

**Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuares**

Par : Arona BOYE

Titre du mémoire : Construction d'une table de mortalité d'expérience pour un portefeuille de rentes viagères

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

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

Membres présents du jury de la filière :

Signature :

Entreprise : Axa France

Nom :

Signature :

Directeur de mémoire en entreprise

Membres présents du jury de l'Institut des Actuares :

Signature :

Nom : CHOCHA YONI

Signature : *Chochalf*

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 :

Chochalf

Signature du candidat :

Arona



INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS

MÉMOIRE ACTUARIAT

Construction d'une table de mortalité d'expérience pour un portefeuille de rentes viagères



Arona BOYE
Août - 2019

Tuteur : M. CHOCHA YONI
Dr Isup : M. LOPEZ OLIVIER

CONFIDENTIEL

Resumé

Ce document présente un ensemble de méthodes pour déterminer la mortalité d'expérience d'une population donnée. L'étape descriptive permet d'avoir une première vision par des indicateurs statistiques du portefeuille considéré. La construction d'une table du moment par la suite est l'étude statistique de la mortalité qui permet d'avoir une vision précise sur une période de neuf ans. Cette étape est nécessaire car grâce à elle, nous pouvons déjà avoir une première indication en ce qui concerne la mortalité du groupe étudié.

L'analyse prospective permet de faire une projection dans le temps de la mortalité afin d'anticiper toute éventuelle dérive de mortalité. Nous implémentons d'une part le modèle paramétrique de Lee-Carter d'autre part trois méthodes d'ajustement via l'approche par positionnement par rapport à une référence externe via le package ELT. Les critères de validation nous permettent par ailleurs de retenir le modèle qui explique le mieux la mortalité de notre portefeuille.

Cette étude étant faite dans le cadre de la gestion du risque de longévité par un assureur, il est intéressant d'utiliser les données de survie déterminées afin de calculer les capitaux à constituer pour une rente viagère annuelle. Ce calcul permet par ailleurs une analyse concrète de l'application des taux d'expérience par rapport aux taux de mortalité des tables réglementaires.

Il est important de noter que le risque de longévité est le principal risque dans un portefeuille de rentiers. Aussi, maîtriser ce risque permet un meilleur pilotage par l'assureur de son portefeuille de rentiers. De plus, les données de mortalité interviennent dans la détermination de l'encours à un instant t donné, et servent de ce fait au provisionnement. Ce document n'a cependant pas vocation à déterminer l'encours ou encore la provision mathématique qui pourrait faire l'objet d'une autre analyse.

Mots clés : Table de mortalité, Lee-Carter, Fonction de survie, Lissage, Extrapolation, Rentes viagères, Capitaux constitutifs.

Abstract

This document presents a set of methods for modelling experienced mortality in a given population. Descriptive analysis step allows a first overview of statistical indicators of considered portfolio. Construction of a period life table allows a statical study of mortality and therefore a precise vision over a period of nine years. This step is necessary since it will allow us to have a first indication about mortality of people studied sample.

The prospective analysis makes it possible to make a projection over time of mortality in order to anticipate any possible drift in mortality. We implement on the one hand the parametric model of Lee-Carter then three methods of adjustment via the approach by positioning compared to an external reference via the ELT package. The validation criteria also allows us to retain the model that best explains the mortality of our portfolio.

Since this report is done in the context of longevity risk management by an insurer, it is interesting to use survival data determined for pricing an annual life annuity. This allows a practical application of experienced death rate compared to death rate of regulatory tables.

It is important to note that longevity risk is the main risk in a pensioner's portfolio. So, control this risk allows a better management by insurer of his annuitants portfolio. Mortality data are used to determine reserve and the amount of stock at any given moment. This report is not intended to determining stock or mathematical provisions that could be the subject of another analysis.

Keywords : Life table, Lee-Carter, Survival function, Smoothing, Extrapolation, Annuities.

Note de synthèse

Contexte

Nous vivons dans une époque mouvementée, caractérisée par des changements politiques, sociaux, économiques et technologiques. Ces évolutions notamment démographiques entraînent des adaptations constantes des ressources et prestations pour les assurés comme pour les assureurs. Certaines tendances sont alors à étudier avec soin comme le vieillissement de la population, qui a un impact non négligeable sur l'estimation des engagements des assureurs.

Ce mémoire s'inscrit dans ce cadre et réalise une étude complète pour le pilotage du risque de longévité d'un portefeuille de rentes viagères de contrats de retraite collective. Le calcul des engagements de l'assureur, dans le cadre de rentes viagères est réalisé à partir de tables réglementaires, TGH TGF 05, issues des données de la population nationale et des données d'assureurs. Elles proposent une évaluation des engagements de manière prudente pour un organisme d'assurance mais qui ne correspond pas en tout point à leurs portefeuilles d'assurés. Pour pallier ce risque technique, la réglementation propose aux assureurs la mise en place de tables de mortalité dites d'expérience.

Ce mémoire s'organise en quatre grandes parties :

- Présentation et étude statistiques des tendances des données de notre portefeuille ;
- Présentation et étude des indicateurs du suivi de risque de longévité de notre portefeuille ;
- Construction d'une table de mortalité d'expérience ;
- Application et analyse d'impact sur la population de notre portefeuille.

Etude de notre portefeuille

Les graphiques ci-dessous représentent la répartition des rentiers exposés au risque entre 2009 et 2017 ainsi que les décès par âge. Nous observons des individus entre 16 et 111 ans dont une exposition quasi nulle avant 60 ans. Les rentiers observés entre 21 et 60 ans correspondent aux réversataires ou à des retraites anticipées. Au vu de leur faible exposition, nous décidons de nous concentrer sur les rentiers entre 60 et 110 ans pour la suite de notre étude.

On observe une tendance à la hausse des décès dans le portefeuille, cette remarque est corrélée par l'augmentation du nombre de rentes présentes dans le portefeuille. Toutefois, cette analyse doit être considérée avec beaucoup de prudence. En effet, nous observons de grandes fluctuations des décès, dues à l'effectif relativement faible du portefeuille considéré.

De plus, le nombre de décès des hommes est beaucoup plus grand que les femmes, soit 65% du nombre de décès, c'est identique avec le nombre de rentes.



Effectif par année d'observation									
	2009	2010	2011	2012	2013	2014	2015	2016	2017
Rentiers	67357	70453	73249	75244	79028	81445	89408	92056	95487
% Homme	67%	67%	67%	67%	66%	66%	65%	65%	65%
Décès	1431	1541	1553	1585	1697	1692	2079	2198	2216

Les indicateurs du suivi de risque de longévité

Pour analyser l'évolution de notre portefeuille en terme de longévité, nous regardons deux indicateurs : le SMR (Standardised Mortality Ratio) et le résultat de mortalité. Le SMR permet de comparer nos observations par rapport à des estimations, la plupart du temps issues des tables de mortalité réglementaires. C'est le même principe pour le résultat de mortalité mais en montant, en comparant ainsi les provisions mathématiques calculées à partir des tables réglementaires par rapport aux décès réellement observés.

Le résultat de mortalité quantifie le gain (ou la perte) dû aux écarts entre la mortalité observée a posteriori et la mortalité théorique (au sens des tables de mortalité réglementaires) qui a servi aux calculs des provisions mathématiques. Il peut être un gain ou une perte : Une perte si le rentier est en vie à la fin de l'année : $q_x^{obs} = 0 < q_x^{th}$, un gain si le rentier décède pendant l'année : $q_x^{obs} = 1 > q_x^{th}$. Il est calculé à l'aide de la formule suivante : $RM = PM_{ouv} * (1 + i) - \frac{R}{k} * (1 + f) * (1 + i)^{\frac{k-1}{k}} * \sum_{j=0}^t v^{\frac{j}{k}} * (1 - \delta_{j,t}) - p^{obs} PM_{clot}$

En observant les deux graphes ci-dessous, nous remarquons une très forte volatilité du résultat de mortalité d'année en année. Ceci est justifié par deux composantes : le montant de la rente et la probabilité de décès estimée.

L'étude sur les rentes montre qu'il y'a une forte corrélation entre le résultat de mortalité et le montant des rentes de sortie du portefeuille (décès). En effet, nous remarquons que lorsque nous avons des résultats proches de 0, nous observons le décès d'un rentier avec

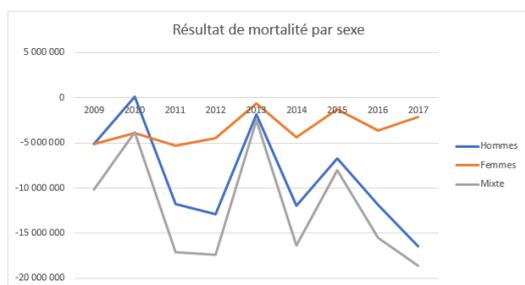


FIGURE 1 – (a)

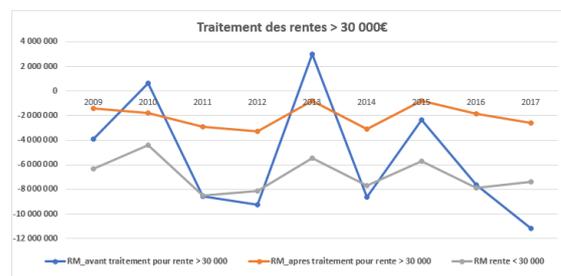


FIGURE 2 – (b)

FIGURE 3 – Résultat de mortalité (a), Réseultat de mortalité après traitement rente (b)

une forte rente. Nous décidons ainsi de jeter un regard sur le résultat lorsque les rentes sont en-dessous d'un certain seuil à définir. A l'aide de l'étude des quantiles, nous capons les rentes à 30000€ et nous définissons ce montant comme le seuil. Nous recalculons un résultat de mortalité par année avec les nouvelles rentes et obtenons les résultats suivants. On remarque alors une baisse de la volatilité des résultats qui reste cependant importante (environ 18 millions initialement contre 8 millions après lissage des montants de rentes par exemple entre 2011 et 2012).

Une deuxième composante influence le résultat de mortalité : les probabilités de décès théoriques. En effet, la réglementation nous impose l'utilisation de tables de mortalité générationnelles pour le calcul de nos engagements mais celles-ci ne correspondent pas en tous points à l'évolution de la mortalité de notre portefeuille de par leurs constructions. Nous avons notamment pu mettre en lumière cette observation grâce à l'étude du SMR.

Tous ces résultats convergent vers une même conclusion : nos rentiers ont une tendance de mortalité différente des tables de mortalité réglementaires et cet effet est démultiplié par le poids des montants de rentes. Ce risque étant prédominant, il est important de pouvoir anticiper l'évolution de cette dérive viagère. Pour cela, nous proposons dans un premier temps une méthode de segmentation du portefeuille pour déterminer une tendance de mortalité au travers de différents profils de risque.

Amélioration du résultat de mortalité : Mise en place d'une classification pour définir des profils de risque

Au vu de l'analyse de données précédente, nous avons remarqué que le résultat global comportait la même tendance que le résultat de la population masculine, que les montants de rentes de base des rentiers ainsi que la présence d'un réversataire. Cette dernière information est non négligeable car lorsque la première tête décède (l'assuré), la provision

mathématique n'est pas entièrement libérée mais seulement d'une certaine quantité, correspondant à la quantité de réversion. Une nouvelle provision est calculée pour la seconde tête (le réversataire). Le gain technique est alors moins important qu'en cas d'absence de réversion. Nous allons donc découper en plusieurs classes notre population selon ces différents critères. Ensuite nous projeterons les résultats pour estimer le résultat de mortalité de l'année en cours.

Pour se faire, nous avons utilisé la méthode des $K - means$ pour dégrossir notre base de données doublée à une classification ascendante hiérarchique CAH pour une segmentation beaucoup plus précise afin d'obtenir entre 7 et 8 clusters.

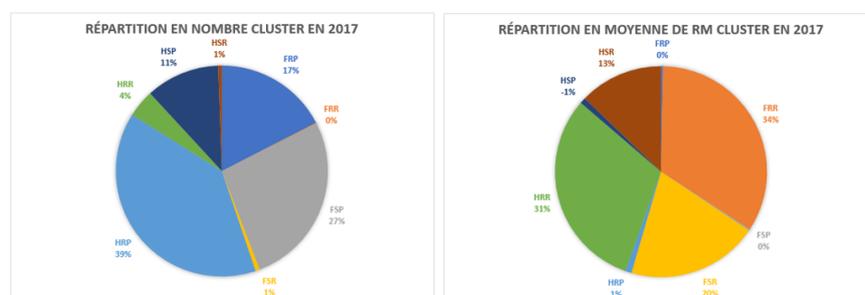


FIGURE 4 – Répartition des clusters pour l'article 39

Chaque année est marquée par huit clusters pour les rentes d'articles 39 et sept clusters pour les rentes d'articles 83. Les clusters obtenus sont également expliqués par les 3 variables SEXE, REVERSION et RICHE. Ainsi, nous décidons d'appeler les sept/huit clusters de la façon suivante : FRP, FRR, FSP, FSR, HRP, HRR, HSP, HSR. La première lettre correspond au SEXE représentatif du cluster (H pour Hommes et F pour Femmes), la seconde représente la présence ou non d'un réversataire (R pour Réversion et S pour seul) et la troisième si les rentiers qui composent le cluster ont des rentes élevées (R pour Riche et P pour Pauvre).

Construction d'une table de mortalité d'expérience

La construction de telles tables est une procédure délicate et deux approches sont disponibles. D'abord, construire une table du moment qui prendra en compte qu'une variable essentielle : l'âge. Ensuite, construire une table de mortalité prospective pour prendre en compte les évolutions futures de la mortalité.

Table du Moment

Dans cette partie, plusieurs méthodes sont confrontées afin de choisir (à l'aide des indicateurs statistiques notamment le SMR et le test des signes) celui qui explique le plus

fidèlement possible les données de notre portefeuille.

Pour estimer les taux bruts, la méthode de Kaplan Meier présente de meilleurs résultats par rapport à la méthode de Hoem. Mais les deux estimateurs convergent vers le même objet, ce qui montre que le résultat est satisfaisant. Pour remédier aux fluctuations des taux estimés deux méthodes ont été choisies : Whittaker-Henderson et lissage Géométrique. Enfin, nous avons ajusté les taux bruts par deux méthodes paramétriques : Makeham et Tatcher.

	Whittaker-Henderson	Makeham	Tatcher	Lissage Géométrique
SMR moyen	102.9%	115.8%	114.8%	102.8%
σ moyen	0.021	0.21	0.03	0.020
Nb dans l'IC	30	14	17	31

TABLE 1 – SMR, σ , Nombre de probabilité de décès l'IC pour chaque modèle

Au premier abord, nous constatons que la moyenne des SMR est supérieure à 1 pour chaque modèle, les taux bruts sont sous-estimés. D'autre part, les lissages de Whittaker-Henderson et Géométrique présentent les meilleurs SMR et plus de valeurs dans l'intervalle de confiance, contrairement aux modèles paramétriques qui sont très éloignés de la réalité. En effet, les méthodes de lissages ne supposent pas de loi a priori afin d'être le plus proche possible des données. Donc, à ce stade le lissage de Whittaker-Henderson et le lissage géométrique sont retenus.

Puis, le test de changement des signes valide l'hypothèse H_0 pour les deux méthodes, avec une précision de 100% pour Whittaker-Henderson. De plus, une analyse de notre portefeuille montre que la méthode géométrique n'est pas adéquate pour notre portefeuille, car ne disposant que de 32 âges, un ordre de 5 implique de prendre en compte un nombre non négligeable d'observations.

En fin de compte, nous avons retenu le lissage par Whittaker-Henderson et avons choisi la méthode de Denuit et Goderniaux pour extrapoler les taux de mortalité aux grands âges car, elle est simple à mettre en œuvre et présente des résultats satisfaisants. Par conséquent, les résultats de la fermeture de la table sont très satisfaisants avec un coefficient de détermination R^2 de 99,73% avec une faible p – *value*.

Afin de valider la table complète construite, nous avons étudié le positionnement de la table construite par rapport aux tables règlementaires. Nous avons donc construit une table de référence en pondérant par le *sexe – ratio* les tables TGH TGF 05 par le poids des pourcentages d'effectifs Homme et Femme. Ensuite, nous émettons l'hypothèse qu'il existe une relation affine entre les logits de notre table et la table de référence : $\text{logit}(q_x) =$

$$a * \text{logit}(q_x^{ref}) + b + \epsilon.$$

Enfin, nous pouvons dire que nous avons une bonne adéquation des logits d'expérience aux logits de référence compte tenue de la valeur élevée du R^2 ajusté, soit 99,64%.

Table de mortalité d'expérience prospective

L'objectif de cette partie est donc de faire une analyse prospective de la mortalité afin de pouvoir anticiper une éventuelle dérive dans le temps. Pour cela, nous allons dans un premier temps, mettre en oeuvre plusieurs méthodes d'ajustement prospectives, ensuite nous ferons une analyse comparative de ces différentes méthodes afin de ne retenir que celle qui correspond statistiquement le mieux à nos données réelles. Rappelons que, nous disposons des données d'un portefeuille de rentiers sur un historique de 2009 à 2017, soit 9 ans. De plus, en suivant la même démarche qu'au chapitre précédent, nous arrivons à une plage d'âge [63,94] ans.

Premièrement, nous avons ajusté les taux bruts par la méthode de Lee Carter. Nous obtenons donc les résultats suivants.

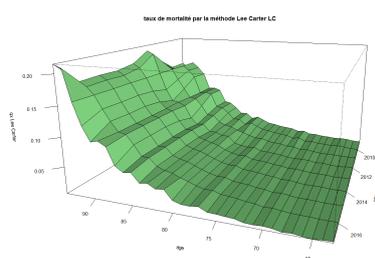


FIGURE 5 – Lee Carter (a)

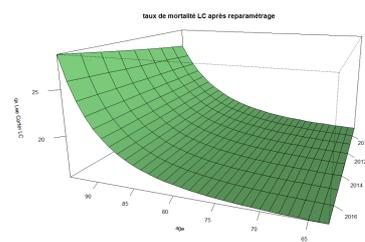


FIGURE 6 – reparamétrage (b)

On peut conclure qu'avec la première approche (sans reparamétrage des coefficients), nous remarquons que le lissage réalisé par la méthode de Lee-Carter n'est pas optimal. Puisque nous sommes dans un contexte où les effectifs (environ 100000 lignes) sont relativement faibles, il subsiste donc des fluctuations d'échantillonnage, qui conduisent à des irrégularités sur les paramètres estimés et donc sur les taux estimés.

Nous remarquons alors une amélioration du modèle de Lee Carter par le reparamétrage des coefficients du modèle dans le sens où nous observons moins de fluctuations des taux estimés. Ainsi, après l'analyse des intervalles de confiance du nombre de décès, les décès prédits par le modèle sont proches des décès observés. Ce qui rend le modèle satisfaisant globalement.

Au vu de la taille limitée de notre portefeuille (environ 98313 rentiers pour 9 ans d'observations), il semble difficile d'envisager uniquement une construction endogène des tables d'expériences comme nous l'avons commencée dans la section précédente. Nous avons alors

construit une table de mortalité d'expérience avec notre portefeuille en se référant à une table externe. Pour se faire, nous avons testé trois méthodes de positionnement via le package ELT suite aux travaux de Tomas et Planchet et Wassim YOUSSEF sur la longévité.

Méthode 1 : $\tilde{q}_x(t) = SMR \times q_x^{ref}(t)$

Méthode 2 : $\text{logit}(\tilde{q}_{x^*}(t^*)) = \alpha + \beta \text{logit}(q_{x^*}^{ref}(t^*)) + \epsilon_{x^*,t^*}$

Méthode 3 : $\mu_{x^*}(t^*) = \beta_0 + \beta_1 \log(q_{x^*}^{ref}(t^*)) + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^*$

Afin de trouver la meilleure méthode qui ajuste nos données, nous utilisons les critères de validation proposés au sein de l'outil ELT. Ils se divisent en trois groupes, le premier consiste à l'analyse des résultats obtenus en ce qui concerne les critères de proximité (R^2 , $MAPE$, SMR , $dviance$, ...), le second concernant la régularité de l'ajustement (on retrouve les tests de runs et des signs) et le troisième vérifie la cohérence des tendances d'évolution de la mortalité extrapolée.

Tous ces critères nous conduisent à la même conclusion que la méthode 3 et la méthode 2 présentent de meilleurs résultats pour notre portefeuille. Toutefois, Il est difficile de juger, à première vue, de la pertinence d'une table ou d'une autre car les résultats obtenus sont très proches entre eux. La fluctuation d'échantillonnage rend l'interprétation des tests statistiques difficiles et contradictoires. Nous décidons donc de conserver les trois méthodes ainsi construites par sexe, et nous choisirons la plus appropriée au vu du SMR de l'espérance de vie et des capitaux constitutifs.

Application et analyse d'impact sur notre portefeuille

Nous avons à présent analysé et calibré différents modèles (modèle de Lee Carter et les trois modèles de positionnement via le package ELT). Il convient d'analyser leur impact sur notre portefeuille permettant ainsi de comparer ces modèles afin de retenir celui qui permet de mieux capter la mortalité de la population de notre portefeuille de rentiers.

SMR et Espérance de vie

Nous remarquons, des mouvements similaires entre les tables d'expériences et les tables réglementaires. On retrouve ici les translations réalisées au travers de l'ajustement par positionnement des tables. De plus, le modèle de positionnement, la méthode 2, présente de meilleurs résultats en termes de SMR car, il est plus proche de la table TGH, tout en surestimant les probabilités de décès. Nous remarquons également que les méthodes implémentées prédisent une tendance de l'espérance de vie générationnelle partielle relativement similaire, qui se situe à un niveau supérieur par rapport aux données brutes.

Au vu des analyses ainsi que le backtesting via l'intervalle de confiance des décès prédits, nous retenons la méthodes 2 (Approche à deux paramètres avec un modèle relationnel) et la méthode 3 (modèle linéaire généralisé de Poisson) comme les meilleurs modèles pour nos données, ils présentent des résultats satisfaisants.

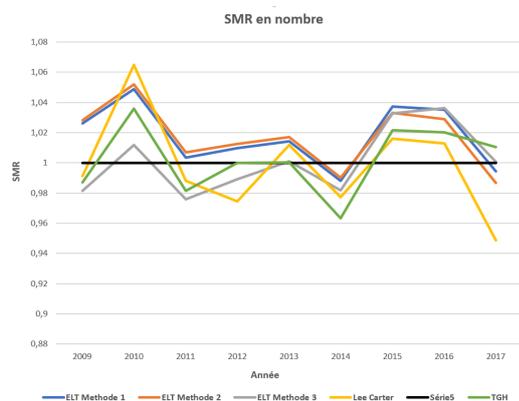


FIGURE 7 – SMR par année

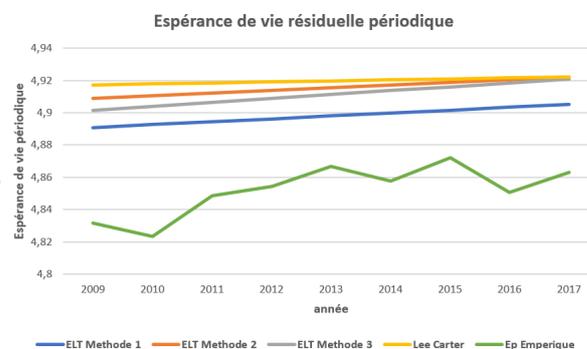


FIGURE 8 – Espérance de vie par année

Capitaux constitutifs de rente

Capitaux constitutifs : Le coût du versement d'un euro de rente annuelle à un rentier à partir d'un âge x ans est défini comme : $C_x = \sum_{k=0}^w v^k k p_x$

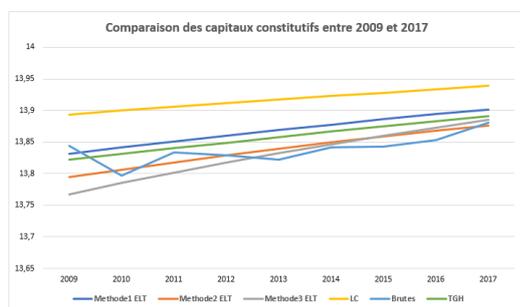


FIGURE 9 – Capitaux entre 2009 et 2017

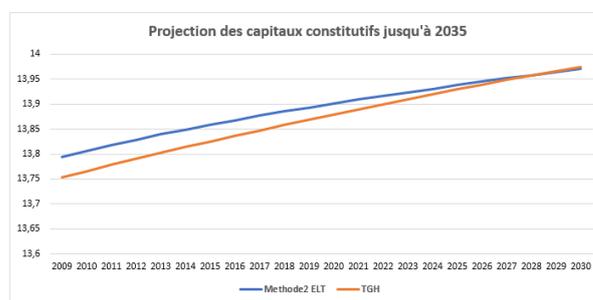


FIGURE 10 – Projection jusqu'à 2030

A l'exception du modèle Lee Carter, les résultats obtenus avec les tables d'expériences sont proches de nos données brutes. La méthode 2 a le double avantage d'être proche des données brutes mais aussi de la table réglementaire, et avec une convergence vers les données brutes au cours du temps. Le coût estimé par la méthode 3 augmente d'une façon très significative au cours du temps (la pente est très forte). Au vu de l'ensemble des résultats, nous conservons les tables d'expériences correspondant au deuxième modèle d'ajustement (Méthode 2 : modèle semi-paramétrique de Brass).

