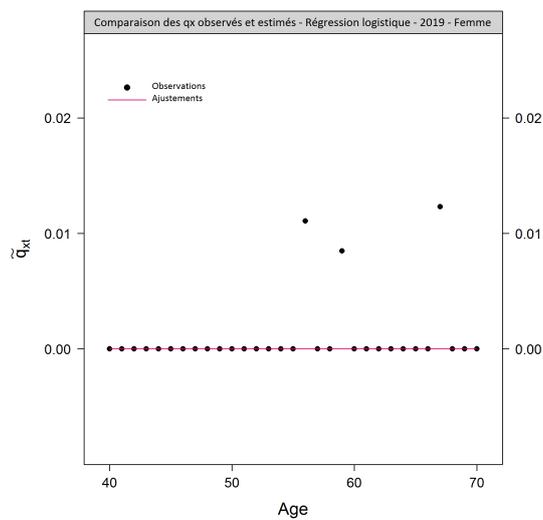
Annexes

Annexe

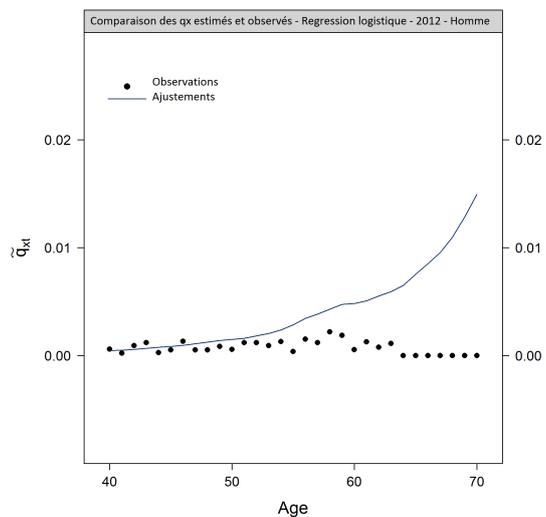
Annexe 1 : Comparaison du nombre de décès observés et estimés en 2014 chez les femmes



Annexe 2 : Comparaison du nombre de décès observés et estimés en 2019 chez les femmes



Annexe 3 : Comparaison des qx observés et estimés en 2012 pour les femmes



Annexe 4 : Comparaison des qx observés et estimés en 2013 pour les femmes