Nous constatons au travers de ces projections, une amélioration constante des taux de décès. Nous remarquons également que les tables réglementaires utilisées dans le cadre du

provisionnement et de l'estimation des engagements resteront prudentes à long terme. Les tables d'expérience ainsi construites permettent le suivi technique de longévité de notre population assurée. Nous remarquons que même si le coût des capitaux obtenu en utilisant la table d'expérience construite présente un écart sensiblement différent de celui obtenu à partir des tables réglementaires, cet écart tend à diminuer fortement à horizon de 30 ans.

L'évolution des taux de décès de notre population est à surveiller, en intégrant au fur et à mesure les nouvelles années d'expérience.

Executive Summary

Context

We live in turbulent times, characterised by political, social, economic and technological changes. These developments, particularly demographic ones, lead to constant adjustments in resources and benefits for both insured people and insurers. Some trends must be carefully studied, such as the ageing of the population, which has a significant impact on the estimation of insurers' commitments.

This thesis is part of this framework and carries out a complete study for the longevity risk management of a life annuities portfolio of group retirement contracts. Calculation of the insurer's commitments in the context of life annuities is based on regulatory tables, TGH TGF 05, derived from national population data and insurers' data. They propose an evaluation of the commitments in a prudent manner for an insurance organisation but which doesn't correspond in all respects to their insureds' portfolios. To mitigate this technical risk, the regulations propose that insurers use so-called experience-based mortality tables.

This brief is organised in four main parts

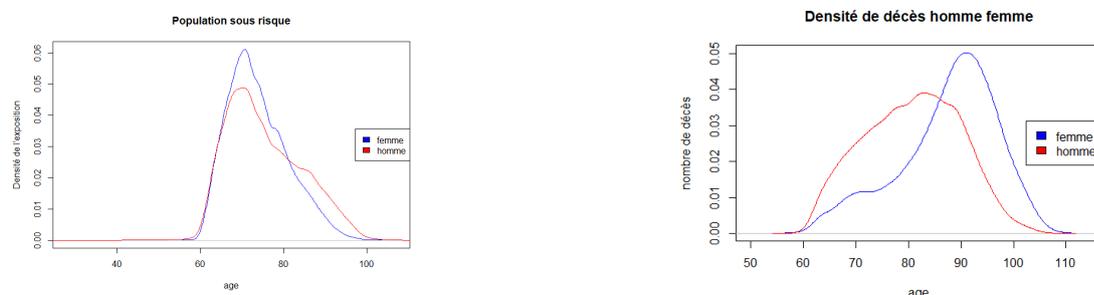
- Presentation and statistical study of trends in our portfolio data ;
- Presentation and indicators study for monitoring the longevity risk of our portfolio ;
- Construction of experiment mortality table ;
- Application and impact analysis on our portfolio population.

Review of our portfolio

The graphs below show the distribution of annuitants exposed to risk between 2009 and 2017 as well as deaths by age. We observe individuals between the ages of 16 and 111 with almost no exposure before the age of 60. The annuitants observed between the ages of 21 and 60 correspond to reversioners or early retirees. In view of their low exposure, we decide to concentrate on annuitants between 60 and 110 years of age for the remainder of our study.

There is an upward trend in the number of deaths in the portfolio, which is correlated with the increase in the number of annuities in the portfolio. However, this analysis must be considered with great caution. Indeed, we observe large fluctuations in deaths, due to the relatively small size of the portfolio considered. Moreover, the number of deaths of men is much higher than women, i.e. 65% of the number of deaths, which is identical with the

number of pensions.



Numbers by year of observation									
	2009	2010	2011	2012	2013	2014	2015	2016	2017
Annuitants	67357	70453	73249	75244	79028	81445	89408	92056	95487
% Male	67%	67%	67%	67%	66%	66%	65%	65%	65%
Deaths	1431	1541	1553	1585	1697	1692	2079	2198	2216

Tracking indicators of longevity risk

To analyse the evolution of our portfolio in terms of longevity, we look at two indicators : the SMR (Standardised Mortality Ratio) and the mortality result. The SMR allows us to compare our observations with estimates, most of the time from regulatory mortality tables. It is the same principle for the mortality result but in amount, thus comparing the mathematical provisions calculated from the regulatory tables with the deaths actually observed.

Mortality result quantifies the gain (or loss) due to the differences between the observed mortality and the theoretical mortality (as defined by the regulatory mortality tables) used to calculate the mathematical provisions. It can be a gain or a loss : A loss if the annuitant is alive at the end of the year : $q_x^{obs} = 0 < q_x^{th}$, a gain if the annuitant dies during the year : $q_x^{obs} = 1 > q_x^{th}$. It is calculated using the following formula.

$$RM = PM_{ouv} * (1 + i) - \frac{R}{k} * (1 + f) * (1 + i)^{\frac{k-1}{k}} * \sum_{j=0}^t v^{\frac{j}{k}} * (1 - \delta_{j,t}) - p^{obs} PM_{clot}$$

Regarding graphs, we notice a very high volatility of the mortality result from year to year. This is justified by two components : the amount of the annuity and the estimated probability of death.

The study on annuities shows that there is a strong correlation between the mortality result and the amount of annuities exiting the portfolio (death). Indeed, we notice that

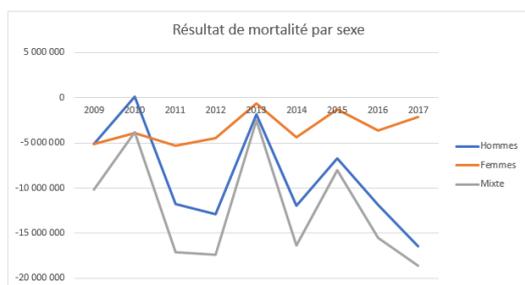


FIGURE 11 – (a)

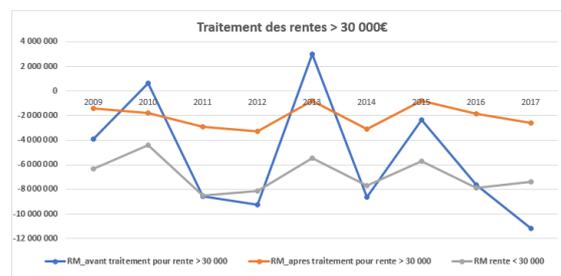


FIGURE 12 – (b)

FIGURE 13 – Mortality result (a), Mortality result after annuity treatment (b)

when we have results close to 0, we observe the death of an annuitant with a high annuity. We therefore decide to take a look at the result when the annuities are below a certain threshold to be defined. Using quantiles' study, we capture annuities at 30000€ and define this amount as the threshold. We recalculate a mortality result per year with the new annuities and obtain the following results. We then notice a decrease in the volatility of the results which remains however important (approximately 18 million initially against 8 million after smoothing of the amounts of annuities for example between 2011 and 2012).

A second component influences the mortality result : the theoretical death probabilities. Indeed, the regulation imposes the use of generational mortality tables for the calculation of our commitments but these don't correspond in all points to our portfolio mortality evolution by their construction. We were able to highlight this observation thanks to the SMR study.

All these results converge towards the same conclusion : our annuitants have a mortality trend that differs from the regulatory mortality tables and this effect is multiplied by the weight of the annuity amounts. This risk being predominant, it is important to be able to anticipate the evolution of this life drift. To do this, we first propose a portfolio segmentation method to determine a mortality trend through different risk profiles.

Improving the mortality result : Implementation of a classification to define risk profiles

From the previous data analysis, we noticed that the overall result had the same trend as the result for the male population, the basic pension amounts of the annuitants and the reversionary presence. This last information is important because when the first head dies (the insured), the provision.

From the previous data analysis, we noticed that the overall result had the same tendency as the result for the male population, the basic pension amounts of the annuitants and the presence of a reversionary. This last information is not negligible because when the first head dies (the insured), the mathematical provision is not fully released but only by a certain amount, corresponding to the reversionary amount. A new provision is calculated for the second life (the reversionary). The technical gain is then less than in the case of no reversion. We will therefore divide our population into several classes according to these different criteria. Then we project the results to estimate the mortality result of the current year.

To do this, we used the K-means method to coarsen our database, backed up by a AHC Agglomeration Hierarchical Clustering for a much more precise segmentation to obtain between 7 and 8 clusters.

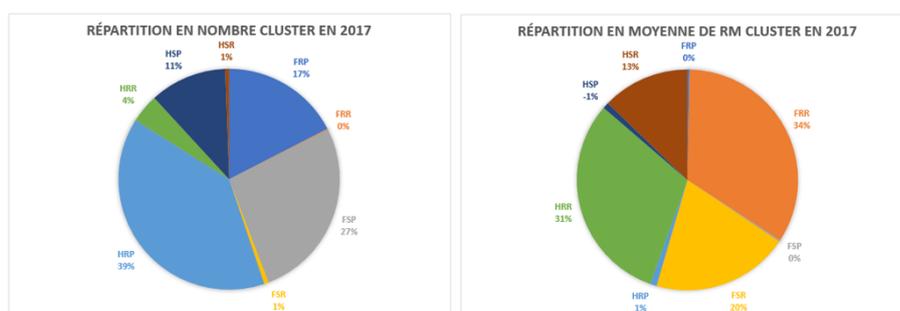


FIGURE 14 – Distribution of clusters for article 39

Each year is marked by eight clusters for Article 39 pensions and seven clusters for Article 83 pensions. The obtained clusters are also explained by the 3 variables SEX, REVERSION and RICH. Thus, we decide to call the seven/eight clusters as follows : FRP, FRR, FSP, FSR, HRP, HRR, HSP, HSR. The first letter corresponds to the representative SEX of the cluster (H for Male and F for Female), the second represents the presence or not of a reversionary (R for Reversionary and S for Single) and the third if the annuitants that make up the cluster have high pensions (R for Rich and P for Poor).

Construction of an experience-based mortality table

The construction of such tables is a delicate procedure and two approaches are available. Firstly, to build a current table which will take into account only one essential variable : age. Secondly, to build a prospective mortality table to take into account future changes in mortality.

Mortality table of moment

In this section, several methods are compared in order to choose (with the help of statistical indicators such as the SMR and the signs test) the one that best explains the data in our portfolio.

For estimating the crude rates, the Kaplan Meier method shows better results than the Hoem method. But both estimators converge to the same object, which shows a satisfactory result. To remedy the fluctuations of the estimated rates two methods were chosen : Whittaker-Henderson and Geometric smoothing. Finally we adjusted the raw rates by two parametric methods Makeham and Tchatcher.

	Whittaker-Henderson	Makeham	Tchatcher	Lissage Géométrique
SMR moyen	102.9%	115.8%	114.8%	102.8%
σ moyen	0.021	0.21	0.03	0.020
Nb dans l'IC	30	14	17	31

TABLE 2 – SMR, σ , Probability death number in CI for each model

At first glance, we see that the average SMR is higher than 1 for each model, so crude rates are underestimated. On the other hand, Whittaker-Henderson and Geometric smoothing present the best SMR and more values in the confidence interval, contrary to the parametric models which are very far from reality. Indeed, the smoothing methods do not assume the a priori law in order to be as close as possible to the data. Therefore, at this stage the Whittaker-Henderson smoothing and the geometric smoothing are retained.

Then, the sign change test validates the H0 hypothesis for both methods, with a 100% accuracy for Whittaker-Henderson. Moreover, an analysis of our portfolio shows that the geometric method is not adequate for our portfolio, as having only 32 ages, an order of 5 implies taking into account a non-negligible number of observations.

In the end, we retained the Whittaker-Henderson smoothing and chose the Denuit and Goderniaux method to extrapolate the mortality rates at older ages because it's simple to implement and has satisfactory results. Consequently, the results of table closure are very satisfactory with a coefficient of determination R^2 of 99.73% with a low p-value.

In order to validate the complete table, we studied the positioning of the table constructed in relation to the regulatory tables. We therefore constructed a reference table by weighting the TGH TGF 05 tables by the sex ratio and by the weight of male and female staffs' percentages. We then assume that there is affine relations between the logits of our table and the reference table : $logit(q_x) = a * logit(q_x^{ref}) + b + \epsilon$.

Finally, we can say that we have a very good match of the experience logits to the reference logits given the high value of the adjusted R, 99.64%.

Prospective based-experience mortality table

The objective of this part is thus to make a prospective analysis of the mortality in order to be able to anticipate a possible drift in time. For it, we will initially implement several methods of prospective adjustment, then we will make a comparative analysis of these various methods in order to retain only that which corresponds statistically best to our real data. Let us recall that we have data from a portfolio of annuitants over a history from 2009 to 2017, i.e. 9 years. Furthermore, following the same approach as in the previous chapter, we arrive at an age range of [63,94] years.

First, we adjusted the raw rates by the Lee Carter method. This gives us the following results.

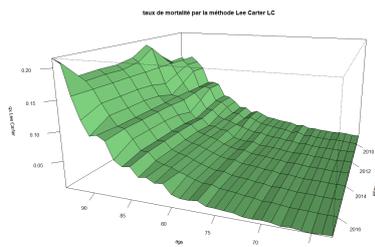


FIGURE 15 – Lee Carter (a)

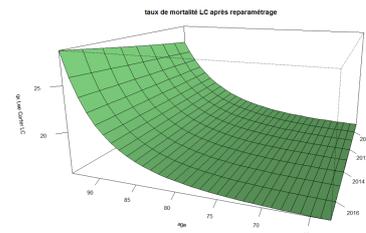


FIGURE 16 – reparamétrage (b)

We can conclude that with the first approach (without reconfiguration of the coefficients), we notice that the smoothing performed by the Lee-Carter method is not optimal. Since we are in a context where the numbers (around 100,000 lines) are relatively small, there are still sampling fluctuations, which lead to irregularities in the estimated parameters and therefore in the estimated rates. We then notice an improvement of the Lee Carter model by reconfiguring model coefficients in the sense that we observe less fluctuations in the estimated rates. Thus, after the analysis of the confidence intervals of the number of deaths, the deaths predicted by the model are close to the observed deaths. This makes the model satisfactory overall.

In view of the our limited portfolio size (about 98313 annuitants for 9 years of observations), it seems difficult to consider only an endogenous construction of experience tables as we started in the previous section. We then constructed an experience table with our portfolio by referring to an external table. To do this, we tested three positioning methods via the ELT package thanks to works of Tomas and Planchet and Wassim YOUSSEF on longevity.

Method 1 : $\tilde{q}_x(t) = SMR \times q_x^{ref}(t)$

Method 2 : $\text{logit}(\tilde{q}_{x^*}(t^*)) = \alpha + \beta \text{logit}(q_{x^*}^{ref}(t^*)) + \epsilon_{x^*,t^*}$

Method 3 : $\mu_{x^*}(t^*) = \beta_0 + \beta_1 \log(q_{x^*}^{ref}(t^*)) + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^*$

In order to find the best method to fit our data, we use the validation criteria proposed in the ELT tool. They are divided into three groups, the first one consists in the analysis of results obtained regarding the proximity criteria (R2, MAPE, SMR, deviance,...), the second one regarding the regularity of the adjustment (we find runs and signs tests) and the third one verifies the coherence of extrapolated mortality evolution trends.

All these criteria lead us to the same conclusion that method 3 and method 2 present better results for our portfolio. However, it is difficult to judge, at first sight, the relevance of one table or another because the results obtained are very close to each other and to our data. The sampling fluctuation makes the interpretation of statistical tests difficult and contradictory. We therefore decide to keep the three methods thus constructed by gender, and we will choose the most appropriate one in view of the SMR of life expectancy and the constituent capitals.

Application and impact analysis on our portfolio

We have now analysed and calibrated different models (Lee Carter model and the three positioning models via the ELT package). It is appropriate to analyse their impact on our portfolio, allowing us to compare these models in order to select the one that best captures population mortality of our annuitant portfolio.

SMR and life expectancy

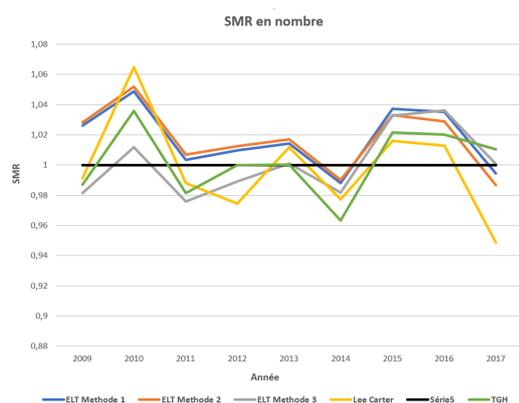


FIGURE 17 – SMR par année

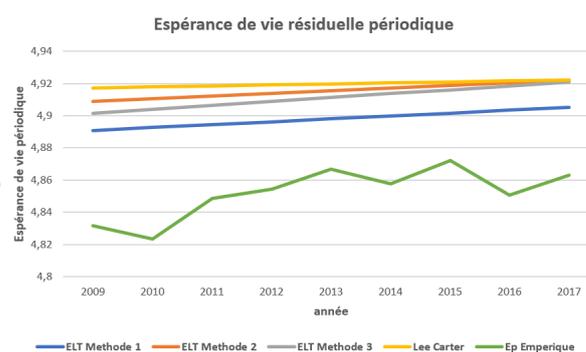


FIGURE 18 – Espérance de vie par année

We note similar movements between experimental and regulatory tables. We find here the translations carried out through the adjustment by tables positioning. Moreover, the positioning model, method 2, presents better results in terms of SRM because it is closer to

the TGH table, while overestimating the probabilities of death. We also note that the implemented methods predict a relatively similar partial generational life expectancy trend, which is at a higher level than the raw data.

In view of the analyses and the back-testing via the confidence interval of the predicted deaths, we retain method 2 (two-parameter approach with a relational model) and method 3 (generalized linear Poisson model) as the best models for our data, they present satisfactory results.

Annuity Capitals

The cost of paying one euro of annual annuity to an annuitant from age x is defined as :

$$C_x = \sum_{k=0}^w v^k {}_k p_x$$

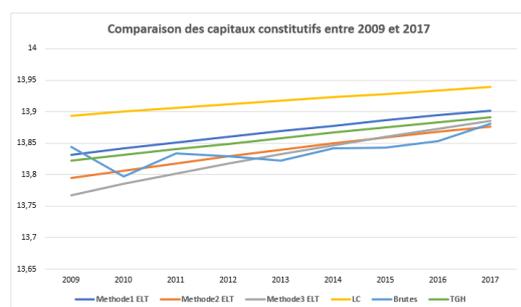


FIGURE 19 – Capitals between 2009 and 2017

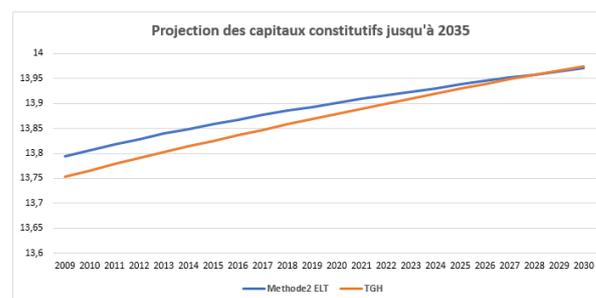


FIGURE 20 – Screening until 2030

Except the Lee Carter model, the results obtained with the experimental tables are close to our raw data. Method 2 has the double advantage of being close to the raw data but also with the regulatory table, and with a convergence towards the raw data over time. Method 3's estimated cost increases very significantly over time (the slope is very steep). In view of all the results, we keep the experience tables corresponding to the second adjustment model (Method 2 : Brass' semi-parametric model).

We note through these projections a steady improvement in death rates. We also note that the regulatory tables used for provisioning and estimating liabilities will remain conservative over the long term. The experience tables thus constructed allow the technical monitoring of our insured population's longevity. We also note that if the cost of capital obtained using the experience table constructed presents a difference significantly different from that obtained from the regulatory tables, this difference tends to decrease significantly over a

30-year horizon.

The evolution of death rates in our population should be monitored, incorporating new years of experience as we go along.



Table des matières

Introduction	5
Remerciement	7
I Contexte de l'étude	8
1 Entreprise et Périmètre d'étude	8
1.1 LA DAAP : Direction Actuariat et Assurance de Personne	8
1.2 La Retraite Collective	9
1.2.1 La législation	9
1.2.2 Les produits de la retraite collective d'Axa France : Pilier 2	10
1.2.3 La rente viagère	11
2 Problématique et Contexte	13
2.1 Risque de longévité	14
2.2 Contexte méthodologique	16
II Etude de notre portefeuille	18
1 Fiabilisation des données et analyse descriptive	18
1.1 Présentation des données	18
1.1.1 La source de données	18
1.1.2 Base pour R	20
1.2 Traitement de données	21
1.3 Analyse descriptive	23
1.3.1 Statistiques des variables pertinentes	23
1.3.2 Analyse des tendances	28



2	Les indicateurs du suivi de risque	31
2.1	Resultat de mortalité	31
2.2	Le SMR	39
2.3	Amélioration du résultat de mortalité : Mise en place d'une classification pour définir des profils de risques	41
2.3.1	Traitement des données	42
2.3.2	Mise en place des clusters	45
2.3.3	Résultats	49
2.3.4	Segmentation supervisée pour estimer le résultat de mortalité d'une année	51
III	Construction d'une table du moment	54
1	Retraite Collective & Modèle de durée	54
1.1	Definitions des notions de bases	54
1.1.1	Fonction de survie	54
1.1.2	Taux de risque instantané	55
1.1.3	Notations actuarielles	56
1.1.4	Phenomene 1 : La Censure	56
1.1.5	Phenomene 2 : La troncature	57
2	Ajustement d'une table du moment	58
2.1	Estimation des taux avec le modèle de Hoem	58
2.1.1	Choix de la tranche d'âge des taux bruts	61
2.2	Estimation des taux avec le modèle de Kaplan Meier	62
2.3	Comparaison des deux estimateurs	64
2.4	Lissage des taux bruts	67
2.4.1	Lissage Whittaker-Henderson	67
2.4.2	Moyenne Géométrique	68
2.5	Ajustements des taux par des modèles paramétriques	70
2.5.1	Ajustement de Makeham	70
2.5.2	Ajustement de Tatcher	72

2.6	Choix du meilleur modèle	73
2.6.1	test de changements des signes	75
2.6.2	Calcul du SMR (Standardized Mortality Ratio)	76
2.7	Fermeture de la table	77
2.8	Validation de la table globale	79
IV	Construction d'une table de mortalité prospective	81
1	Ajustement par Lee Carter	83
1.1	Estimation des paramètres	85
1.2	Résultat du modèle Lee Carter	88
1.3	Recalibrage des coefficients du modèle	89
1.4	Fermeture de la table	90
1.5	Validation du modèle	91
1.5.1	Analyse par espérance de vie générationnelle	91
1.6	Intervalle de confiance	94
2	Ajustement des taux brutes par des méthodes de positionnement	96
2.1	Méthode 1 : Approche à un paramètre avec le SMR	97
2.2	Méthode 2 : Approche à deux paramètres avec un modèle relationnel	98
2.3	Méthode 3 : Modèle linéaire généralisé de Poisson incluant des interactions avec l'âge et l'année calendaire	99
2.4	Analyse des résultats obtenus	100
2.5	Fermeture des tables	111
2.6	Comparaison des méthodes	117
2.6.1	Comparaison du SMR	117
2.6.2	Comparaison des espérances de vie résiduelles périodiques	119
V	Application dans un portefeuille de rentes viagères	120
1	Définition des capitaux constitutifs	120



2 Application	121
Conclusion	123
Bibliographies	125

Introduction

L'allongement de la durée de vie est un sujet d'actualité soulevant de nombreuses questions. En France, le maintien d'un système de retraite montre à quel point ce phénomène est lourd de conséquences sur les équilibres économiques et financiers.

Nous vivons dans une époque mouvementée, caractérisée par des changements politiques, sociaux, économiques et technologiques. Ces évolutions notamment démographiques entraînent des adaptations constantes des ressources et prestations pour les assurés comme pour les assureurs. Certaines tendances sont alors à étudier avec soin comme le vieillissement de la population, qui a un impact non négligeable sur l'estimation des engagements des assureurs.

Dans le cadre de l'assurance vie, la sortie pour l'assuré peut se faire sous forme de capital ou de rente. Dans ce dernier cas, et plus particulièrement dans le cadre de rentes viagères, l'assureur s'engage à verser périodiquement à partir d'une date fixée, un montant déterminé jusqu'au décès de l'assuré. C'est donc un contrat où le risque porte sur la durée de vie humaine. Le risque viager porté par l'assureur est dans ce cas un risque de longévité, c'est-à-dire un risque que la durée de vie des assurés soit supérieure à la durée de vie attendue (estimée par les tables de mortalité). Il s'avère donc nécessaire pour l'assureur, de pouvoir évaluer le plus précisément possible la durée de vie des rentiers et les éventuelles dérivées futures de mortalité afin de maîtriser au mieux ses engagements à chaque instant.

Ce mémoire s'inscrit dans ce cadre et réalise une étude complète pour le pilotage du risque de longévité d'un portefeuille de rentes viagères de contrats de retraite collective. Le calcul des engagements de l'assureur, dans le cadre de rentes viagères est réalisé à partir de tables réglementaires, TGH TGF 05, issues des données de la population nationale et des données d'assureurs. Elles proposent une évaluation des engagements de manière prudente pour un organisme d'assurance mais qui ne correspond pas en tout point à leurs portefeuilles d'assurés. Pour pallier ce risque technique, la réglementation propose aux assureurs la mise en place de tables de mortalité dites d'expériences.

La construction de telles tables est une procédure délicate et deux approches sont disponibles. D'abord, construire une table du moment qui prendra en compte qu'une variable essentielle : l'âge. Ensuite, construire une table de mortalité pour prendre en compte les évolutions futures de la mortalité. Pour cette dernière deux approches sont possibles : une

approche par positionnement à des références de mortalité externes ou une approche dite endogène, permettant d'ajuster aux données brutes un modèle paramétrique, semi paramétrique, non paramétrique ou même stochastique. Le choix de l'une ou l'autre approche n'est pas anodin. C'est pourquoi, une étude approfondie des tendances de mortalité du portefeuille est nécessaire.

Afin de comprendre au mieux le contexte de l'étude, nous présentons dans un premier temps le périmètre de la retraite, les évolutions démographiques observées ces dernières années ainsi que les outils existants permettant l'évaluation de la mortalité d'une population. Nous réaliserons par la suite l'étude complète du portefeuille au travers de deux indicateurs : le SMR (ou Standardised Mortality Ratio) et le résultat de mortalité. Nous détectons par la suite des groupes homogènes de rentiers avec des profils dits "à risque" pour notre étude. Afin d'approfondir notre étude sur les tendances de mortalité de notre portefeuille, nous mettons en place une table du moment et une table de mortalité d'expérience prospective à partir d'approches exogènes au vu du faible échantillon de données disponible.

Remerciement

Je souhaite remercier et témoigner toute ma reconnaissance à mon tuteur d'entreprise Yoni CHOCHA pour le suivi et pour tout son soutien moral et technique.

Je voudrais également remercier mes collègues d'AXA France et toute la DAAP, plus particulièrement l'équipe Retraite Collective, Ran, Ilyes et Camille Renard pour tout leur soutien et leur aide dans la construction de ce mémoire et leurs différentes idées.

Je tiens à remercier l'équipe pédagogique de l'Institut Statistique des Universités de Paris (ISUP) ainsi que les intervenants professionnels pour la formation de qualité qu'ils ont délivrée et qui m'ont été d'une grande aide durant la construction de ce mémoire et plus particulièrement M. Olivier LOPEZ.

Enfin, je remercie chaleureusement mes amis et ma famille, qui m'ont soutenue durant toutes mes années d'études.

Première partie

Contexte de l'étude

"La retraite est la période de la vie d'un travailleur où, après s'être retiré de la vie active, il continue de toucher régulièrement une somme d'argent à titre de pension." *DefWikipedia.*

Pour certaines personnes, c'est la deuxième phase de sa vie, pour l'Etat, c'est un acte de solidarité intergénérationnelle . Elle a de multiples définitions, c'est une partie absolue de la société et de l'économie, mais également une source de risques.

1 Entreprise et Périmètre d'étude

La retraite est un système très complexe, une composante majeure de l'économie de la France, toutefois son architecture peut se diviser en deux parties. D'abord, la retraite de base et la retraite complémentaire, gérées par les Régimes de base de l'Etat, ensuite s'ajoute la retraite entreprise et l'épargne retraite individuelle généralement gérées par des organismes externes comme les assureurs. Ainsi, la retraite devient un secteur très convoité du privé qui développe des activités très diversifiées de la retraite.

1.1 LA DAAP : Direction Actuariat et Assurance de Personne

La Direction Actuariat et Assurance de Personne (DAAP) est une sous direction de la Direction Finance Stratégie et Juridique et elle a deux fonctions : la fonction actuarielle et la partie inventaire. Elle réalise sur une année, plusieurs exercices prévisionnels, permettant d'estimer et de projeter les marges techniques et financières afin d'anticiper, au mieux, le résultat des arrêtés de comptes.

Comme exercices prévisionnels nous réalisons des visées (exercices annuels permettant d'estimer le résultat de fin d'année N) et les arrêtés trimestriels.

Il s'ajoute à cela la fonction actuarielle, nous livrons chaque année, un cahier de charges (le rapport actuariel) qui est un résumé de l'ensemble des études actuarielles sur les sujets de la retraite, de l'épargne et de la prévoyance. Ce cahier est suivi par les autorités de contrôle (ACPR). Ce rapport doit être, alors, le plus rigoureux et le plus pointu possible.

1.2 La Retraite Collective

Le système de la retraite en France a plusieurs traits caractéristiques, on peut citer le caractère législatif et démographique.

1.2.1 La législation

Le système de retraite dépend de la législation car c'est elle qui fixe le cadre réglementaire du système. En France, plusieurs régimes dits de base ont notamment été mis en place. Ils dépendent du métier mais également du secteur de l'entreprise. Ainsi, deux actuaires peuvent ne pas dépendre du même régime de base : un actuair e chez AXA dépend de l'Assurance Retraite du Régime Général tandis qu'un actuair e chez Crédit Agricole dépend de la MSA (Mutuelle des Salariés Agricoles).

Pour visualiser le système de retraite, nous pouvons l'assimiler à un édifice à 3 piliers : Le premier représente la base, un socle commun à tous les retraités. Cette base peut être complétée des retraites publiques de base (dites complémentaires) qui sont obligatoires et collectives. Ce pilier est géré par répartition : en tant que salarié, nous cotisons obligatoirement, avec notre employeur pour constituer des droits à la retraite. Ces cotisations sont directement utilisées pour payer les retraites actuellement versées. Ce système de répartition est un mécanisme de solidarité. Le second pilier se compose des retraites professionnelles (collectives) à adhésion obligatoire ou facultative. Le troisième et dernier pilier est défini par les retraites individuelles, facultatives. Ces deux derniers piliers possèdent une gestion par capitalisation, dépendant de la capacité d'épargne de chacun.

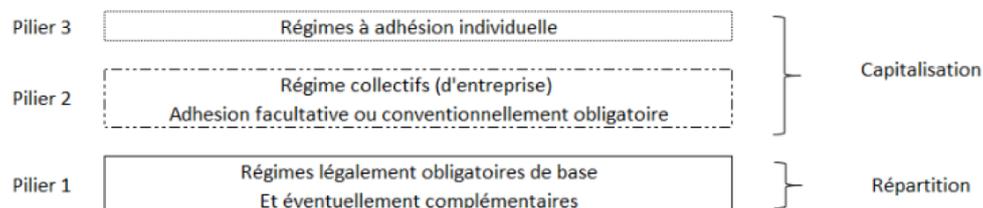


FIGURE 21 – Représentation de l'organisation de la retraite en France

Prenons un exemple : soit un salarié cadre du privé. Le premier pilier (Régimes légaux obligatoires) est composé : de son régime de base qui est le régime général de Sécurité Sociale et des régimes complémentaires qui sont l'AGIRC et ARRCO. Le second pilier (les régimes supplémentaires) est composé des régimes d'entreprise à cotisations ou prestations définies. Enfin, le troisième pilier (la capitalisation

individuelle) comporte les assurances individuelles du salarié.

Les entreprises ont donc un véritable rôle à jouer pour aider leurs salariés dans la préparation de leur retraite. Pour aider les entreprises dans la mise en place de compléments de retraite, les sociétés d'assurance proposent des contrats permettant le versement d'une rente viagère lors de la cessation définitive d'activité professionnelle, à condition d'avoir atteint l'âge minimum de départ à la retraite.

Dans le cadre de ce projet, nous étudions les rentiers issus d'un régime supplémentaire collectif et plus précisément les contrats relatifs aux régimes à prestations et cotisations définies. Les régimes de retraite à cotisations définies (ou encore appelés Article 83) se caractérisent par le fait que l'engagement porte sur le montant de la cotisation et non sur le montant de la pension. A contrario, les régimes de retraite à prestations définies (ou Article 39) se caractérisent, quant à eux, par le fait que l'engagement porte sur le montant de pension, fixé à l'avance.

1.2.2 Les produits de la retraite collective d'Axa France : Pilier 2

Les produits de Retraite Collective sont des contrats d'assurance proposés aux entreprises, afin qu'elles garantissent à leurs salariés le versement d'une rente viagère lors de leur cessation définitive d'activité professionnelle, à condition que ceux-ci aient atteint l'âge minimum de départ à la retraite.

A ce titre, le contrat d'assurance en retraite collective se divise en deux phases :

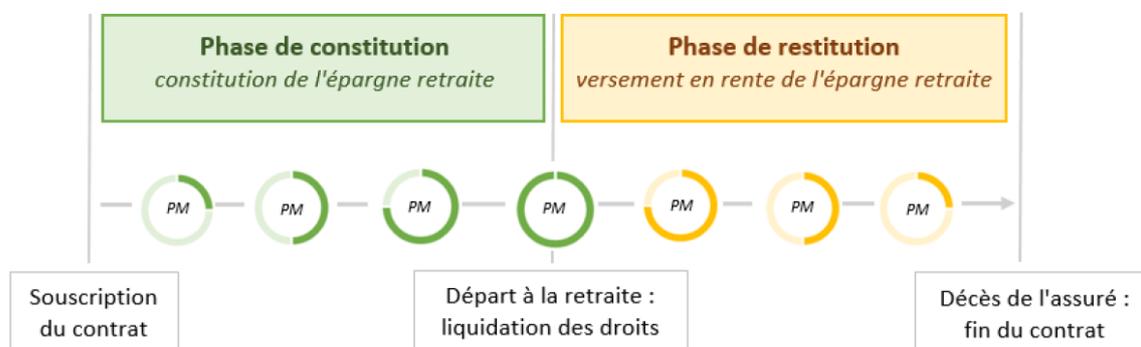


FIGURE 22 – Les différentes phases de la retraite collective

1. La phase d'épargne appelée **phase de constitution** : c'est la phase pendant la-

quelle l'assuré verse ses cotisations pour alimenter un compte propre. La phase de constitution se termine au moment de la liquidation des droits à la retraite.

2. La phase de rente appelée **phase de restitution** : c'est la phase pendant laquelle une rente est versée à l'assuré par l'assureur jusqu'au décès de l'assuré. S'il y a un reversataire prévu par les termes du contrat, la rente lui est alors versée jusqu'à son décès.

D'autre par, il existe plusieurs produits proposés par les entreprises d'assurance, les tableaux suivants résument ces différents produits.

	Contrat dit Art 82/83 : à cotisations définies <i>seul le taux de cotisation est fixé au contrat, non le montant de la prestation</i> <i>Obligatoire, ce contrat est souscrit par l'entreprise au profit du personnel</i>	Contrat dit Art 39 : à prestations définies <i>pourcentage du salaire annuel ou liées à l'ancienneté du salarié dans l'entreprise</i> <i>Obligatoire, ce contrat est souscrit par l'entreprise au profit du personnel</i>
Pendant la phase de Constitution	Constitution progressive pour chaque salarié d'une épargne sur un compte individualisé Droits acquis définitivement, même en cas de changement d'entreprise en cours de carrière Les versements sont obligatoires et correspondent aux cotisations patronales et salariales. Ils peuvent être complétés de versements individuels facultatifs Particularité pour les contrats dits RVD : la cotisation est immédiatement transcrite en pied de rente	Il existe deux types de régimes : - Additif, le niveau de prestations prédéterminé est garanti et vient s'ajouter à celui des autres régimes de retraite - Différentiel (ou chapeau), le niveau global de retraite est garanti, déduction faites des autres régimes de retraite obligatoire Le financement est uniquement réalisé par l'employeur et les droits ne sont pas acquis définitivement
Pendant la phase de Restitution	L'épargne constituée est distribuée sous forme de rente viagère au salarié, puis le bénéficiaire de la réversion	Au moment du départ en retraite, l'organisme assureur débite du fonds collectif les montants constitutifs du versement des rentes

FIGURE 23 – Les produits de la retraite collective 1

	Contrat IFC : Indemnité de Fin de Carrière <i>indemnité versée par l'employeur à un salarié au départ à la retraite</i> <i>Obligatoire, ce contrat est souscrit par l'entreprise au profit du personnel et à prestations définies</i>	Contrat dit L441 <i>contrat en points à cotisations définies</i> <i>Facultatif (exemple : Prefon, retraite complémentaire des fonctionnaires)</i>
Pendant la phase de Constitution	Afin d'anticiper, l'entreprise peut externaliser les indemnités de fin de carrière en souscrivant un contrat d'assurance dit IFC Les sommes versées sur le contrat alimentent un fond collectif	En contrepartie des versements, les cotisants acquièrent des points de retraite L'assureur constitue une provision technique spéciale (PTS) où la totalité des produits et charges financières générés y sont affectés
Pendant la phase de Restitution	Pour bénéficier des prestations, le salarié doit être présent dans l'entreprise au moment de son départ à la retraite et remplir les conditions requises pour le versement des IFC Ces conditions requises sont définies contractuellement ou par les conventions collectives et dépendent de son ancienneté En cas de départ à la retraite, l'organisme assureur débite du fond collectif les sommes dues et les verse à l'employeur qui les reverse ensuite au salarié concerné	A l'âge de la retraite, les assurés perçoivent une rente viagère dont le montant est calculé en fonction du nombre de point acquis et de la valeur de service du point

FIGURE 24 – Les produits de la retraite collective 2

Remarque

C'est important de mettre en évidence cette différence de phases car pour nos données, nous allons uniquement étudier les tables de mortalité des rentiers. Donc, nous nous concentrerons sur la phase de restitution où les comptes des assurés sont déjà liquidés (départ à la retraite) et ainsi observerons les décès qui sont survenus durant cette phase.

1.2.3 La rente viagère

Pour garantir à tout moment le règlement intégral de ses engagements, l'assureur doit constituer une certaine réserve, appelée provision mathématique *PM*. Elle correspond à

la valeur actuelle probable des prestations futures (des rentes immédiates et des rentes différées), autrement dit, notre dette vis à vis de l'assuré, compte tenu d'un taux d'actualisation et d'une table de mortalité fixés par le code des assurances. Elle est calculée par la formule suivante

$$PM = Rente * Annuité$$

L'*annuité* correspond à la valeur actuelle probable du versement de 1€ de pension annuelle d'un assuré, du premier au dernier versement. Elle est calculée à partir des taux de décès estimés dans la table de mortalité et du taux d'actualisation. Nous allons donc, établir une table de mortalité pour enfin calculer la rente (voir Application).

La *Rente* c'est la *PM* de base, extraite depuis le système de gestion Axa, qui est un montant fixe, calculée en fonction du salaire de l'assuré.

La rente viagère est un revenu périodique contre un transfert de capital à l'assureur. Les contrats de rentes viagères constituent pour l'assureur un engagement sur toute la durée de vie de l'assuré. Il est donc nécessaire d'évaluer la durée de vie moyenne probable des assurés pour calculer plus précisément les provisions mathématiques.

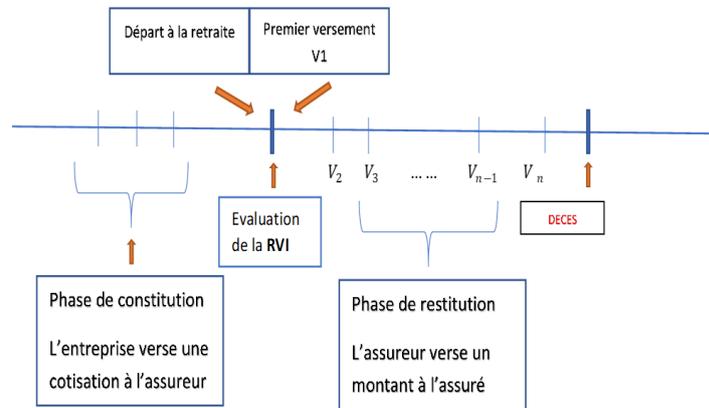
On distingue les rentes viagères à effet immédiat et les rentes viagères à effet différé selon l'âge et le début de la rente.

— RVI : La Rente Viagère Immédiate

Le versement de la rente commence juste après la date de liquidation. Et donc l'évaluation commence au moment du premier versement.

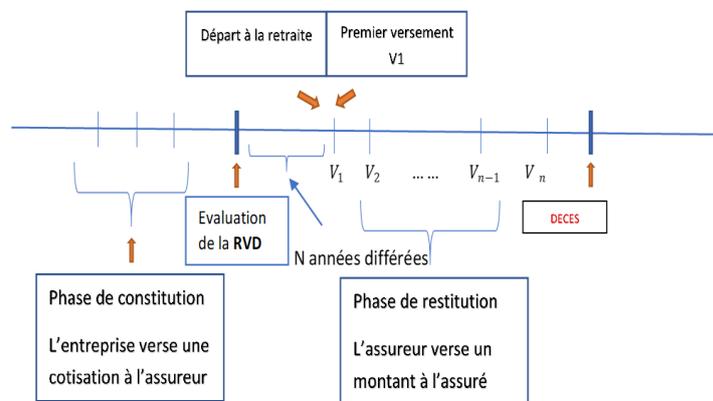
— RVD : La Rente Viagère Différée

Le versement de la rente commence à une date future après la date de liquidation. La RVD se calcule en actualisant la RVI en N années.



Classification : AXA FR Confidentiel

FIGURE 25 – Rente Viagère Immédiate



Classification : AXA FR Confidentiel

FIGURE 26 – Rente Viagère Différée

2 Problématique et Contexte

Dans la partie suivante, nous allons analyser les données de notre portefeuille puis définir un modèle de mortalité. Il s'avère donc nécessaire avant de mettre en place un modèle, de définir les problématiques de notre étude afin de mieux cerner les raisons pour lesquelles il est nécessaire d'établir une table de mortalité d'expérience.

2.1 Risque de longévité

L'espérance de vie à la naissance ne cesse de progresser depuis le début du 20^{ème} siècle, elle a atteint 79,4 ans pour les hommes et 85,3 ans pour les femmes en 2018 en France métropolitaine, selon l'Insee¹. Cette progression, est principalement due aux progrès de la médecine et à l'amélioration de la condition de vie des Hommes avec une tendance plus évidente au cours 20^{ème} siècle. En effet, Les statistiques ont montré que, depuis les années 1960, l'espérance de vie a progressé en moyenne de dix ans.²

Cependant, il faut noter que le progrès scientifique n'est pas le seul facteur explicatif de cette hausse. L'amélioration des conditions socio-économiques et la hausse des dépenses de santé sont, notamment, certains des principaux moteurs de l'allongement de l'espérance de vie observé au cours des dernières décennies. Ainsi, une hausse de 10% des dépenses de santé par habitant (en valeur réelle) est associée à un allongement de 3.5 mois.³

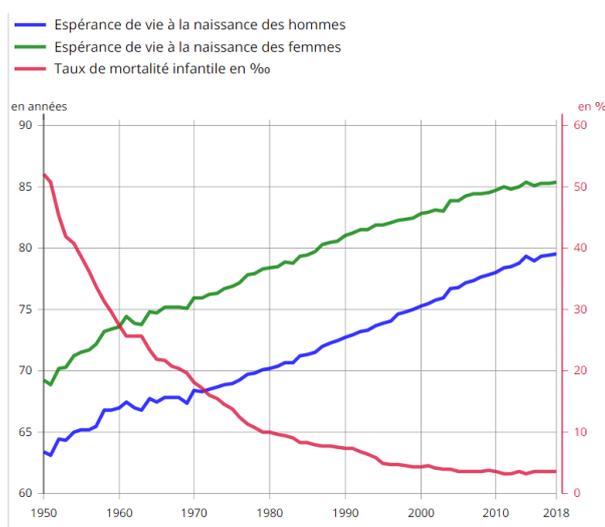


FIGURE 27 – Espérance de vie à la naissance

L'allongement de la durée de vie influence de façon non négligeable toutes les activités liées à la sécurisation des revenus des personnes âgées. L'un des secteurs le plus impacté est celui des retraites et des rentes viagères. En effet, une baisse des taux de mortalité implique que les prestations doivent être payées plus longtemps que la durée prévue initialement, augmentant ainsi l'engagement de l'assureur.

1. Insee : Institut national de la statistique et des études économiques Français

2. Pour avoir plus d'information, vous pouvez consulter : <https://www.insee.fr/fr/statistiques/4277640?sommaire=4318291>

3. Source : Rapport OCDE : panaroma de la santé 2017

De plus, certaines décisions politiques prises par les gouvernements viennent souvent changer la manière de concevoir la longévité. Par exemple le gouvernement Français met en place une nouvelle réforme (appelée la réforme des retraites 2019) qui compte instaurer un nouveau système universel où la retraite serait calculée de la même manière quel que soit le régime. La réforme vise également à ce que l'âge moyen de départ à la retraite soit reculé.

Outre cela, récemment, la loi n° 2019-486 du 22 mai 2019, dite loi PACTE, prévoyait la transposition (dans un délai de 6 mois) de la directive européenne du 16 avril 2014 dite directive portabilité retraite. L'ordonnance a été publiée au Journal officiel du 4 juillet. Elle met fin aux régimes de retraite à prestations définies dits "à droits aléatoires" conditionnant les droits des bénéficiaires à leur présence en entreprise et à la date de leur départ à la retraite. En effet, l'idée étant d'accroître la mobilité des travailleurs entre les Etats-membres, la directive 2014/150/UE du Parlement européen et du Conseil du 16 avril 2014 a réduit les obstacles à cette mobilité, créés par les règles d'acquisition et de préservation des droits à la pension de retraite complémentaire (dont les régimes de retraite supplémentaires à prestations définies aléatoires font parties). Par ailleurs, la même disposition est arrivée récemment aux Etats-Unis, où la plupart de ces contrats sont transformés en régimes à cotisations définies. Ces dispositifs impliquent un transfert du risque de marché et de l'assureur vers l'assuré.

Tout de même, le risque de longévité est difficile à anticiper sur le long terme. Les tables de mortalité prospectives, permettant de projeter la tendance de l'amélioration de la durée de vie, sont utilisées afin de mieux gérer le risque de longévité. Elles s'avèrent très utiles pour le calcul des provisions mathématiques, notamment dans le secteur de l'assurance vie. Il est important de mettre à jour les tables, afin d'éviter le sous provisionnement des engagements, comme ce fut le cas en France en 2006. En effet, l'amélioration de la longévité fut plus importante que celle prévue par les anciennes tables de 1993. Les assureurs ont donc été contraints d'augmenter les réserves de 8% en moyenne, pour prendre en compte ce phénomène.⁴

Afin de mieux maîtriser le risque de longévité et au vu des différents phénomènes de longévité cités ci-dessus, il est important que les assureurs définissent, contrôlent et suivent ce risque lié à la durée de vie humaine pour absorber toutes les spécificités et la nature de leur propre portefeuille d'assurés. En effet, les problèmes liés à la durée de vie n'ont pas les mêmes impacts si nous considérons différents portefeuilles des compagnies d'assurance. Autrement dit, les taux de mortalité et la vitesse de l'amélioration du niveau de vie ne sont pas homogènes entre des groupes d'assurés différents. Il est ainsi difficile pour une compagnie d'assurance de s'appuyer sur les indicateurs nationaux de mortalité, ou même

4. Pour plus d'information : Vous pouvez consulter *ACPR PDF*

ceux du secteur, dans sa gestion du risque de longévité.

2.2 Contexte méthodologique

Dans le cadre du pilotage de son portefeuille de rentiers, il est essentiel pour l'assureur de maîtriser le risque de longévité lié à ce portefeuille. Le but ultime de notre étude est de déterminer une donnée en entrée essentielle au pilotage des rentes de l'assureur. Ceci suppose de capter rigoureusement la mortalité du portefeuille concerné et ensuite de se servir de cette mortalité pour la tarification des rentes. Cette table pourrait aussi servir à la projection des taux futurs de prestations à verser par l'assureur. Notre analyse sera consacrée au traitement des données et à la modélisation de la mortalité qui seront essentiels pour une bonne projection des futures rentes.

En outre, une table de mortalité d'expérience est l'outil par excellence permettant d'expliquer la mortalité d'une population donnée. Pour parvenir à une maîtrise fine du risque de longévité, il est très important d'avoir l'information la plus pertinente et la plus rigoureuse en ce qui concerne la mortalité de la population. Ceci a pour incidence directe sur notre travail de donner une importance particulière à la recherche et à la mise en oeuvre de différents modèles de mortalité, afin de ne retenir que le modèle qui explique le mieux la mortalité du portefeuille considéré.

Pour ce faire, plusieurs modèles ont été développés depuis le début du 20^{ème} siècle. Il s'agit des modèles non-paramétriques (Kaplan Mayer, Nelson-Aalen...), des modèles paramétriques (Gompertz, Makeham...) et des modèles semi-paramétriques (modèle de Cox). En se basant sur des critères de validation tels que le SMR et le pouvoir prédictif, le but est de voir quel modèle explique le plus efficacement possible la mortalité de notre portefeuille.

Toutefois un certain nombre d'assureurs se heurtent à des problématiques techniques liées à la taille de leurs portefeuilles et à leur hétérogénéité en termes de garanties (pour un même risque). Par exemple, en termes de risque en cas de décès, un assureur peut disposer de quelques dizaines de milliers d'assurés bénéficiant de contrats de prévoyance décès, de garanties-plancher sur des contrats en unités de comptes et de contrats d'assurance de prêt. Dans de tels contextes, il est délicat de construire des tables de mortalité au moyen de la seule expérience de chaque type de produit.

Enfin, il est à noter que les tables que nous obtiendrons ici ne feront pas l'objet d'une certification et ne constituent en aucun cas les tables Best Estimate de référence pour l'assureur. Il s'agit d'un outil d'aide au pilotage du risque de longévité. Ainsi, avant toute



chose, il est souhaitable, de porter un regard statistique sur les données qui seront utilisées tout au long de cette étude.

Deuxième partie

Etude de notre portefeuille

La maîtrise des données de notre portefeuille et de ses risques sont des étapes indispensables pour la construction d'une table de mortalité d'expérience. Nous allons donc présenter au sein de cette partie les données et les différentes caractéristiques de notre portefeuille ainsi que deux indicateurs utilisés pour le suivi du risque de longévité : le résultat de mortalité et le SMR.

1 Fiabilisation des données et analyse descriptive

La fiabilisation et la maîtrise de nos données sont les premières étapes les plus prenantes et les plus déterminantes car tous les résultats dépendront de la qualité des données. Il s'avère donc nécessaire, avant de mettre en place des modèles d'estimation des taux de mortalité, d'étudier avec attention notre portefeuille. Pour cela, nous allons commencer par présenter les données, ensuite les principales corrections qui ont été faites (en tenant compte des censures et troncatures) et nous finirons par une analyse descriptive des variables importantes qui sont susceptibles d'influencer nos résultats.

1.1 Présentation des données

Dans le cadre d'une analyse sur la mortalité, deux critères sont à prendre en compte : la profondeur de l'historique et le volume d'informations disponibles.

Lors d'une étude actuarielle, un des aspects majeurs de la pertinence des résultats est la qualité des données utilisées pour réaliser l'étude, mais aussi d'étudier le portefeuille et de faire ressortir les spécificités.

1.1.1 La source de données

Les données sont issues du périmètre de rentes collectives, c'est à dire que nous allons travailler dans le cadre d'assurance de groupe. Les assurés sont les salariés d'entreprises appartenant à différentes catégories socioprofessionnelles. Pour construire notre base de mortalité, nous disposons d'une base de données constituée à partir de deux fichiers.

Un premier fichier contenant toutes les informations liées aux contrats et aux données caractéristiques des assurés. C'est une grosse base de données contenant plus d'un million

d'assurés dont chaque enregistrement contient plus de 40 variables. Nous retenons les 21 variables suivantes et nous notons cette table inforente.

- Numéro de rente
- Titre du titulaire
- Nom du titulaire
- Prénom du titulaire
- Catégorie personnel
- Nature de la rente
- Date de liquidation
- Date de création de la rente
- Date de clôture de la rente
- Date de fin théorique pour les rentes temporaires
- Code motif de clôture
- Date de naissance
- Quotité de réversion
- Terme de paiement
- Périodicité du paiement
- Etat du réversataire
- Date de décès
- Date de suspension
- Motif de suspension
- Raison sociale

Le Terme explique si le paiement de la rente est effectué (en avance ou en échu), la Périodicité nous explique comment le paiement est effectué pendant l'année (annuel, mensuel trimestriel ou semestriel) et la date de suspension indique la date où la rente a été suspendue. cette suspension peut avoir des causes autres que le décès, par exemple les rachats.

Le deuxième fichier utilisé nous donne les informations liées aux contrats, aux assurés et aux montants de rente. Par ailleurs, nous avons récupéré les PM du système et calculé les bénéfices de mortalité⁵. Nous pouvons récupérer pour chaque contrat présent en portefeuille les variables telles que :

- la table de mortalité (TGH, TGF ou TPG)
- Arrérage
- frais de gestion
- PM du système

5. le gain (ou la perte) dû aux écarts entre la mortalité observée à postériori et la mortalité théorique (au sens des tables de mortalité) qui a servi aux calculs des provisions mathématiques (PM).

- RM : résultat de mortalité

En résumé, les données concernent un régime de rentes viagères immédiates du périmètre retraite collective d'AXA France, sur une période d'observation $N_t=[2009,2017]$ soit 9 années. Les individus sont âgés de 21 à 105 ans, $N_x=[21,105]$.

C'est en partant de ces fichiers que nous allons construire notre base de données brutes et fiables. Cette dernière sera ensuite utilisée pour l'analyse de la mortalité des rentiers.

Par ailleurs, nous avons préparé une base de données simplifiée et adaptée aux modèles développés avec le logiciel R, contenant parmi les variables citées ci-dessus les plus importantes pour les modèles que nous allons étudier.

1.1.2 Base pour R

Le but ici est de créer une table telle que pour une personne donnée, nous avons son âge, sa date d'entrée et de sortie de l'étude, Une telle table permet de prendre en compte les censures et troncatures et s'adapte facilement à toute modélisation (par période, par cohorte). Les données recueillies doivent présenter la forme suivante :

Champ	Date d'effet	Motif de sortie	Date de sortie	Date de naissance	Sexe
Format	Date	0 ou 1	Date	Date	0 ou 1

Du fait de la nomenclature inhérente à R, les noms de champs doivent répondre à deux critères :

- ne pas contenir d'espace
- ne pas contenir de signe particulier

De plus, R étant sensible à la casse, l'application et l'utilisation des noms doivent être rigoureuses concernant les majuscules et les minuscules. Du coup, nous changeons la forme des champs :

Champ	Datedeeffet	Motifdesortie	Datedesortie	Datedenaissance	Sexe
Format	Date	0 ou 1	Date	Date	0 ou 1

Le champ « Datedeeffet » indique la date à partir de laquelle l'individu concerné entre dans la période d'observation.

Le champ « Datedesortie » renseigne la date de sortie de risque, plus précisément de la période d'observation, notamment en cas de censure ou de décès.

Le champ « Sexe » indique le sexe de l'individu. Le sexe permettra de créer des tables segmentées selon les caractéristiques. Ce champ vaut « 0 » pour un homme et « 1 » pour une femme.

1.2 Traitement de données

La pertinence et la qualité des données vont conditionner la qualité de la table d'expérience, c'est l'étape la plus prenante et la plus déterminante pour la construction de la table. Il est donc indispensable de s'assurer que toutes les précautions ont été prises en compte. De ce point de vue, nous avons fait des mises à jour et certaines incohérences ont été supprimées. Ainsi, nous détaillons ici toutes les corrections qui ont été apportées pour obtenir une base solide.

Par ailleurs, la norme réglementaire Solvabilité II, a bien pris conscience de l'importance de cette problématique et impose des exigences accrues en matière de gouvernance de la qualité des données. Elle stipule dans son article 48 que la fonction actuarielle est tenue "d'apprécier la suffisance et la qualité des données utilisées dans la construction d'une table de mortalité d'expérience". Les données fiables sont donc un élément essentiel. Il est possible d'affirmer de façon plus spécifique qu'une modélisation n'a de sens qu'à partir du moment où l'on est capable de retracer de façon rigoureuse et fiable les données utilisées.

D'abord il faut savoir que l'extraction est faite par année d'observation de 2009 à 2017, ensuite nous corrigeons les données et enfin nous avons agrégé les données.

Mise en forme des différentes dates utilisées

Nous avons vérifié la cohérence des différentes dates utilisées. Nous avons environ six variables de date qui sont : la date de naissance, la date de liquidation, la date de clôture, la date de suspension, la date de décès (s'il a lieu) et la date de création de la rente. Pour toutes les dates on fait un test pour vérifier la cohérence :

On vérifie que :

- les jours sont compris entre 1 à 31 pour les mois 1, 3, 5, 7, 8, 10, 12 et entre 1 à 30 pour les mois 4, 6, 9, 11.
- et entre 1 à 29 pour le mois de février et pour une année bissextile.

Traitement des incohérences

Outre mettre à jour les dates, d'autres corrections ont été faites.

- On exclut toutes les clauses ou rentiers n'intervenant pas dans l'exercice afin d'éviter les doublons lors de l'agrégation des données.
- on supprime également les rentiers nés après l'année 2005 car c'est l'année de fin de la table de mortalité.

- On supprime les rentes spécifiques (les rentes individuelles, CARCO, UNAPEI et les rentes des moniteurs de Ski. Puisque nous sommes dans un portefeuille de rentes collectives, il est prudent d'éliminer toutes les rentes qui peuvent biaiser le calcul du provisionnement. En effet, ces rentes sont souvent chères de plus on tient compte des caractéristiques dans le cas individuel que nous ne pouvons pas avoir dans le cas collectif.
- On supprime les rachats, car dans le cas d'une étude de mortalité c'est hors catégorie (censure et troncature).
- Les erreurs de saisie, car la création d'une nouvelle rente n'est pas systématique et parfois celle-ci ne réapparaît que des années après.
- Les décès lorsqu'il y a suspension pour autre cause que décès.
- La date de clôture lorsqu'il y a suspension pour toute cause autre que le décès mais que la rente n'est pas clôturée.
- On supprime également les rentes temporelles.
- On met à jour les décès après la fin de l'année pour éliminer ces phénomènes de censures et de troncatures (définies dans la partie modélisation)

Parmi les corrections qui ont été faites, on a relevé :

- 5,87% des rentiers supprimés dont la date d'effet est postérieure à la date de clôture.
- 0,76% des rentiers suspendus avant l'exercice.
- le nombre de décès avant la date d'effet est négligeable soit 0,03% pareil que le nombre de décès avant l'exercice soit 0,07%.
- 7,69% de rentes spécifiques(Individuelles, SKI...).
- 0,42% de nouveaux réversataires pendant l'exercice ont été supprimés.
- 5,58% de rentes temporaires.

Ainsi, au total nous avons éliminé 7,8% de nos données dans cette première partie de nettoyage. Nous verrons plus tard qu'en agrégeant les données et en éliminant les doublons nous allons nous retrouver avec environ 100000 lignes.

De plus, les âges sont compris entre 21 et 105 ans pour des rentiers alors que l'âge moyen de départ à la retraite observé dans notre portefeuille est de 65 ans. En effet, l'assuré à l'option de désigner un ou plusieurs bénéficiaires qui continuent à toucher une rente après la mort de celui-ci. Ces rentes peuvent être des rentes dites d'éducation, de conjointe etc. Nous avons donc supprimé les rentiers âgés de moins de 55 ans car ces rentiers ont une faible exposition et représentent une minorité dans la base de données.

1.3 Analyse descriptive

Dans la première partie, nous avons défini la problématique du sujet, à savoir la conception d'une table de mortalité d'expérience, ensuite nous avons commencé cette partie en présentant les données et les traitements qui ont été faites visant à fiabiliser notre base de données.

1.3.1 Statistiques des variables pertinentes

Dans la suite, pour mieux comprendre notre portefeuille et connaître le modèle adéquat pour expliquer la mortalité de notre périmètre, nous allons tout d'abord décrire les principales caractéristiques dans notre base de données susceptible de contribuer à la construction d'une table de mortalité d'expérience.

Analyse des âges de la population

La population étudiée est constituée de personnes âgées, avec une concentration de l'effectif entre 55 et 100 ans. Ceci s'explique par le fait que nous étudions un portefeuille de rentiers.

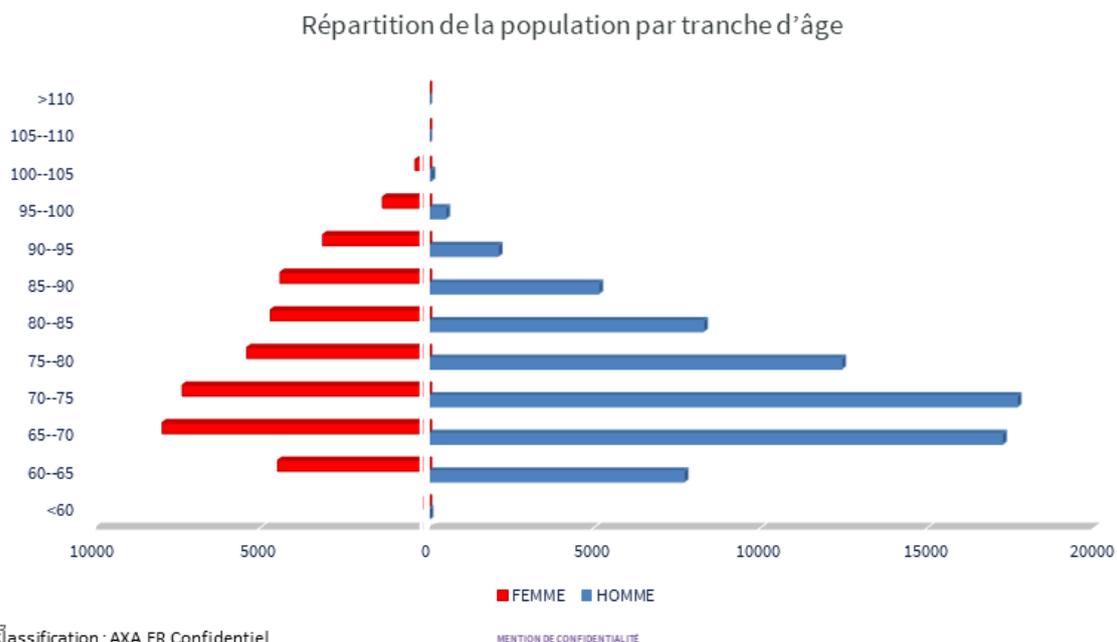


FIGURE 28 – Répartition de la population par tranches d'âge

Les rentiers restent très faibles pour les âges inférieurs à 60 ans (Figure 9), avec une majorité entre 60 et 95 ans. Au-delà de cet âge, l'effectif est très faible. Nous avons une population majoritairement âgée ce qui est normale car nous étudions ici des données de rentes.

Effectif par année d'observation									
	2009	2010	2011	2012	2013	2014	2015	2016	2017
Hommes	45 208	47 281	49 097	50 406	52 455	53 953	57 896	59 508	61 894
Femmes	22 149	23 170	24 149	24 834	26 570	27 487	31 504	32 545	33 649
Total	67 357	70 451	73 246	75 240	79 025	81 440	89 400	92 053	95 543
% de Homme	9,5%	9,9%	10,3%	10,6%	11,0%	11,3%	12,1%	12,5%	13,0%
% de Femme	9,0%	9,4%	9,8%	10,1%	10,8%	11,2%	12,8%	13,2%	13,7%

TABLE 3 – Répartition des assurés en fonction du sexe et de l'année d'observation

De plus, la proportion d'hommes est clairement supérieure à la proportion de femmes au sein du portefeuille avec un sex-ratio de 66%.

Cependant cette tendance se renverse lorsque l'analyse est portée sur les âges les plus élevés c'est à dire à partir de 90 ans. Cette tendance est cohérente avec la mortalité générale de la population française qui relate une longévité plus élevée chez les femmes que chez les hommes.

Age moyen des assurés

Année	2009	2010	2011	2012	2013	2014	2015	2016	2017	Moy/sexe
Hommes	70,81	71,06	71,37	71,79	72,14	72,51	73,14	73,51	73,87	72,35
Femmes	74,06	74,04	74,09	74,29	74,40	75,58	75,38	75,63	75,79	74,79
Moy/année	71,88	72,04	72,27	72,62	72,90	73,20	73,93	74,26	74,54	73,07

TABLE 4 – Age moyen des assurés par année

Le tableau ci-dessus donne l'âge moyen des assurés de l'ensemble du portefeuille par année d'observation et par sexe. Les femmes sont en moyennes plus âgées que les hommes, avec une augmentation des hommes et des femmes au cours des années d'observation.

Nous observons une évolution de l'âge moyen en portefeuille globale (passage de 71,88 à 74,54) suite à une augmentation croissante du nombre d'observations par année et de

l'allongement de la durée de vie humaine. Cette croissance est d'autant plus marquée chez les hommes (+3ans) que chez les femmes (+1,5 ans).

Age de liquidation

Nous analysons l'âge de départ à la retraite dans cette partie.

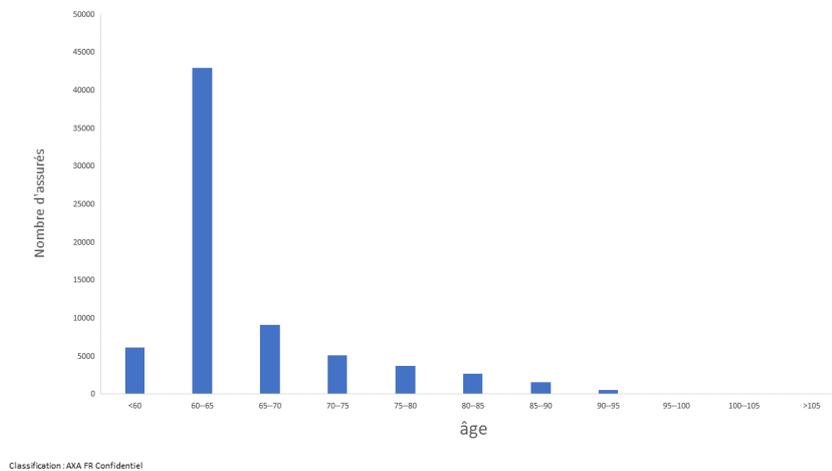


FIGURE 29 – age de liquidation

Les âges de liquidations des rentes nous semblent cohérents. En effet, ils se situent en majorité entre 60 et 70 ans, ce qui est conforme à la réglementation en vigueur en ce qui concerne l'âge de départ à la retraite. De plus, on constate qu'une partie non négligeable des rentiers partent à la retraite au-delà de 70 ans, ceci n'est lié à aucune incohérence de nos données c'est juste que la plupart des contrats présentent des clauses permettant la prorogation de maximum 15 ans afin que les assurés qui veulent continuer à exercer leurs activités professionnelles puissent le faire.

Age de décès

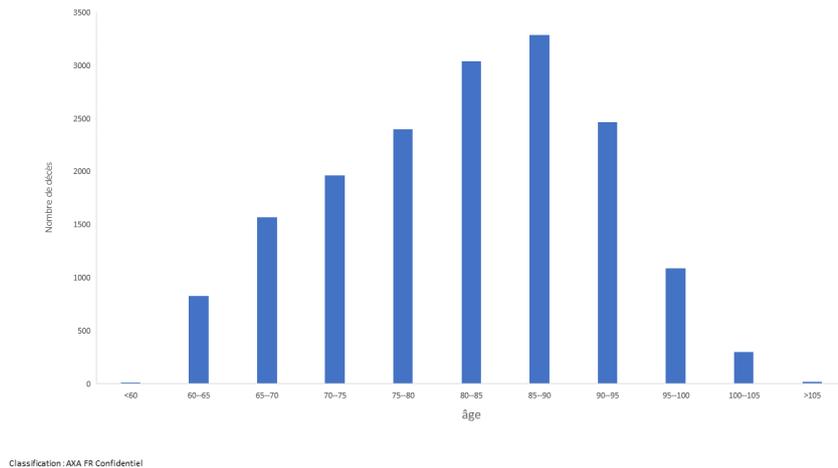


FIGURE 30 – Age de décès

L'âge moyen pondéré de décès est de 84,63 ans. La répartition des âges de décès nous semble correcte puisque nous observons que la majorité des décès survient entre 80 et 95 ans. Autrement dit, la population de notre portefeuille vit longtemps, ceci est en cohérence avec les données nationales car la population française a une espérance de vie au-delà de 80 ans. Ce qui nous permet de valider a priori les dates de décès renseignées .

Nature des rentes

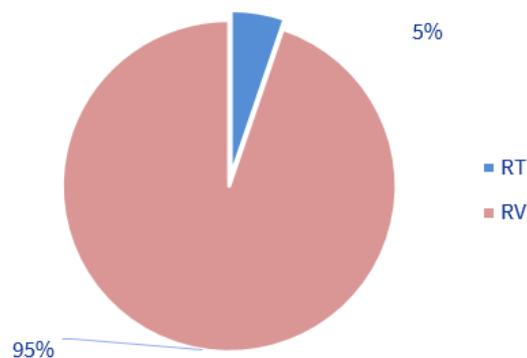


FIGURE 31 – Répartition de la population par la nature des rentes

Les rentes viagères (RV) représentent 95% des rentes du portefeuille contre 5% pour

les rentes temporaires (RT). Retenir uniquement les rentes viagères pour notre étude ne change pas de façon significative la structure des données en portefeuille.

Catégorie socioprofessionnelle

La variable catégorie socioprofessionnelle a beaucoup de valeurs manquantes. Nous avons omis l'étude sur la catégorie socioprofessionnelle des assurés.

Pour l'instant cette information ne sera pas pertinente dans notre étude .

Conclusion

L'analyse que nous venons de présenter nous montre que nous disposons de suffisamment d'informations sur notre portefeuille pour pouvoir réaliser une modélisation de la mortalité. En effet, pour pouvoir faire une étude sur la mortalité, certaines variables sont indispensables. On peut citer dans ce cas :

- La date de naissance
- La date de liquidation
- La date de sortie ou encore la date de décès dans le cas de notre étude

Les dates de naissance et de sortie du portefeuille vont en effet permettre de déterminer les âges de vie et décès de chaque assuré, ce qui permettra par la suite la construction d'une table de mortalité en calculant par âge le taux de mortalité qui est le rapport entre le nombre de décédés et le nombre de vivants (approche classique pour le calcul des taux bruts).

En ce qui concerne la date de liquidation, elle est primordiale dans le cadre de notre étude car nous sommes dans le contexte d'un portefeuille de rentiers et cette date marque l'entrée de chacun des assurés en phase de restitution.

D'autres variables telles que le sexe et la CSP peuvent s'avérer intéressantes pour une étude plus fine et segmentée. Nous avons vu qu'en ce qui concerne la CSP nous avons un fort déficit d'informations. Cependant, ce déficit n'aura pas d'impact sur notre étude car nous modélisons la mortalité de la population entière sans distinction de CSP.

Une fois le nettoyage de la base de données effectué, la suppression des doublons et des

lignes aberrantes faites, nous obtenons une table avec 98 883 assurés. Ce qui représente 92% de la table d'assurés initiale.

Nous pouvons donc conclure après le traitement que nos données retenues pour l'étude répondent aux critères de précision (pas d'erreurs, d'omissions, de doublons) et de justesse (représentatives du risque analysé) tels que définis par l'EIOPA1.

1.3.2 Analyse des tendances

Après avoir montré les principales caractéristiques de la base de données, nous allons analyser la population du portefeuille, les décès en fonction des années d'observations dans notre portefeuille.

Population sous risque

La population sous risque c'est à dire l'effectif vivant et présent au sein du portefeuille considéré augmente assez rapidement d'une année de l'autre :

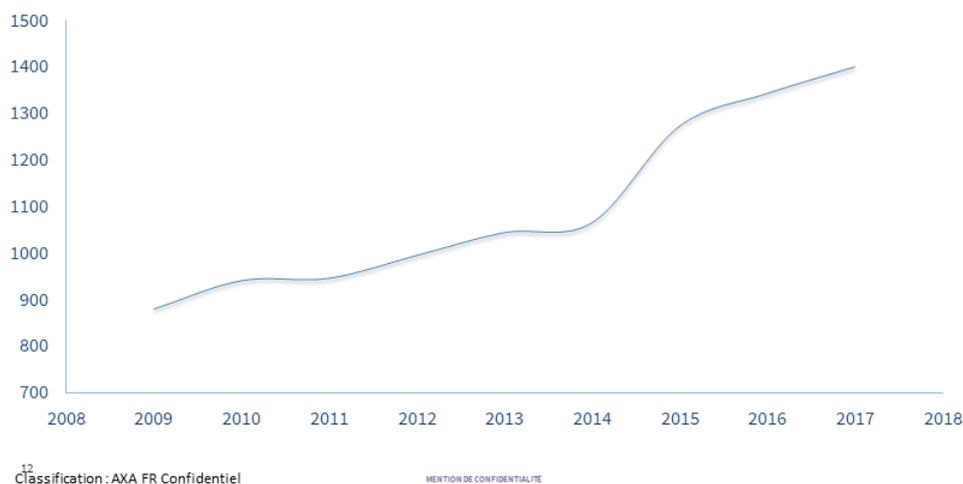


FIGURE 32 – Population vivante

En observant le graphe, nous constatons une augmentation du nombre de personnes vivantes présentes dans notre portefeuille par année d'observation. Cette tendance est cohérente avec la mortalité générale de la population française qui relate une longévité. Cependant, on note une hausse brutale entre l'année 2014 et 2015. On constate la même chose avec les décès, nous y reviendrons.

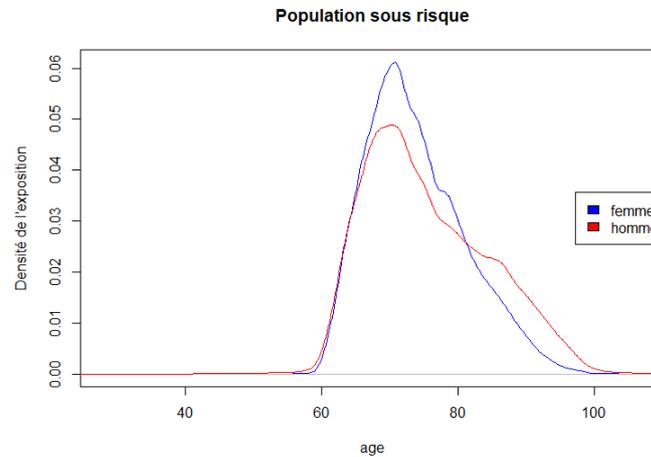


FIGURE 33 – Densité de la population vivante homme/femme

La densité de présence des personnes vivantes entre les hommes et les femmes est quasiment la même jusqu'à environ l'âge de 60 ans.

Nombre de décès par sexe

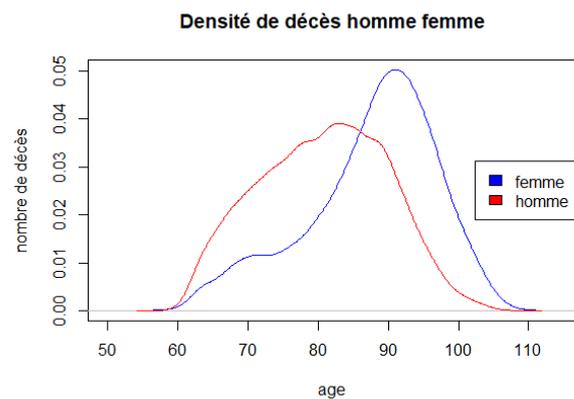
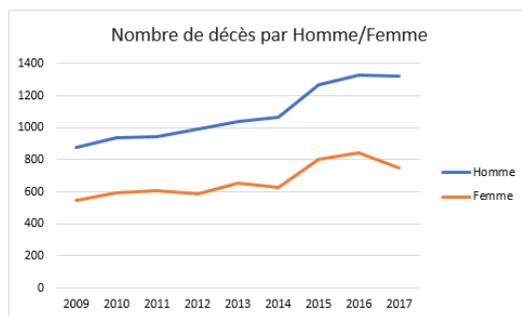


FIGURE 34 – Analyse du nombre de décès homme/femme

On observe une tendance à la hausse des décès dans le portefeuille. Toutefois, cette analyse doit être considérée avec beaucoup de prudence. En effet, nous observons de grandes fluctuations des décès, dues à l'effectif relativement faible du portefeuille considéré. Aussi, détecter une tendance avec un tel effectif peut s'avérer trompeur.

Nous constatons une augmentation très forte entre 2014 et 2015, elle est due à une loi

mise en place en 2014 par le gouvernement qui exige aux assureurs à enquêter sur l'état des rentiers dans l'année. En effet, l'assureur ne disposait pas d'assez de moyens pour connaître, dans les plus brefs délais si un client est décédé. Donc, les assureurs ont inscrit beaucoup de décès durant cette année.

Le nombre de décès est plus élevé chez les hommes avec une différence en moyenne de 24% soit 3 743. D'autre part, la densité pour les femmes aux grands âges est plus élevée, ceci est dû à une faible présence des femmes de cette tranche d'âge.

Article 39 et 83

Les deux principaux produits de notre portefeuille de rentiers Axa France sont l'article 39 qui consiste à garantir à l'assuré le versement d'une rente après le départ à la retraite dont le montant est un pourcentage du dernier salaire . Pour l'article 83, l'entreprise paie pour ses salariés une cotisation à l'assureur qui s'engage à verser une rente lors de la cessation définitive d'activité professionnelle. Ce dernier produit est le plus répandu en France de même que dans notre portefeuille.

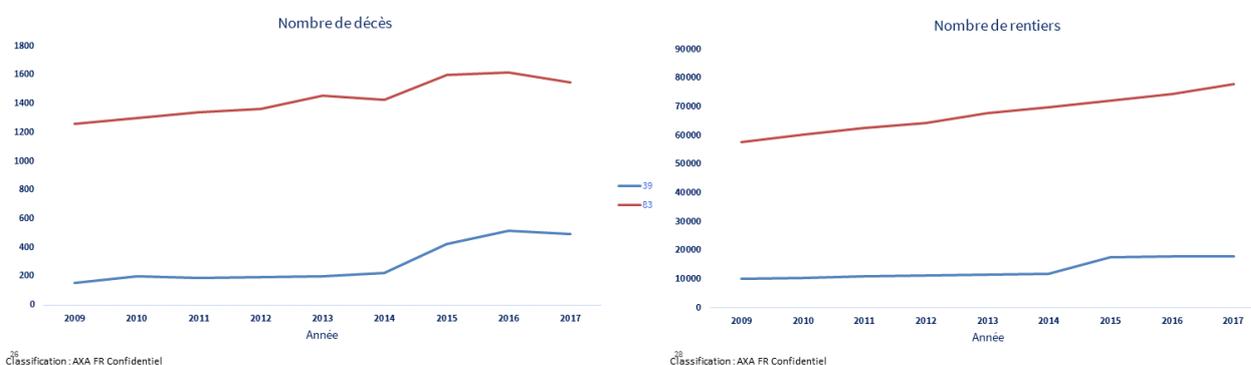


FIGURE 35 – Le nombre de décès en année d'observation des différents produits

On observe (figure 14) une très grande différence entre le nombre de décès pour l'article 39 que 83. l'article 83 constitue la majorité du portefeuille et du coup cette différence est un effet de volume. C'est-à-dire qu'on constate un faible nombre de décès pour le 39 car leur présence est minimale dans notre portefeuille comparé au 83 mais avec la même tendance dans le temps.

Nous constatons également, sur la figure de droite que les articles 83 sont en constante augmentation alors que le nombre d'article 39 reste plat. Ceci est une bonne chose pour notre portefeuille. En effet, d'après les nouvelles réformes de la retraite, notamment la loi

PACTE, qui sont entrées en vigueur, la politique des assureurs est de supprimer au fur et à mesure les articles 39 et favoriser les articles 83.

Dans ce contexte, nous n'allons pas distinguer les deux produits dans la construction des tables de mortalité. Mais, une analyse particulière sera faite sur l'analyse des indicateurs du risque de longévité dans la section suivante .

2 Les indicateurs du suivi de risque

Pour analyser l'évolution de notre portefeuille en terme de longévité, nous regardons deux indicateurs : le SMR (Standardised Mortality Ratio) et le résultat de mortalité. Le SMR permet de comparer nos observations par rapport à des estimations, la plupart du temps issues des tables de mortalité réglementaires. C'est le même principe pour le résultat de mortalité mais en montant, en comparant ainsi les provisions mathématiques calculées à partir des tables réglementaires par rapport aux décès réellement observés.

2.1 Resultat de mortalité

Le résultat de mortalité quantifie le gain (ou la perte) dû aux écarts entre la mortalité observée a posteriori et la mortalité théorique (au sens des tables de mortalité réglementaires) qui a servi aux calculs des provisions mathématiques.

Calcul du résultat de mortalité

La provision mathématique constituée en fin d'année sert à payer les arrérages des années suivantes sous la probabilité théorique d'être en vie au cours de ces années. La provision mathématique (PM) est définie comme la différence entre les valeurs actuelles des engagements respectivement pris par l'assureur et par les assurés. Notre étude portant exclusivement sur les rentes viagères immédiates, la PM correspond uniquement aux engagements de l'assureur (après liquidation de la retraite, la valeur de l'engagement de l'assuré est nulle).

Pour les rentes viagères, la PM d'ouverture, complétée des intérêts techniques de l'année doit financer : les arrérages de l'année, les chargements sur arrérages de l'année et la PM de clôture. Elle intègre donc des probabilités de décès théoriques de l'assuré issues des tables de mortalité réglementaires.

Quelques notations

- BM le bénéfice de mortalité. Il correspond à la perte ou au gain pour un état donné. Le résultat de mortalité est alors une généralisation du bénéfice de mortalité ;
- PM_n la provision mathématique de clôture de l'année n calculée avec la table de mortalité servant au calcul de la PM de clôture $n - 1$;
- i le taux d'intérêt technique annuel ;
- f le taux de frais des arrérages ;
- R le montant annuel d'arrérages ;
- k le fractionnement de la rente ;
- p_x^{obs} la probabilité de survie observée d'un rentier d'âge x déduit de la table de mortalité servant au calcul de la PM de clôture $n - 1$;
- t la période de décès du rentier. $t = k$ s'il n'y pas de décès observé ;
- $\delta_{j,t}$ le symbole de Kronecker. Il vaut 1 si $j = t$; 0 sinon
- $v^{\frac{j}{k}} = \left(\frac{1}{1+i}\right)^{\frac{j}{k}}$.

Nous définissons le bénéfice de mortalité d'un rentier en cas de vie à la fin de l'année n comme la différence entre la PM d'ouverture revalorisée au taux d'intérêt technique i , les arrérages versés pendant l'année revalorisés et la PM de clôture.

$$BM = PM_{ouv} * (1 + i) - \frac{R}{k}(1 + f) \sum_{j=0}^{k-1} (1 + i)^{\frac{j}{k}} - PM_{clot} \quad (1)$$

Si le rentier décède pendant la période $\left[t\frac{12}{k}; (t+1)\frac{12}{k}\right]$ où $0 \leq t \leq k - 1$, le bénéfice de mortalité ne correspond plus qu'à la différence entre la PM d'ouverture revalorisée et les arrérages versés pendant l'année jusqu'au décès :

$$BM = PM_{ouv} (1 + i) - \frac{R}{k}(1 + f) \sum_{j=0}^t (1 + i)^{\frac{k-1-j}{k}} (1 - \delta_{j,t}) \quad (2)$$

Dans le cadre général, nous pouvons donc écrire que :

$$RM = PM_{ouv} * (1 + i) - \frac{R}{k} * (1 + f) * (1 + i)^{\frac{k-1}{k}} * \sum_{j=0}^t v^{\frac{j}{k}} * (1 - \delta_{j,t}) - p^{obs} PM_{clot} \quad (3)$$

Considérons une rentière, âgée de 96 ans et née en 1914, avec une rente de base de 401. Nous appliquons des frais de gestion de 3% et un taux technique de 4,5%. Elle reçoit sa rente trimestriellement et à terme échu. A la fin de l'année, Madame Y est toujours en vie.

Au vu des caractéristiques observées, nous pouvons définir les variables suivantes :

- i le taux d'intérêt technique annuel = 0.045 ;
- f le taux de frais des arrérages = 0.03 ;
- R le montant annuel d'arrérages = 401 ;
- k le fractionnement de la rente = 4 (trimestriel) ;
- $p_x^{obs} = 1$ correspond à la probabilité d'être en vie à la fin de l'année. Dans notre exemple, pas de décès observé à la fin de l'année ;
- $t = 4$, t correspond au nombre de rentes versées pendant l'année. Ainsi, un rentier qui perçoit une rente trimestriellement et décède au mois de juillet, n'aura perçu que 2 versements pendant l'année. Dans notre exemple, Madame Y est toujours en vie à la fin de l'année et a donc perçu sa rente annuelle en 4 fois. Idem pour la variable tr mais associé à la seconde tête ;
- $v = \frac{1}{1+i} = 0,957$

Grâce aux tables de mortalité réglementaires, nous pouvons déterminer les probabilités de décès, de survie et l'annuité associées à la rente. Nous utilisons dans ce cas la table TGF 05 :

- $a_{96}^{1914} = \sum_{k=1}^{120-96} \frac{L_{96+k}}{L_{96}} * (1+i)^{-k} = 2,9085$ correspond à l'annuité pour une rente viagère payable à terme échu d'une personne âgée de 96 ans.
- $-a_{96}^{(per)} = a_{96}^{1914} + \frac{per-1}{2*per} = 3,2836$ correspond à l'annuité pour une rente viagère payable à terme échu avec une périodicité per qui dans notre cas vaut 4, pour une personne âgée de 96 ans.

$$PM_{ouv} = (1+f) * a_{95}^{(per)} * R = 1440,76$$

$$PM_{clo} = (1+f) * a_{96}^{(per)} * R = 1356,37.$$

$$RM = PM_{ouv}(x) * (1+i) - \frac{R}{4} * (1+f) * (1+i)^{\frac{3}{4}} * \sum_{j=0}^4 v^{\frac{j}{4}} - PM_{clo}(x)$$

$$RM = -270,75 = -20\% \text{ de la PM.}$$

Interprétation des résultats

Le résultat de mortalité est soit une perte, soit un gain selon l'écart entre la mortalité observée et la mortalité théorique. Il sera notamment une perte technique lorsque le rentier est en vie à la fin de l'exercice⁶ : sa probabilité de décès constatée est nulle et est donc

6. Un exercice est une année complète par exemple du 01 janvier au 31 décembre 2009

inférieure à sa probabilité de survie théorique. Et inversement, un gain lorsque le rentier est décédé au cours de l'exercice, sa probabilité de décès observée étant égale à 1 et est donc supérieure à sa probabilité de décès théorique.

Voici les résultats obtenus entre 2009 et 2017 :

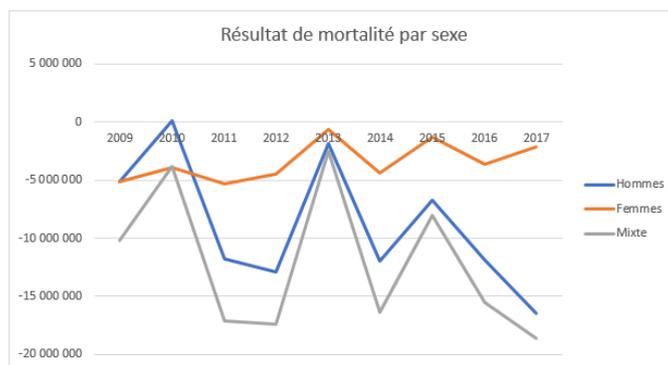


FIGURE 36 – Résultats de mortalité par sexe

Nous remarquons que le résultat de mortalité est très volatile d'année en année contrairement à la proportion de décès observés qui reste stable.

Lorsque nous détaillons le calcul du résultat de mortalité, nous remarquons que celui-ci dépend fortement du montant de la rente de base et de la différence entre les probabilités de décès théoriques et les probabilités de décès réellement observées. Nous décidons d'étudier ces deux composantes plus en détails.

En découpant le résultat de mortalité selon la probabilité de décès observée ($QXOBS = 0$ si le rentier est en vie à la fin de l'année et $QXOBS = 1$ si le rentier est décédé pendant l'année), nous retrouvons la volatilité du résultat de mortalité pour les rentiers décédés (figure 16). Nous notons également une décroissance du résultat de mortalité pour les rentes en cours de service due à l'augmentation de rentes observées au cours du temps.

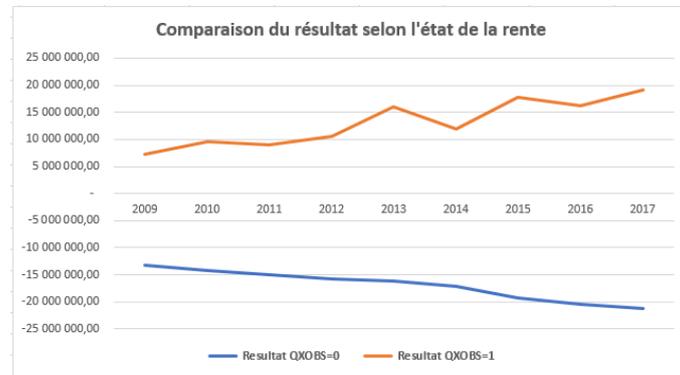


FIGURE 37 – Evolution du résultat de mortalité selon le statut de la rente

Lorsque nous regardons la répartition des montants de rentes en moyenne, selon les probabilités de décès observées (figure 38), nous remarquons une évolution proche entre les deux probabilités de décès observées, avec tout de même un pic en 2013 pour les rentiers décédés. Concernant les rentes maximales observées (figure 39), une nouvelle fois la volatilité est plus forte pour les rentiers décédés (rentes sortie), passant du simple au triple selon les années (en 2012, la rente maximale observée est de 192 188 contre 723 999 en 2013).

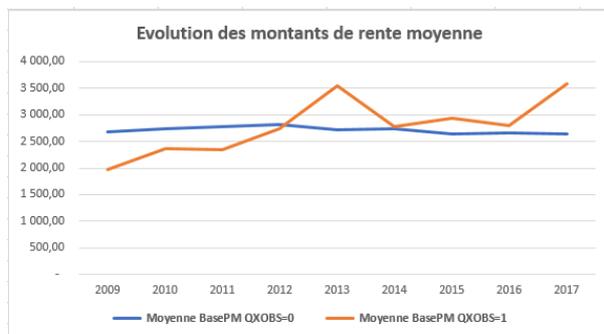


FIGURE 38

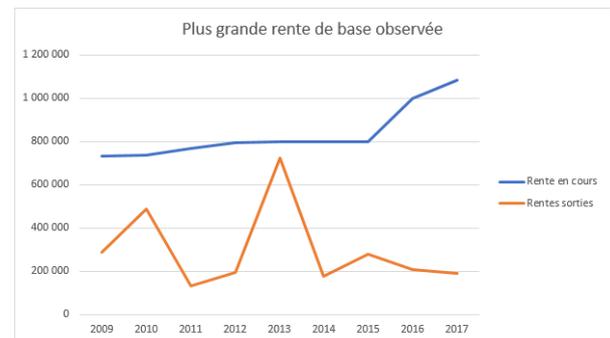


FIGURE 39

FIGURE 40 – Evolution de la répartition des rentes

Nous remarquons que les montants de rentes et les probabilités de décès agissent en complémentarité sur le résultat de mortalité. En effet, une probabilité de décès mal estimée a un impact multiplié par le montant de rentes. Ainsi, il est primordial de suivre l'évolution de notre portefeuille tant en montant qu'en probabilité.

Zoom sur les montants de rentes

Nous allons nous concentrer dans un premier temps sur le montant de la rente de base. Pour exclure l'effet "âge" du rentier, nous nous concentrons sur la rente de base et non la

provision mathématique.

Nous remarquons avec la figure 36 que le résultat de mortalité est fortement influencé par les hommes. Nous avons vérifié cette hypothèse en observant le résultat de mortalité des rentiers décédés par sexe.

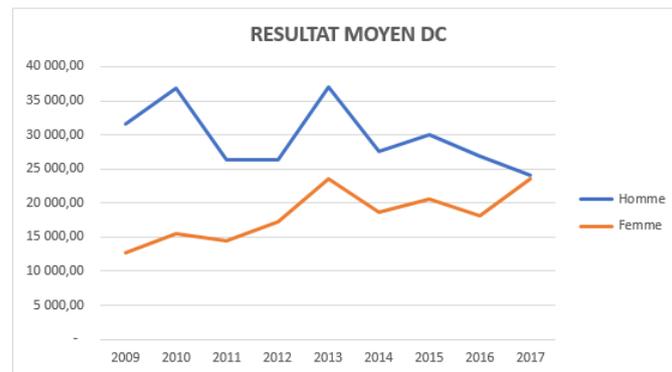


FIGURE 41 – Résultat observé pour les rentiers décédés entre 2009 et 2017

Nous remarquons que les hommes ont un résultat de mortalité plus important que les femmes, avec une tendance semblable au résultat de mortalité globale (figure 41). Mais cette tendance s'inverse en 2017.

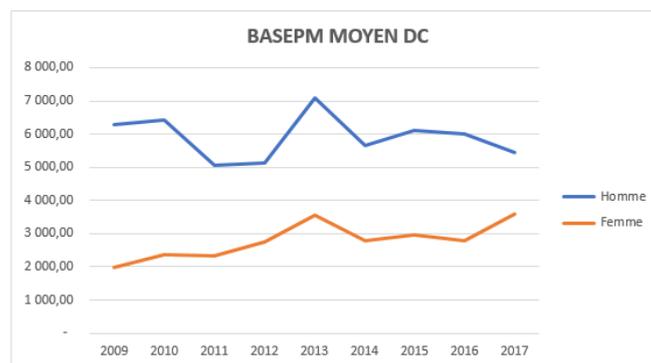


FIGURE 42 – Rentes de base moyennes observées pour les rentiers décédés entre 2009 et 2017

Lorsque nous regardons l'évolution moyenne des montants de rentes au cours du temps, nous remarquons que celle-ci est quasi identique à celle du résultat de mortalité moyenne (Figures 41 et 42). Le montant de rente influence le résultat de mortalité et ceci est d'autant plus marqué avec la répartition des montants de rentes.

ANNEE	QXOBS	q0	q5	q10	q25	q50	q75	q90	q95	q99	q100
2009	1	7	126	195	498	1 209	3 256	11 529	20 624	40 196	287 595
2010	1	18	117	192	514	1 429	4 015	11 647	20 599	43 607	487 301
2011	1	12	125	225	542	1 310	3 177	9 026	17 800	42 845	131 335
2012	1	13	136	225	542	1 312	3 360	10 525	18 505	46 259	192 188
2013	1	6	155	280	639	1 515	4 121	12 467	23 934	61 733	723 999
2014	1	4	180	313	635	1 459	3 988	11 009	18 666	46 662	177 662
2015	1	13	170	290	615	1 412	4 009	12 596	20 106	43 582	276 960
2016	1	13	166	317	630	1 382	3 914	10 986	20 279	47 670	206 162
2017	1	2	200	342	635	1 479	4 020	13 191	21 671	41 361	187 811

FIGURE 43 – Répartition des montants de rentes de base pour les rentiers décédés entre 2009 et 2017

En zoomant sur la répartition des montants de rentes des rentiers décédés, nous remarquons que lorsque nous avons des résultats proches de 0, nous observons le décès d'un rentier avec une forte rente. La figure 36 indique trois pics du résultat de mortalité, en 2010, 2013 et 2015. Pour ces trois années, les quantiles à 99% des montants de rentes au décès sont proches aux autres années mais le 100^{me} quantile nous indique des montants de rentes au décès 2 à 4 fois supérieurs aux autres années (figure 43).

Lorsque la rente est toujours en cours de service à la fin de l'année, les quantiles des montants de rentes sont stables comme le montre la Figure 44.

ANNEE	QXOBS	q0	q5	q10	q25	q50	q75	q90	q95	q99	q100
2009	0	0	213	360	687	1 609	4 081	10 660	17 927	42 845	731 381
2010	0	0	225	374	704	1 613	4 035	10 432	17 611	42 994	738 182
2011	0	0	234	386	715	1 616	3 998	10 229	17 265	42 677	768 000
2012	0	0	245	399	728	1 629	3 992	10 107	17 182	42 896	793 806
2013	0	0	259	417	731	1 583	3 853	9 725	16 540	41 048	797 839
2014	0	0	267	427	736	1 573	3 794	9 445	16 135	40 237	797 839
2015	0	0	260	420	721	1 504	3 626	8 978	15 447	39 200	797 839
2016	0	0	269	433	728	1 504	3 579	8 775	15 103	38 601	1 000 000
2017	0	0	201	398	717	1 487	3 515	8 470	14 598	37 686	1 084 573

FIGURE 44 – Répartition des montants de rentes de base pour les rentiers survivants entre 2009 et 2017

La médiane reste autour de 1600 et 99% des rentiers ont une rente inférieure à 42 000. De même, la rente maximale observée sur l'année reste identique dans le temps. Nous remarquons alors l'impact du décès d'un rentier avec une forte rente avec les années 2010, 2013 et 2015.

Nous nous proposons de regarder l'impact des montants de rentes hors rentes élevées, afin de savoir si une tendance du résultat de mortalité se dégage des observations. Pour cela, nous capons le montant de rente de base et regardons comment évolue le résultat de mortalité en dehors de ce phénomène.

Nous choisissons le montant seuil en fonction de la répartition des rentiers en général. Nous décidons de sélectionner les rentiers avec une rente supérieure à 20 000 € (quantile 97,5) et choisissons la médiane des montants de rentes pour ces rentiers comme nouvelle rente.

ANNEE	q0	q5	q10	q25	q50	q75	q90	q100
2009	20 008	20 658	21 347	23 730	29 090	41 724	71 812	731 381
2010	20 008	20 666	21 347	23 725	29 132	42 720	73 346	738 182
2011	20 002	20 618	21 346	23 692	29 133	42 899	73 530	768 000
2012	20 012	20 655	21 337	23 674	29 204	43 369	75 374	793 806
2013	20 002	20 663	21 363	23 682	29 333	43 748	76 481	797 839
2014	20 002	20 656	21 363	23 632	29 243	43 653	77 183	797 839
2015	20 017	20 662	21 381	23 712	29 318	44 404	77 465	797 839
2016	20 017	20 665	21 363	23 707	29 384	44 833	79 439	1 000 000
2017	20 017	20 675	21 379	23 731	29 445	45 155	80 774	1 084 573

FIGURE 45 – Répartition des montants de rentes supérieures au seuil définis entre 2009 et 2017

Nous décidons alors de caper les rentes à 30 000 € et nous définissons ce montant dans la suite de l'étude, le 'seuil'. Nous recalculons un résultat de mortalité par année avec les nouvelles rentes et obtenons les résultats suivants :

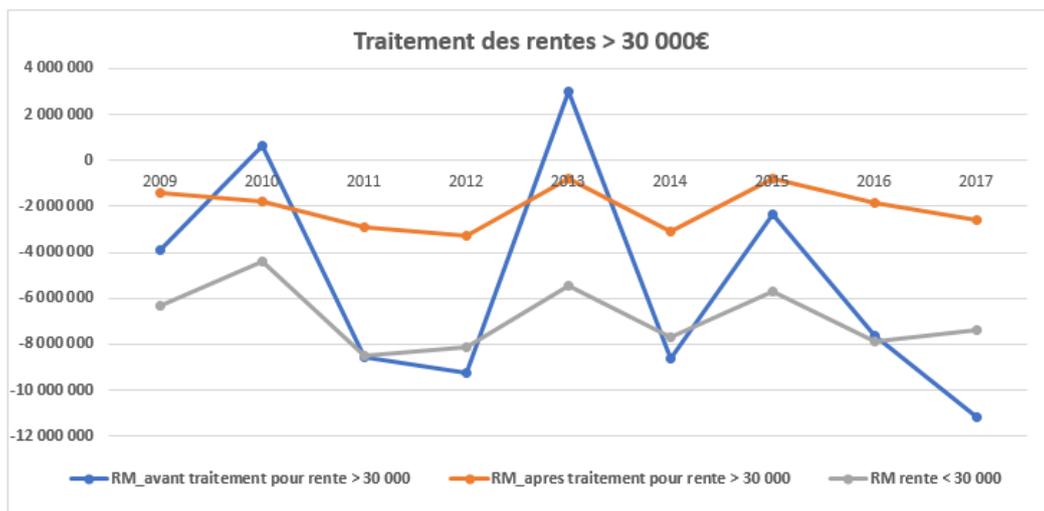


FIGURE 46 – Répartition des montants de rentes supérieures au seuil définis entre 2009 et 2017

On remarque alors une baisse de la volatilité des résultats qui reste cependant importante (environ 18 millions initialement contre 8 millions après lissage des montants de rentes par exemple entre 2011 et 2012). Aucune tendance n'est détectée.

Nous venons de voir la forte corrélation entre les montants de rentes et le résultat de mortalité. Une deuxième composante influence le résultat : les taux de décès théoriques. Les probabilités de décès sont estimées par âge et par sexe. Nous regardons dans un premier temps, la répartition des montants de rentes par âge au vue d'une éventuelle corrélation entre les montants de rentes et les rentiers âgés où le risque de longévité est plus important.

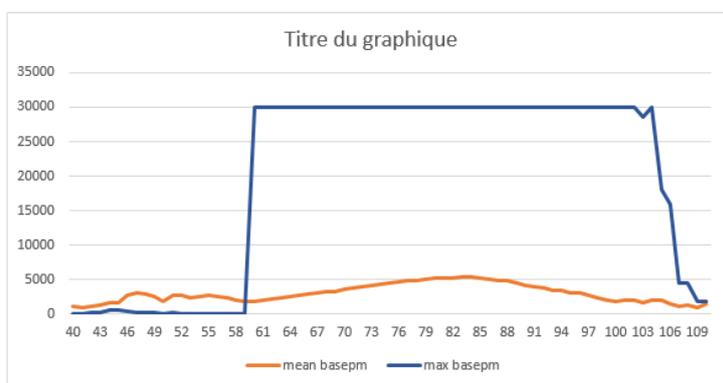


FIGURE 47 – La classe d'âges ayant atteint le seuil de 30 000

Nous remarquons alors que les rentiers riches (qui ont une rente supérieure à 30 000 €) sont âgés entre 60 et 100 ans. Il est important de savoir si l'estimation des probabilités de décès par les tables réglementaires est bien réalisée sur cette tranche d'âges. Pour cela, nous étudions l'impact de ces taux de décès par rapport aux taux observés au travers d'un second indicateur, le SMR.

2.2 Le SMR

Le SMR ou Standardised Mortality Ratio est un indicateur qui compare le nombre de décès observés avec le nombre de décès attendus. Le nombre de décès attendus est estimé à partir des taux de mortalité d'une population de référence (dans notre cas, les taux de mortalité des tables réglementaires TGH TGF 05).

Le SMR s'interprète de la manière suivante :

- S'il est inférieur à 1, le nombre de décès observés est inférieur au nombre de décès attendus (ou définis par les tables réglementaires). Il y a donc une sous mortalité de notre portefeuille ;

- S'il est égal à 1, le nombre de décès observés est égal au nombre de décès attendus. La mortalité de notre portefeuille est donc parfaitement représentée ;
- S'il est supérieur à 1, le nombre de décès observés est supérieur au nombre de décès attendus. Il y a donc une surmortalité de notre portefeuille.

Grâce à cet indicateur, nous pouvons voir si les tables de mortalité utilisées pour le calcul des engagements sont prudentes par rapport à nos données.

Dans notre cas, les provisions mathématiques de rentes sont estimées à partir de tables réglementaires TGH TGF 05. Notre population ne correspondant pas en tout point à celle utilisée pour la construction de la table réglementaire, nous étudions le SMR afin de savoir s'il existe une dérive viagère de notre portefeuille.

Remarque : Depuis le début, notre étude inclue les reversataires. Pour la suite, nous ne distinguerons plus les individus "première tête" et "deuxième tête", la population assurée ayant la même probabilité de décès. L'impact de la réversion intervient en appliquant un coefficient à la rente de base pour le provisionnement et est non négligeable.

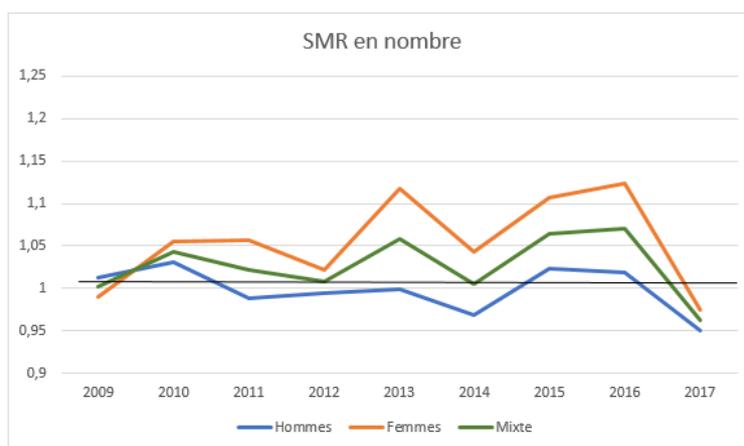


FIGURE 48 – La classe d'âges ayant atteint le seuil de 30 000 €

Ce graphique (figure ??) représente le ratio des décès observés entre 2009 et 2017 sur les taux de décès de la table de mortalité TGH TGF 05. Le calcul des taux de décès réglementaires prend en compte l'exposition au risque des rentiers et non le nombre de rentiers observés.

On remarque alors que les tables sont prudentes pour les femmes et au global : on observe un SMR supérieur à 1. Elles sont prudentes dans le sens où nous prédisons moins de décès que ce qui est observé et provisionnons donc plus que nécessaire, mais ceci doit être

modérée. En effet, il ne faut pas trop sous-estimer les taux de décès car cela impliquerait une surprovision des rentes. Concernant la population masculine, nous remarquons des années où le SMR est inférieur à 1 (par exemple en 2014). Ceci est alors très dangereux car les taux de décès sont surestimés et ce qui représente un sous-provisionnement potentiel pour ces années.

Lorsque l'on représente le SMR en montant (nous pondérons les probabilités de décès par les montants de rentes associés), la volatilité du ratio est encore plus marquée et les résultats obtenus sont très en dessous de 1 (figure 49). Cette variabilité du ratio s'explique par la taille réduite du portefeuille et l'hétérogénéité des montants de rentes des rentiers.

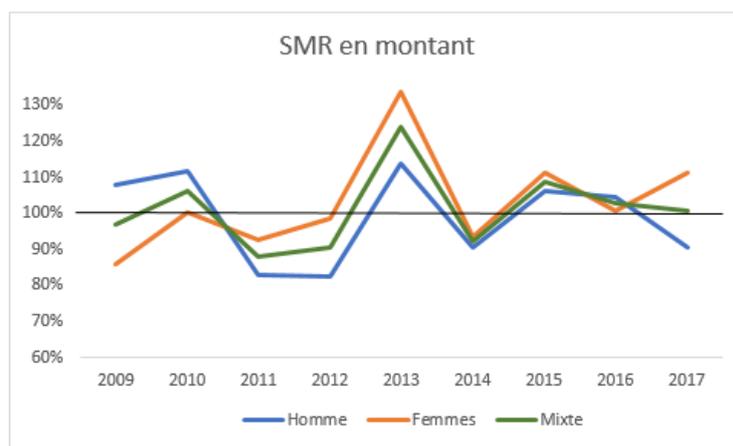


FIGURE 49 – La classe d'âges ayant atteint le seuil de 30 000 €

Tous ces résultats convergent vers une même conclusion : nos rentiers ont une tendance de mortalité différente des tables de mortalité réglementaires et cet effet est démultiplié par le poids des montants de rentes. Ce risque étant prédominant, il est important de pouvoir anticiper l'évolution de cette dérive viagère. Pour cela, nous proposons dans un premier temps une méthode de segmentation du portefeuille pour déterminer une tendance de mortalité au travers de différents profils de risques.

2.3 Amélioration du résultat de mortalité : Mise en place d'une classification pour définir des profils de risques

Le résultat de mortalité est un indicateur très utile pour estimer la dérive viagère d'un portefeuille. Au vu des résultats précédents, nous allons essayer de définir des profils de rentiers à risque à l'aide de la classification. C'est une technique très répandue lorsque

nous disposons d'un grand volume de données et que nous souhaitons y distinguer des sous ensembles homogènes.

Au vu de l'analyse de données précédente, nous avons remarqué que le résultat global comportait la même tendance que le résultat de la population masculine, que les montants de rentes de base des rentiers ainsi que la présence d'un réversataire. Cette dernière information est non négligeable car lorsque la première tête décède, la provision mathématique n'est pas entièrement libérée mais seulement d'une certaine quantité, correspondant à la quantité de réversion. Une nouvelle provision est calculée pour la seconde tête. Le gain technique est alors moins important qu'en cas d'absence de réversion. Nous allons donc découper en plusieurs classes notre population selon ces différents critères.

2.3.1 Traitement des données

Avant de procéder à notre classification, nous allons sélectionner nos données. En effet, il est primordial de traiter les données car elles déterminent la qualité des résultats obtenus autant que le modèle sélectionné. Lors de la mise en place des bases de données sous SAS, un premier traitement des données aberrantes (comme un rentier avec une date de naissance postérieure à la date de liquidation de ces droits) a été réalisé. Nous allons donc nous concentrer sur la structure des données.

Nous commençons par analyser les différentes variables disponibles. Grâce aux extractions SAS, nous avons à notre disposition les données associées à un rentier comme sa date de naissance, son âge, son sexe, la présence d'un réversataire ainsi que ces propres caractéristiques. Certaines de ces données se rejoignent (ou sont fortement corrélées) comme la date de naissance et l'âge du rentier, la *PM* et la rente de base. Ainsi, nous décidons de conserver :

- L'identifiant du rentier afin de distinguer un rentier d'un autre. L'identifiant correspond à un nombre aléatoire entre 0 et 1 000 000 ;
- L'âge ;
- Le sexe ;
- La présence d'un réversataire ;
- Le montant de rente ;
- Le résultat de mortalité observée pendant l'année ;
- L'article associé au contrat.

Nous décidons de renseigner pour chaque rentier la nature de l'article (39 ou 83) car après un premier test sur la classification hors de cette distinction, nous avons remarqué

une trop forte mutualisation du résultat pour chaque cluster. Nous avons donc regardé plus en profondeur la répartition de nos rentiers et nous avons remarqué une forte disparité entre les rentiers pour ces deux articles. Voici ci-dessous une présentation succincte des données pour chacun des deux articles :

ARTICLE 39	2009	2010	2011	2012	2013	2014	2015	2016	2017
Exposition	9 886	10 356	10 767	11 048	11 456	11 773	17 467	17 679	17 760
Rente moyenne	4 382	4 576	4 538	4 586	4 413	4 423	3 424	3 477	3 558
age moyen	71	71	72	72	72	73	76	76	76
%Hommes	62%	62%	61%	60%	60%	60%	56%	55%	55%
% Réversion	63%	62%	63%	63%	64%	64%	60%	60%	61%
% Rente > 30000 (seuil)	8%	8%	8%	8%	8%	8%	6%	6%	6%

ARTICLE 83	2009	2010	2011	2012	2013	2014	2015	2016	2017
Exposition	57 471	60 095	62 479	64 192	67 569	69 667	71 933	74 374	77 783
Rente moyenne	2 298	2 346	2 396	2 433	2 393	2 390	2 383	2 402	2 387
age moyen	72	72	72	72	73	73	73	74	74
%Hommes	68%	68%	68%	68%	67%	67%	67%	67%	67%
% Réversion	37%	38%	38%	38%	38%	38%	38%	37%	37%
% Rente > 30000 (seuil)	1%	1%	1%	1%	1%	1%	1%	1%	1%

FIGURE 50 – Répartitions des rentiers par article

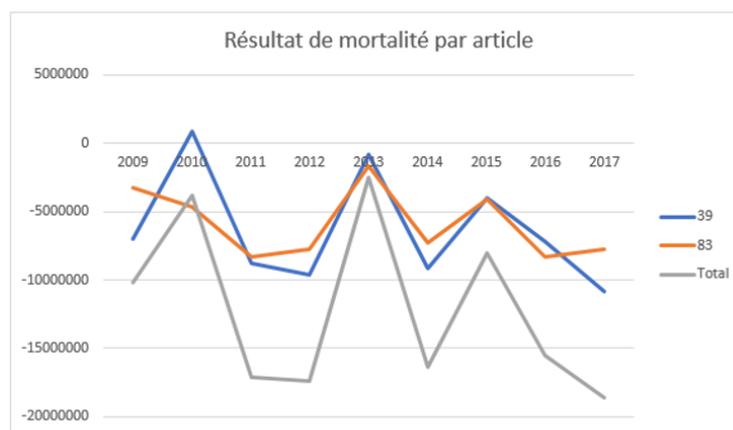


FIGURE 51 – Evolution du résultat de mortalité par article

On remarque au travers de ces tableaux que les rentiers associés aux articles 39 ont des rentes de base deux fois supérieures aux rentiers des articles 83 et représentent cependant environ 15% d'exposés au risque. La répartition Hommes/Femmes ainsi que l'âge moyen observé restent équivalents entre les deux. De même, lorsque nous regardons l'évolution du résultat de mortalité, nous pouvons constater que les résultats sont semblables. Cela

signifie que les rentiers des articles 39 sont des profils à surveiller car une nouvelle fois, ils représentent moins d'un sixième des rentiers mais une grande partie du résultat de mortalité.

Le traitement des données implique également l'analyse de données extrêmes. Une valeur extrême correspond à une donnée rare. Pour notre étude, les données caractérisées comme extrêmes correspondent aux rentiers avec une très forte rente de base. Par année, ces rentiers représentent moins d'un pour-cent de la population mais nous l'avons vu précédemment, ils ont un impact non négligeable sur le résultat de mortalité. Nous décidons donc de les conserver.

Suite à cette hypothèse, nous devons choisir une méthode de classification adaptée. En effet, alors que les valeurs extrêmes ne posent pas de problèmes significatifs dans la mise en place d'arbres de décision, elles affecteront certaines méthodes de classification comme celles basées sur le calcul de variance, les régressions logistiques ou encore l'analyse discriminante. Nous décidons de réaliser une standardisation des données afin de réduire la sensibilité à cette trop grande dispersion sans perdre leur impact. Au vu de la répartition des montants de rentes, nous n'appliquons pas de standardisation à partir de la moyenne et de l'écart-type observés sur la base mais nous centrons les données par rapport au minimum observé et nous réduisons par rapport à l'écart entre la plus petite et la plus grande valeur. Ainsi, les montants des rentes de base ne varient plus entre 1 e et 1000 0000 mais entre 0 et 1. Nous réalisons cette étape pour chaque variable explicative utilisée dans le modèle afin d'être cohérent et pouvoir comparer les individus dans l'espace.

Pour choisir le modèles adéquat à nos données, nous conjuguons plusieurs critères :

- Nous souhaitons conserver toutes les valeurs, mêmes extrêmes. C'est pourquoi nous favoriserons l'utilisation des arbres de décision ;
- Pour la méthode descriptive des données, nous souhaitons regrouper dans tout l'espace les données en classes homogènes. Nous utiliserons donc les méthodes de partitionnement comme la méthode des centres mobiles ou des K-means ou les méthodes hiérarchiques comme la CAH ;
- Nous avons en notre possession un nombre de données important pour réaliser une segmentation. Et il est à noter que certains modèles ont une complexité qui s'accroît avec l'effectif. Nous utiliserons alors une méthode de segmentation mixte. Nous alierons la méthode des centres mobiles avec la classification ascendante hiérarchique afin de réduire la complexité de celles-ci ;
- Pour la méthode prédictive, n'ayant pas de modèle déterminé en amont, nous utiliserons une méthode d'analyse probabiliste basée sur les plus proches voisins (k-plus proches voisins).

2.3.2 Mise en place des clusters

Pour réaliser la segmentation de notre population, nous utilisons une méthode mixte : nous allions ainsi une méthode hiérarchique avec une méthode de partitionnement.

On réalise dans un premier temps une première classification sur l'ensemble de notre portefeuille, par article, selon la méthode des centres mobiles en fixant le nombre de classes à 50 pour la base des adhérents aux articles 83 et 25 pour la base des adhérents aux articles 39. En effet, deux segmentations distinctes sont réalisées car les deux bases ont des profils différents ainsi qu'une composition différente (cf. section précédente) ce qui implique des hypothèses de base différentes comme le nombre de classes initiales pour le premier partitionnement. Le nombre de classes est fixé de manière assez large afin de limiter le risque que des classes naturelles fusionnent. On effectue ensuite une classification ascendante hiérarchique sur les centres de ces classes et l'on coupe le dendrogramme à une hauteur optimale c'est-à-dire après une forte perte d'inertie.

Avant d'explicitier chacune de ces méthodes, nous allons définir l'inertie inter et intra classe. On appelle inertie totale d'une population (ou de manière équivalente d'une classe), la moyenne pondérée des carrés des distances des individus au centre de la population (ou de la classe).

Si la population est segmentée en k classes, on définit l'inertie intraclasse comme la somme des inerties de chacune des classes. On dit qu'une classe est homogène si l'inertie de celle-ci est faible. Une bonne classification cherche à minimiser l'inertie intraclasse. A contrario, l'inertie interclasse est définie comme la moyenne des carrés des distances des centres de chaque classe au centre global de la population. On cherche alors à maximiser l'inertie interclasse qui représente la séparation entre classe.

La méthode des centres mobiles est une méthode de partitionnement qui se déroule comme suit :

- On choisit k individus comme centre initiaux des classes ;
- On calcul ensuite les distances entre chaque individu aux centres des classes et on affecte l'individu au centre le plus proche ;
- On redéfinit les centres des classes en fonction des nouveaux individus qui les composent ;
- On regarde si les centres sont restés stables ou si le nombre d'itération fixé en amont est atteint auquel cas, on arrête alors l'algorithme et k classes ont été définies. Sinon, on recommence à partir du deuxième point

L'avantage de cette méthode est qu'elle est applicable pour de grands échantillons de données et l'algorithme de réaffectation améliore la qualité ce qui n'est pas le cas des méthodes hiérarchiques. L'inconvénient de cette méthode porte sur l'initialisation des centres ce qui contraint à ne pas converger vers un optimum global. Pour pallier à cela, nous réitérons l'initialisation plusieurs fois et croisons les classes obtenues. L'utilisation d'une méthode mixte permet de conserver les avantages de chacune des méthodes sans leurs inconvénients. Ainsi, en utilisant la méthode des centres mobiles, il est difficile de choisir un nombre optimal de classes de la partition. Dans le cadre de la segmentation mixte, l'objectif n'est pas de définir le nombre optimal mais de définir une partition plus grossière pour ensuite procéder à une classification ascendante hiérarchique plus fine.

La seconde étape consiste à réaliser des partitions, sans fixer de nombres a priori, mais de manière hiérarchique produisant ainsi des partitions emboîtées. Voici l'algorithme de classification :

- La classe initiale est composée de tous les objets ;
- On calcul ensuite les distances entre classes ;
- On regroupe les deux classes les plus proches en une seule ; reprend le deuxième point jusqu'à avoir qu'une seule classe contenant l'ensemble des individus

On représente les partitions par un dendrogramme. Le dendrogramme est un arbre avec pour hauteur de branche la perte d'inertie interclasse. Pour choisir un nombre de classes, nous coupons le dendrogramme au dernier niveau où la hauteur des branches est élevée. Pour notre étude, nous avons choisis d'utiliser la distance euclidienne et la distance de Ward au vu du fort volume de données que nous avons et de la proximité de celles-ci. La distance de Ward permet de fusionner deux classes qui font le moins baisser l'inertie interclasse et est définie par :

$$d(A, B) = \frac{d(a, b)^2}{\frac{1}{n_A} + \frac{1}{n_B}} \quad (4)$$

où A et B sont deux classes de centres a, b et d'effectifs n_A, n_B .

Pour expliquer chacune des méthodes et les indicateurs statistiques nous permettant de conclure sur la pertinence de la classification, prenons l'exemple de l'année 2009.

Nous commençons tout d'abord par définir la base de données et récupérons pour cela les bases précédemment définies où une ligne représente un rentier. Nous définissons les

variables REV, SEXE et RICHE telle que :

$$\text{variable} = \begin{cases} 1 & \text{si vrai} \\ 0 & \text{sinon} \end{cases} \quad (5)$$

Pour la variable RICHE, elle vaut 1 si le montant de rente associé au rentier est supérieur au seuil (30 000) que nous avons fixé lors de l'étude des montants de rentes élevés. Nous séparons ensuite la base en deux, une partie pour les articles 39 et une seconde pour les articles 83. Ainsi, pour l'année 2009, l'une contient environ 10 000 rentiers avec une rente moyenne de 4 382 et 8% des rentiers considérés comme riches tandis que la seconde contient moins de 58 000 rentiers avec une rente moyenne de 2 298 et moins de 1% de rentiers riches (supérieur au seuil).

Nous procédons ensuite à la standardisation des données suivie de la première segmentation avec la méthode des centres mobiles. Pour cette étape, nous rappelons que nous ne cherchons pas les classes optimales mais seulement à réduire la dimension des données. En observant les clusters, nous remarquons des dissimilarités en terme de structures entre eux. Par exemple, un cluster est composé d'un seul rentier et un autre de plus de 3 000. Ceci est dû aux rentiers "extrêmes" c'est à dire les rentiers avec de très fortes rentes. Ces résultats sont satisfaisants car notre objectif est de construire des sous groupes homogènes de rentiers afin de définir des profils à risques pour notre étude.

Nous utilisons les centres obtenus pour réaliser une classification ascendante hiérarchique. Contrairement à la méthode des centres mobiles, nous ne choisissons pas le nombre optimal de classes a priori mais en étudiant le dendrogramme. Voici celui obtenu pour chacun des deux articles :

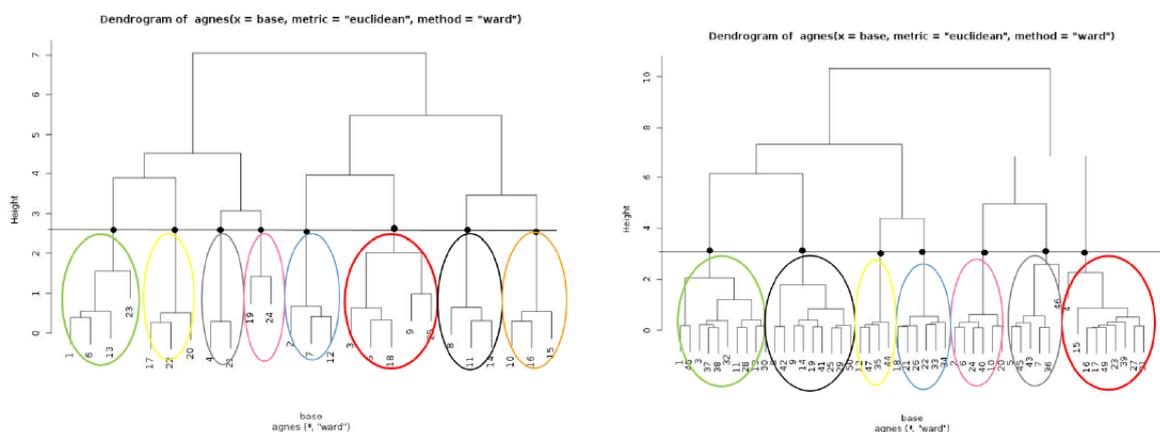


FIGURE 52 – Dendrogramme article 39 à gauche et Article 83 à droite

Nous coupons l'arbre à huit clusters pour la base des articles 39 et 7 pour les articles 83 et réalisons différents tests statistiques afin de vérifier la cohérence des clusters. Nous nous basons pour cela sur l'indice de la silhouette et le coefficient d'agglomération.

Le coefficient d'agglomération permet de regarder si la structure des groupes est cohérente. Plus celui-ci est proche de 1, plus la structure (ou l'appartenance de tous les individus à un même groupe) est justifiée. Dans le cadre de notre segmentation, nous obtenons un très bon coefficient pour les deux bases (autour de 0,9).

L'indice de la silhouette permet quant à lui de voir si l'appartenance d'un individu à un groupe est justifiée. Contrairement au coefficient d'agglomération qui regarde l'ensemble de la structure, l'indice de la silhouette regarde si un individu est correctement classé en mesurant la distance suivante :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i); b(i))} \quad (6)$$

où $a(i)$ représente la distance de l'individu i avec les autres individus et $b(i)$ la plus petite distance avec les autres clusters. Si l'indice est compris entre 0,51 et 1, on peut dire que l'appartenance de l'individu au cluster est justifiée. Nous retrouvons ci-dessous les graphiques récapitulatifs des indices de la silhouette en fonction des clusters.

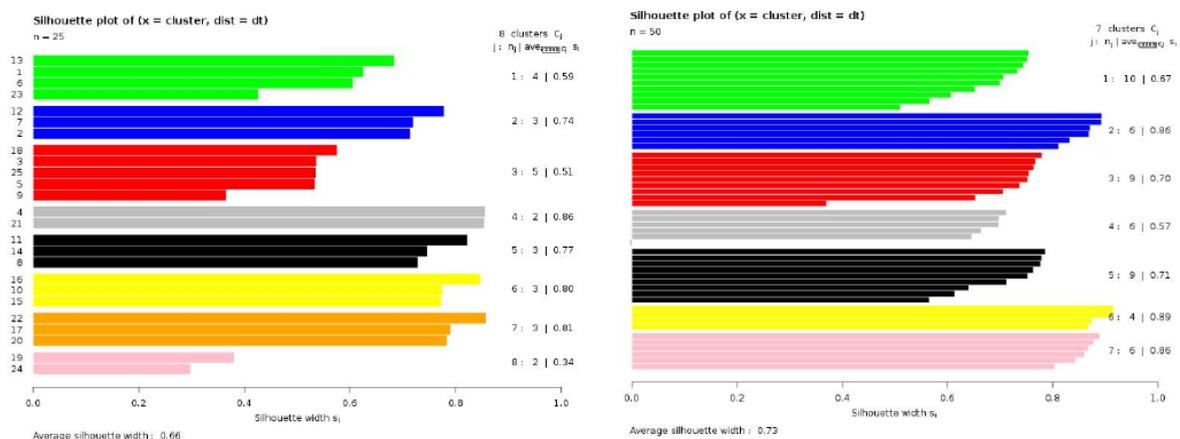


FIGURE 53 – Indice de silhouette

Nous remarquons alors que pour la segmentation sur la base 'Article 83', nous avons en général de très bons indices (entre 0,67 et 0,89). Concernant la segmentation de la base

'Article 39', un cluster comporte une structure faible au vu de l'indice. Ceci est dû aux rentes qui composent le cluster. En effet, seulement 2 centres déterminés via la méthode des centres mobiles composent le cluster. Cependant lorsque nous regardons la composition initiale de ces centres, nous remarquons que le cluster représente les femmes avec de très fortes rentes et la présence d'un réversataire.

Nous validons cette segmentation et regardons la structure des clusters obtenus. Nous remarquons alors que les classes sont essentiellement définies par 3 variables : SEXE, REV et RICHE. Ainsi, nous décidons d'appeler les sept/huit clusters de la façon suivante : FRP, FRR, FSP, FSR, HRP, HRR, HSP, HSR. La première lettre correspond au SEXE représentatif du cluster (H pour Hommes et F pour Femmes), la seconde représente la présence ou non d'un réversataire (R pour Réversion et S pour seul) et la troisième si les rentiers qui composent le cluster ont des rentes élevées (R pour Riche et P pour Pauvre). Entre les deux bases segmentées, nous remarquons que ce sont les mêmes types de clusters qui sont définis avec des moyennes d'âges et de rentes différentes. Le cluster manquant au sein de la base 'Article 83' correspond aux femmes avec réversions et riches. Lorsque nous regardons la structure de la base et zoomons sur cette catégorie, nous remarquons la présence d'un seul rentier décédé la première année. La segmentation en sept classes est donc justifiée.

2.3.3 Résultats

Nous avons réalisé la même segmentation pour chaque année d'étude, en suivant les étapes vues précédemment. Chaque année est marquée par huit clusters pour les rentes articles 39 et sept clusters pour les rentes articles 83. Les clusters obtenus sont également expliqués par les 3 variables SEXE, REVERSION et RICHE. Nous avons donc conservé nos notations en trois lettres pour chaque cluster. Voici graphiquement les différentes classes obtenues chaque année. Nous distinguons pour chaque graphique les rentes supérieures au seuil de rentes élevées de celles inférieures afin de comparer leurs évolutions et non leurs valeurs car nous savons que celles-ci sont très disparates.

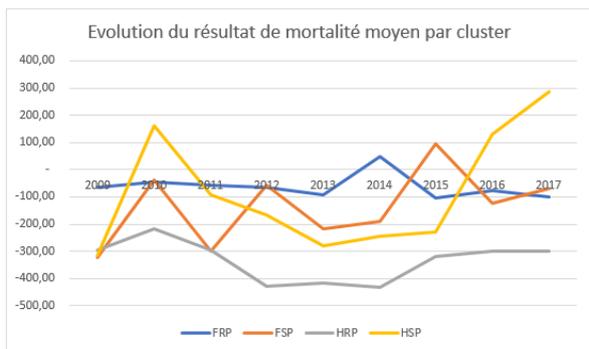


FIGURE 54 – Rentes < 30 000

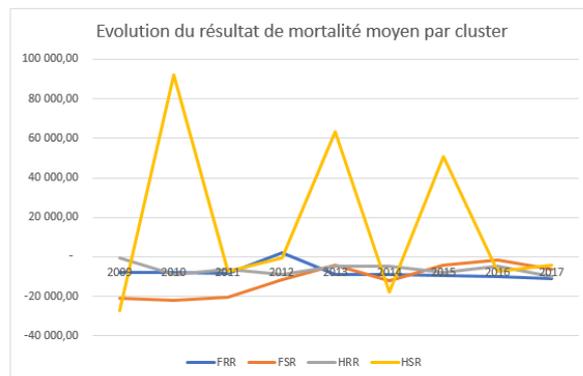


FIGURE 55 – Rentes > 30 000

FIGURE 56 – Évolution du résultat moyen observé par cluster entre 2009 et 2017 pour les articles 39

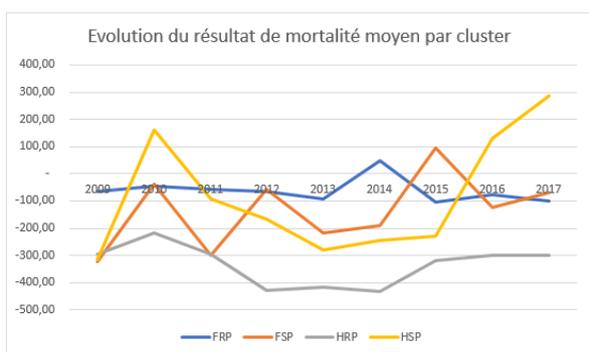


FIGURE 57 – Rentes < 30 000

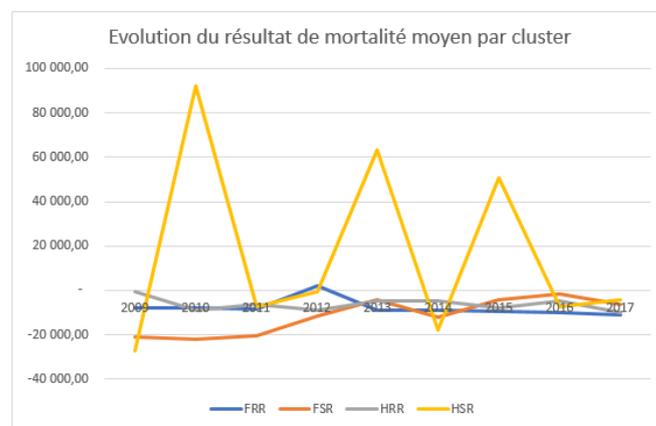


FIGURE 58 – Rentes > 30 000

FIGURE 59 – Évolution du résultat moyen observé par cluster entre 2009 et 2017 pour les articles 83

On constate une très forte volatilité des résultats de mortalité moyens, quel que soit l'article et le type de rente (supérieure ou inférieure au seuil). Cette volatilité est d'autant plus marquée pour le cluster HSR correspondant aux hommes sans réversion et riches.

2.3.4 Segmentation supervisée pour estimer le résultat de mortalité d'une année

Pour estimer le résultat de mortalité de l'année (2018), nous allons réaliser une segmentation supervisée de nos rentiers, afin de leur attribuer un cluster de référence selon ceux définis lors de notre segmentation mixte.

Nos classes étant définies par les axes SEXE, REVERSION et RICHE, nous définissons ces variables à notre base actuelle. Ce sont en effet des données connues dès le début de l'année ou lorsque les rentiers liquident leurs droits durant l'année. Nous appliquons ensuite la méthode du plus proche voisin afin de définir les clusters de références à nos rentiers. Cette méthode va comparer nos données (SEXE, REVERSION, RICHE) aux centres des clusters et attribuer le plus proche.

Afin de tester la robustesse de la méthode, nous avons réalisé entre 2010 et 2017 un back testing de la méthode en incorporant au fur et à mesure les nouveaux résultats de mortalité obtenus des clusters. Ainsi, pour estimer le résultat de l'année 2017, nous utilisons les résultats observés en 2009 et 2016.

Concernant le résultat de mortalité à attribuer par classe, nous utilisons la moyenne pondérée des résultats de mortalité observés par classe depuis 2009. Ainsi, le résultat du cluster HSR en 2017 est déterminé par la moyenne pondérée (par l'effectif) des résultats de mortalité observés du cluster HSR en 2009 et 2016. Nous obtenons les résultats suivants :

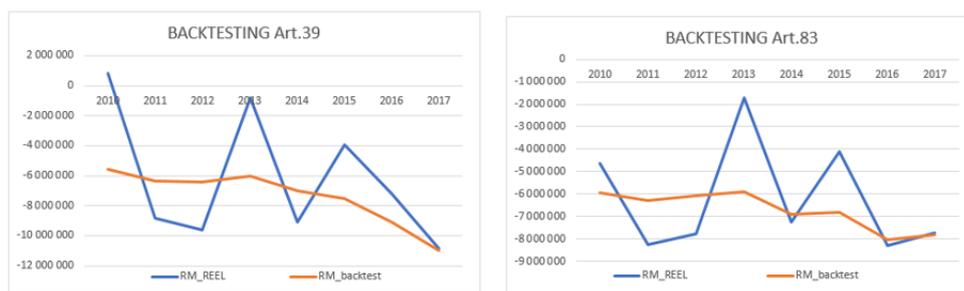


FIGURE 60 – Évolution du résultat de mortalité estimé au fur et à mesure des années par article

Nous remarquons alors que pour les articles 83 la méthode appliquée lisse bien les résultats de mortalité des années précédentes avec la même tendance.

Grâce à cette segmentation, nous avons pu détecter les profils de rentiers dits "à risque" pour l'analyse du résultat de mortalité. Une estimation basée sur les clusters est cependant

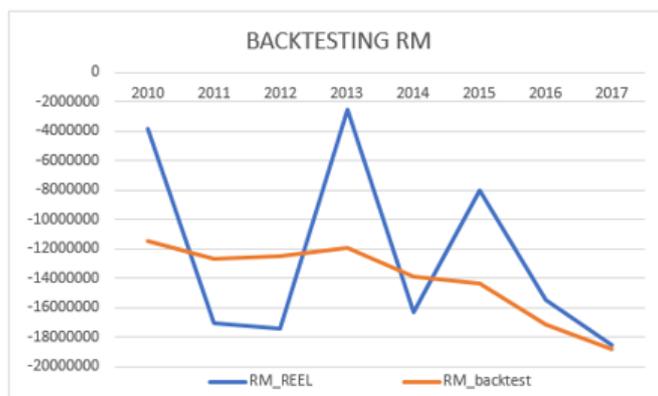


FIGURE 61 – Évolution du résultat de mortalité estimé au fur et à mesure des années - Back test global

délicate au vu du caractère aléatoire des décès des rentiers au sein du portefeuille et de l'hétérogénéité des montants de rentes observés.

Conclusion

Pour suivre l'évolution de la longévité de notre portefeuille, nous utilisons des indicateurs comme le résultat de mortalité et le SMR.

Au travers du résultat de mortalité, nous remarquons une forte volatilité des pertes techniques constatées au cours du temps. Après analyse de celui-ci, nous remarquons une influence prédominante des montants de rentes qui se trouvent très hétérogènes. Cette variable est cependant difficile à estimer ou maîtriser car elle dépend, avant tout, du rentier et de son appétence à l'assurance vie et donc des capitaux assurés. Dans le cadre de la retraite collective, l'hétérogénéité des montants de rentes assurés est très présente.

Une deuxième composante influence le résultat de mortalité : les probabilités de décès théoriques. En effet, la réglementation nous impose l'utilisation de tables de mortalité générationnelles pour le calcul de nos engagements mais celles-ci ne correspondent pas en tous points à l'évolution de la mortalité de notre portefeuille de par leurs constructions. Nous avons notamment pu mettre en lumière cette observation grâce à l'étude du SMR.

Afin de suivre l'évolution de la mortalité de notre portefeuille au cours du temps, nous avons découpé nos rentiers en différentes classes dites à risques. Nous pouvons ainsi suivre l'évolution passée du résultat de mortalité au travers de ces groupes homogènes de rentiers. Il est cependant difficile d'estimer à l'avenir l'évolution de ces groupes et les résultats



présents restent dépendants des années passées.

Afin de suivre de manière plus précise notre portefeuille et sa dérive viagère par rapport aux tables réglementaires, nous mettons en place une table de mortalité d'expérience d.

Troisième partie

Construction d'une table du moment

Nous allons tout d'abord présenter les modèles de survie dans leur globalité. En effet ceux-ci recouvrent d'une terminologie spécifique qu'il convient de présenter. Les modèles de survie sont des classes de modèles largement utilisés par les assureurs pour créer un outil de pilotage dans le cadre de la maîtrise du risque de longévité. Ces modèles sont utilisés dans de nombreux domaines de l'assurance : durée de la vie humaine, durée de l'arrêt de travail, durée de chômage, durée avant la ruine, etc. L'objectif des modèles de durée est d'étudier le délai de survie jusqu'à la survenance d'un évènement.

1 Retraite Collective & Modèle de durée

Un des risques majeurs en assurance vie, notamment en retraite collective et le risque de longévité c'est à dire le risque lié à la durée de vie humaine. En effet, l'assureur est soumis au code des assurances, qui impose aux assureurs d'évaluer, d'une manière fiable leur engagement à tout moment, vis-à-vis de leurs assurés. Dans le cadre de la retraite collective, l'assureur évalue la provision mathématique à partir des tables réglementaires où des tables d'expériences construisent à partir de son historique certifié par l'autorité nationale.

Ainsi, l'assureur peut provisionner plus ou moins par rapport aux PM obtenues avec les tables réglementaires. Cette différence représente un coût et donc un enjeu majeur pour l'entreprise.

1.1 Définitions des notions de bases

A partir de la loi de la variable aléatoire T , nous pouvons déterminer les fonctions suivantes.

1.1.1 Fonction de survie

Soient T la variable aléatoire du temps de survie et S la fonction de survie de la variable aléatoire T . La fonction de survie S de la variable aléatoire T correspond à la probabilité de survivre jusqu'à l'instant x . Elle est définie par :

$$S(x) = P(T \geq x) = 1 - F(x-)$$

Où F correspond à la fonction de répartition de la variable T qui est définie par :

$$F(x) = P(T \leq x), \quad x \in \mathbb{R}^+$$

1.1.2 Taux de risque instantané

Le taux de risque instantané fournit un indicateur adapté aux modèles de durée, plus visuel, que la fonction de survie. Ce taux de risque instantané μ de la variable aléatoire T définit la probabilité de survivre au-delà de T_i sachant que l'individu a déjà vécu jusqu'à T_{i-1} . Celui-ci permet de visualiser plus facilement les propriétés liées aux durées. Il est défini en deux temps, d'abord dans un cas discret ensuite dans un cas continu

— Cas discret

On rappelle que, dans ce cas, la variable de durée T prend ses valeurs dans un espace dénombrable, on définit :

$$\mu(t) = P(T = t | T \geq t) = \frac{P(T=t)}{P(T \geq t)}, \quad \forall t \in \mathbb{R}^+$$

On retrouve une forme de la fonction de survie en fonction du taux de risque instantané dans le cas discret :

$$S(t) = \prod_{i|t_i < t} (1 - \mu(t_i))$$

A noter que le taux de risque instantané est nul si t n'appartient pas à l'ensemble des valeurs prises de T .

— Cas continu

Dans le cas continu, le taux de risque instantané est défini par :

$$\mu(t) = \lim_{dt \rightarrow 0^+} \frac{P(T \in [t, t+dt] | T \geq t)}{dt}$$

$$\forall t \in \mathbb{R}^+, \mu(t) = -\ln S(t)$$

Remarque : Nous pouvons retrouver à partir du taux de risque instantané, le taux de risque cumulé qui est défini selon le cadre (discret ou continu) :

Cas discret	Cas continu
$\sum_{t_i \leq t} \mu(t_j)$	$\int_{-\infty}^t \mu(u) du$

1.1.3 Notations actuarielles

Nous utiliserons dans cette partie les notations propres à la littérature actuarielles. Parmi celles-ci :

l_x : nombre de survivants à l'âge x .

Puisque l'on se place dans le cadre d'une loi de mortalité, celui-ci peut être définie comme le nombre de personnes vivant à l'âge x . En effet, la définition de ce paramètre est d'autant plus importante qu'elle permettra de déterminer une relation d'abord avec la fonction de survie définie précédemment, ensuite par la probabilité à estimer.

A partir de cette relation, on retrouve la fonction de survie :

$$S(x) = \frac{l_x}{l_0}$$

$p_x(t)$: Probabilité que la durée de survie soit supérieure à t sachant x Dans la notation actuarielle, elle est calculée en utilisant une table mortalité et est définie par :

$$p_x(t) = P(T \geq x + t | T \geq x) = \frac{l_{x+t}}{l_x}$$

q_x : probabilité que l'évènement d'intérêt survienne, elle est définie par :

$$q_x = \frac{S(x) - S(x+1)}{S(x)}$$

$$q_x = 1 - \frac{S(x+1)}{S(x)}$$

Dans le cadre des études relatives aux durées, on est confronté le plus souvent à deux phénomènes liés à des observations incomplètes que nous tenterons de définir. Ainsi au lieu d'observer des réalisations indépendantes et identiquement distribuées de T , les données de l'échantillon sont soumises à diverses perturbations qui empêchent d'observer complètement le phénomène. Nous allons donc définir les notions de censures et de troncatures

1.1.4 Phenomene 1 : La Censure

Communément appelée censure à droite, la censure est un phénomène qui affecte plus généralement la partie droite de la distribution.

Soit T le temps de survie. Pour n observations indépendantes et identiquement distribuées, on détermine (T_1, T_2, \dots, T_n) de même loi que T , les temps de survie.

Soit (C_1, C_2, \dots, C_n) les temps censurées pour n observations (iid) de même loi que C (variable de censure).

Soit (Y_1, Y_2, \dots, Y_n) les durées de survie observées pour les n observations.

On définit ensuite un indicateur de censure qui prend la valeur 0 si l'observation est incomplète et 1 si l'observation est complète.

On dit qu'il y'a censure pour l'échantillon (T_1, T_2, \dots, T_n) , si au lieu d'observer cet échantillon, nous observons : $(Y_1, \delta_1), (Y_2, \delta_2), \dots, (Y_n, \delta_n)$ définit par.

$$\begin{cases} Y_i = \inf(T_i, C_i) \\ \delta_i = 1_{T_i \leq C_i} \end{cases} \quad (7)$$

Nous observons donc la sortie uniquement si elle a lieu avant la date C .

Ainsi, nous distinguons 3 types de censure : la censure à gauche, la censure à droite et la censure par intervalle. Puisque nous nous plaçons sur les nouveaux décès, c'est-à-dire ceux survenus entre la période de début et de fin d'observation, nous sommes confrontés à un phénomène de censure à droite.

1.1.5 Phenomene 2 : La troncature

On distingue comme pour la censure, trois types de troncature : la troncature à gauche, la troncature à droite et la troncature par intervalle. Dans le cadre de notre étude, nous nous plaçons sur une troncature à gauche impactant négativement la partie gauche de la distribution. Soient,

(T_1, T_2, \dots, T_n) de mêmes loi que le couple (T, C)

$(\tau_1, \tau_2, \dots, \tau_n)$ iid de même lois qu'une variable de troncature noté τ

Dans ce cas de figure, on observe : $(Y_1, \delta_1, \tau_1, \dots, Y_n, \delta_n, \tau_n)$ avec $Y_i \geq \tau_i$

$$\begin{cases} Y_i = \inf(T_i, C_i) \\ \delta_i = 1_{T_i \leq C_i} \end{cases} \quad (8)$$

T_i : âge de l'individu i au moment du décès

C_i : âge de l'individu i quand celui-ci quitte le portefeuille pour une autre cause que le décès

Y_i : âge de l'individu i quand il quitte le portefeuille, quelque soit la raison

τ_i : âge de l'individu i au début de la période d'observation.

Les deux phénomènes sont très différents, car dans le cas de la troncature, on perd complètement l'information sur les observations en dehors de la plage alors que dans le cas de la censure, on a connaissance du fait qu'il existe une information, mais on ne connaît sa valeur précise, simplement le fait qu'elle excède un seuil.

Une fois, les troncatures et les censures prises en compte, nous pouvons désormais étudier les modèles d'estimations des taux bruts.

2 Ajustement d'une table du moment

Il est nécessaire de commencer par estimer les taux bruts de décès du portefeuille. La qualité des travaux d'ajustement dépendent en partie de cette estimation. Pour ce faire les estimateurs de Hoem et Kaplan Meier présentés dans la littérature actuarielle vont faire l'objet d'une étude et d'une application aux données du groupe de rentiers.

2.1 Estimation des taux avec le modèle de Hoem

La méthode des moments de Hoem est une méthode qui permet d'estimer de manière paramétrique les taux de décès en présence de censure et de troncature. Cette méthode ne calcule pas la fonction de survie, elle estime directement les taux bruts de décès.

La méthode repose sur deux hypothèses : les décès sont indépendants les uns des autres et le nombre de décès à un âge x suit une loi binomiale $B(n_x, q_x)$.

On complète l'estimateur binomial en y introduisant la prise en compte de censures et de troncatures. Il repose sur les mêmes hypothèses. Il est appliqué dans les échantillons où tous les individus sont vivants à l'âge x et où la seule cause de sortie possible est le décès.

Principe

Considérons qu'un assuré k n'est exposé au risque de mortalité dans la classe d'âge considérée ($[x, x + 1[$) qu'entre les dates α_k (date du début d'observation de l'assuré d'âge x) et β_k (date de fin d'observation de l'assuré d'âge x , par décès, à défaut par censure, ou par changement d'âge). En effet la date de fin du contrat ou fin de la période d'observation ou résiliation est connue. Ainsi, seules les périodes où il y avait effectivement un risque pour l'assureur de cet assuré i sont prises en compte.

Notation

- n_x : nombre d'individus survivant à l'âge x
- D_x : variable aléatoire traduisant le nombre d'individus décédés entre l'âge x et $x + 1$
- i_x : nombre d'individus décédés entre l'âge x et l'âge $x + 1$
- q_x : taux de décès à l'âge x

- $[\alpha_k, \beta_k] \subset [x, x + 1[$: intervalle dans lequel on observe l'assuré i dans le portefeuille entre l'âge x et $x + 1$.
- $(X_1, X_2, \dots, X_{n_x})$: n_x variables de Bernoulli indépendantes, qui représente l'état de l'assuré (sinistré ou non).

Hypothèse

On suppose que la probabilité qu'un individu d'âge x décède entre x et $x+t$ est une fonction linéaire du temps des X_k , ces variables sont supposées indépendantes et identiquement distribuées, $D_x = \sum_{k=0}^{n_x} X_k$ comptabilise le nombre de décès observés sur la classe d'âge $([x, x + 1[)$. Ceci implique que le nombre de décès D_x suivra une loi binomiale :

$$l_x \sim B(n_x, q_x * (b_k - a_k))$$

Enfin, en passant par la maximisation de la log vraisemblance de cette loi binomiale avec q_x comme inconnu, nous pouvons en déduire la formule du taux de décès associé à chaque âge.

$$\frac{d \ln \hat{q}_x}{d q_x} = \sum_{i=1}^{n_x} \frac{i_x * (1 - (b_k - a_k) q_x) - (n_x - i_x) (b_k - a_k) \hat{q}_x}{(b_k - a_k) \hat{q}_x (1 - (b_k - a_k) \hat{q}_x)}$$

On maximise maintenant cette quantité pour obtenir la valeur optimale de q noté \hat{q} qui annule la dérivée ci-dessus. Après calculs, nous arrivons au résultat ci-dessous traduisant ainsi le taux de décès bruts par la méthode classique de Hoem :

$$\hat{q}_x = \frac{i_x}{\sum_{k=1}^{n_x} (b_k - a_k)}$$

Notons que l'estimateur des moments de Hoem est sans biais. Il s'avère que cet estimateur est particulièrement bien adapté si les hypothèses paramétriques adoptées sont proches de la loi de mortalité.

En appliquant cette méthode à notre portefeuille, nous pouvons alors tracer les \hat{q}_x bruts homme, femme et ensemble (unisexe). Ces taux sont déterminés pour la période allant du 01/01/2009 au 31/12/2017 et pour tout âge $x \in [55; 100]$ ans. Le graphique ci-dessous les présente.

Une première observation est le fait que la courbe unisexe soit en dessous de la courbe homme et au dessus de la courbe femme. Cela semble en adéquation avec les statistiques nationales sur la mortalité. De plus, les courbes nationale ont une allure exponentielle, ce qu'on retrouve ici jusqu'à environ 95 ans. Au delà, les q_x deviennent beaucoup plus volatils. Cela s'explique en partie par le fait que l'exposition est moindre et donc qu'un

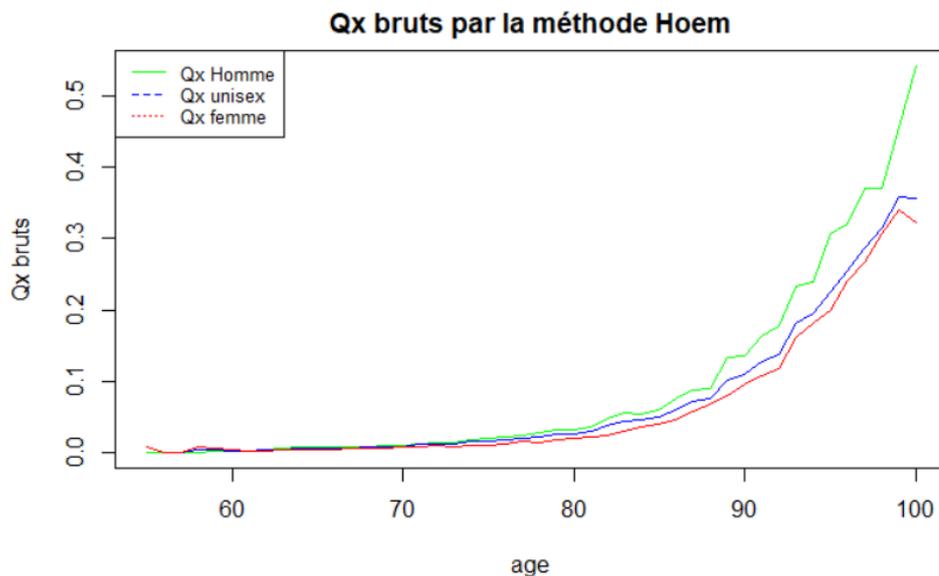


FIGURE 62 – Comparaison des qx brut homme, femme unisexe par âge

décès supplémentaire peut avoir un impact significatif sur les q_x .

Il est intéressant de construire un intervalle de confiance asymptotique de cet estimateur afin d'en apprécier la précision.

Nous avons vu que les variables $(X_1, X_2, \dots, X_{n_x})$ sont n_x variables de Bernouille indépendantes telles que :

$$\begin{aligned} \forall i \in [1, E_x], E(X_i) &= q_x \\ \forall i \in [1, E_x], Var(X_i) &= q_x * (1 - q_x) \end{aligned}$$

Dès lors, nous allons pouvoir appliquer le théorème central-limite en posant E_x l'exposition au risque d'âge x :

$$E_x \frac{q_x - \hat{q}_x}{\sqrt{q_x(1-q_x)}} \hookrightarrow \mathcal{N}(0, 1)$$

Ainsi, l'intervalle de confiance asymptotique de niveau α pour q_x est donné par :

$$I_\alpha = \left[\hat{q}_x - u_{\alpha/2} \sqrt{\frac{q_x(1-q_x)}{E_x}}; \hat{q}_x + u_{\alpha/2} \sqrt{\frac{q_x(1-q_x)}{E_x}} \right]$$

ou u_α représente le quantile d'ordre α de la loi normale centrée réduite. On peut montrer qu'en remplaçant dans l'expression précédente q_x , inconnu, par son estimateur \hat{q}_x , que l'intervalle de confiance obtenu asymptotiquement est également de niveau α des taux bruts unisexes.

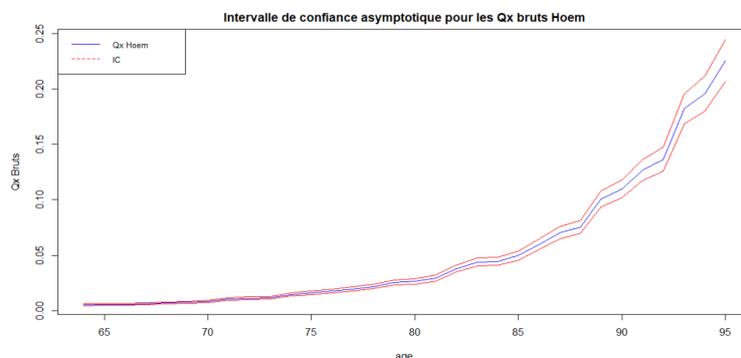


FIGURE 63 – Intervalle de confiance de niveau α des Q_x bruts de Hoem

Nous remarquons que l'intérieur de l'intervalle formé par les intervalles de confiance à 95% est beaucoup plus large aux âges élevés (au delà de 85 ans). Ceci s'explique par le fait que la construction de ces intervalles dépend de la variance de l'estimateur, qui est le rapport entre $\hat{q}_x(1 - \hat{q}_x)$, E_x au dénominateur. Aux âges élevés, l'exposition devient de plus en plus faible du coup la variance augmente.

Avant de voir d'autres méthodes de construction des taux bruts, nous allons restreindre notre plage d'étude en sélectionnant les taux bruts fiables.

2.1.1 Choix de la tranche d'âge des taux bruts

Nous avons vu précédemment que nous pouvons calculer les taux bruts de mortalité à partir de la fonction de survie estimée. On calcule donc ces taux bruts pour chaque âge x .

Cependant, il se pose le problème de savoir si pour chaque âge donné, la taille de la population est statistiquement significative pour le calcul du taux brut correspondant, ce qui explique les fluctuations pour les grands âges.

Nous allons, pour répondre à cette question, nous baser, d'une part, sur le critère de Cochran et le quantile à 95% de la distribution de la population sous risque, et d'autre part, imposer la contrainte de croissance des taux de mortalité. Le critère de Cochran est vérifié lorsque pour un âge x , les relations suivantes sont vraies :

$$L_x * \hat{q}_x \geq 5 \text{ et } L_x * (1 - \hat{q}_x) \geq 5$$

Dans notre cas, ce critère est vérifié pour la tranche d'âge [64, 104] ans.

En imposant la croissance à la courbe des taux bruts, et après avoir considéré le critère de Cochran, nous retenons la tranche d'âge [64 ; 95] ans.

Enfin, pour les âges supérieurs à 95 ans, une extrapolation des taux sera faite pour compléter la table de mortalité.

Regardons cependant une autre méthode de construction des taux bruts en considérant toujours cette plage d'âges, la méthode de Kaplan Meier

2.2 Estimation des taux avec le modèle de Kaplan Meier

La méthode d'estimateur de Kaplan Meier est un estimateur non paramétrique (c'est-à-dire ne partant d'aucune loi a priori) de la fonction de survie, qui permet, tout comme Hoem, de gérer les phénomènes de censure et de troncature. La fonction de survie S est définie par :

$$S(x) = 1 - F(x)$$

Comme vu à la section précédente, on peut déterminer la distribution de la variable T à partir du taux de risque instantané

$$F(t) = 1 - \prod_{i|t_i < t} (1 - \mu(t_i))$$

La construction de l'estimateur de Kaplan Meier repose sur l'expression du taux de risque instantané en fonction de H_0 , H_1 et H définies par.

$H(t) = P(Y \leq t)$: la fonction de répartition de la variable Y . Donc un estimateur de cette variable est $\hat{H}(t) = \frac{1}{n} \sum_{i=1}^n 1_{Y \leq t}$ et donc.

$H_1(t) = P(Y \leq t, \delta = 1)$ et $H_0(t) = P(Y \leq t, \delta = 0)$

A partir de ces quantités, nous pourrions obtenir \hat{F}

Soient C : variable aléatoire de censure et $G(t) = P(C \leq t)$, la fonction de répartition de la variable de censure. Si T est indépendant de C , pour quelque soit ϕ telle que $E[|\phi(Y)|]$

$$E[\delta \phi(Y)] = E[(1 - G(T-))\phi(T)]$$

Introduisons les quantités H_1 et H

H_1 : probabilité de sortie pour une cause décès

H : probabilité de sortie pour n'importe quelle cause

Dans le cas où la troncature n'est pas considérée, F peut être estimée par H_1 . Ainsi le taux de risque instantané dans le cas continu est :

$$\mu(t) = \frac{P(T=t)}{P(T \geq t)} = \frac{dF(t)}{1-F(T-)}$$

$$\mu(t) = \frac{dH_1(t)}{1-\hat{H}(t-)}$$

Un estimateur naturel de celui-ci serait de compter le nombre d'individus d'âge X sortis pour une cause décès sur le nombre total d'individus d'âge X , ce qui conduit à la formule suivante :

$$\mu(t) = \frac{dH_1(t)}{1-\hat{H}(t-)} = \frac{\{i: X_i=t, \delta_i=1\}}{\{i: X_i \geq t\}}$$

L'estimateur de Kaplan est défini comme :

$$F(t) = 1 - \prod_{Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^n 1_{Y_i \leq j}}\right)$$

Une fois la fonction de survie estimée, les taux bruts sont définis par :

$$\hat{q}(x, t) = 1 - \frac{\hat{S}_t(x+1)}{\hat{S}_t(x)}$$

Passons maintenant à l'application de Kaplan-Meier grace au package *Survival* du logiciel R. Pour l'ensemble de notre population, nous pouvons alors tracer la courbe des taux bruts de mortalité estimés par Kaplan-Meier dans l'intervalle [64 : 95] ans.

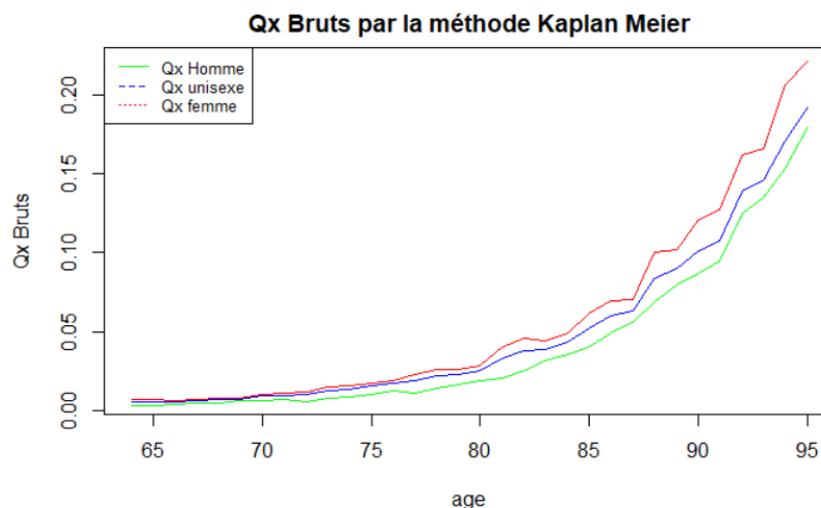


FIGURE 64 – Taux bruts de mortalité par âge estimés par Kaplan Meier

Nous remarquons, comme avec la méthode de Hoem, que la courbe unisexe de Kaplan-Meier est entre la courbe homme et la courbe femme. Nous obtenons ainsi avec le même package survival un intervalle de confiance de l'estimateur.

D'abord, notons bien que ce n'est pas un intervalle de confiance asymptotique. Ceci est un intervalle de confiance pour mesurer l'erreur d'estimation de la fonction de survie. Après une transformation de cette fonction par la formule $q_x = 1 - \frac{s(x+1)}{s(x)}$, nous obtenons un intervalle mesurant l'erreur d'estimation.

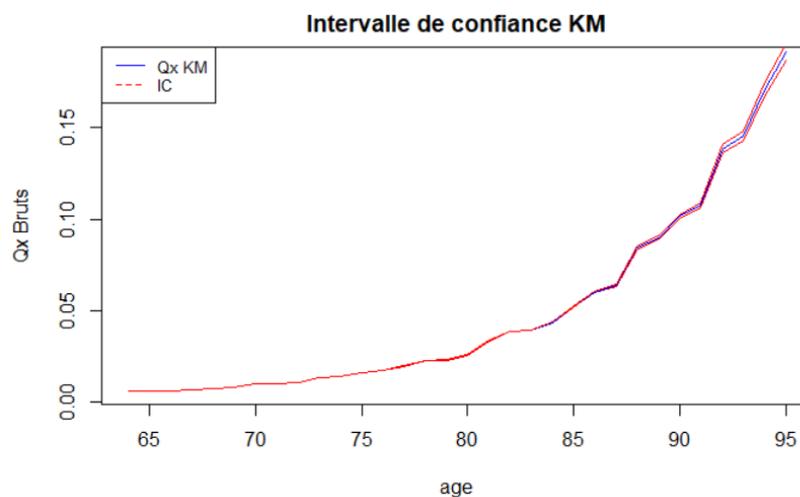


FIGURE 65 – Intervalle de confiance KM du package survival

Nous observons que les q_x se trouvent bien dans l'intervalle, de plus, les écarts sont presque confondus avec les q_x entre les âges 64 et 90 ans. Au delà, l'amplitude de l'intervalle augmente. En effet, on constate que plus l'exposition diminue plus l'écart augmente, mais ça reste stable avec la plage d'âges choisit (entre 64 et 95) et explose avec les âges élevés. Tout de même, on peut conclure Kaplan-Meier estime bien les taux bruts au vu des faibles écarts. Il reste maintenant de comparer Kaplan-Meier avec les taux obtenus avec la méthode classique (Hoem).

2.3 Comparaison des deux estimateurs

Nous pouvons observer Kaplan-Meier sous-estime les taux obtenus avec la méthode de Hoem. De plus, elles sont toutes deux croissantes d'une manière exponentielle entre les âges 64 et 95 ans, avec des taux beaucoup plus faibles pour les femmes comme nous pouvons le remarquer dans le cas des tables réglementaires TGH, TGF 05.

Nous allons maintenant calculer le ratio entre les deux estimateurs et représenter le rapport sur un graphique.

Cela consiste à évaluer les quotients $\frac{q_x^{KM}}{q_x^{HOEM}}$ où q_x^{KM} désignent les taux bruts estimés par la méthode de Kaplan Meier et, q_x^{HOEM} désignent les taux bruts estimés par la méthode de HOEM.

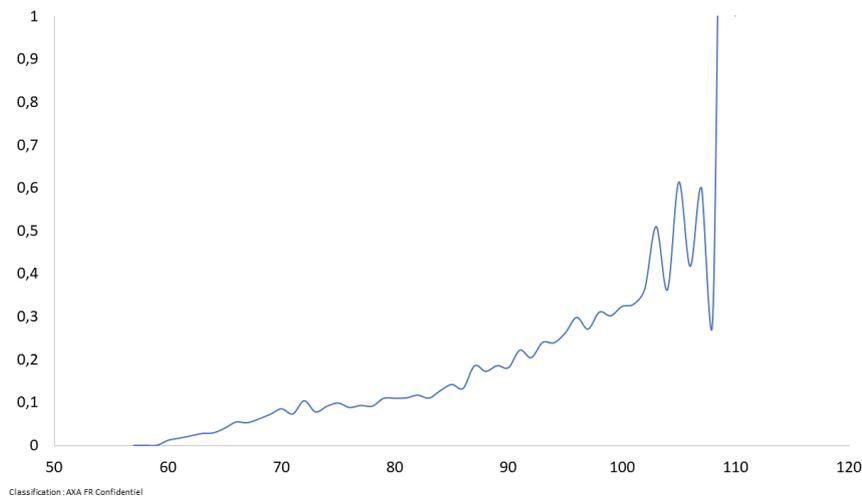


FIGURE 66 – Comparaison des taux bruts, estimés par les méthodes de HOEM et Kaplan Meier

Nous constatons que le ratio a une tendance croissante avec l'âge, et une fluctuation pour les grands âges. C'est à dire que les q_x de Kaplan-Meier explose avec l'âge par rapport aux q_x de Hoem qui semble beaucoup plus stable.

Pour approfondir cette remarque, voici un tableau qui resume la moyenne des ratios sur la période d'observation (entre 2009 et 2017). Nous rappelons que l'exposition augmente en fonction de l'année, nous avons donc plus d'assurés dans notre portefeuille dans les dernières années.

Les moyennes du rapport des taux bruts Kaplan Meier/HOEM sont les suivantes.

Année	2009	2010	2011	2012	2013	2014	2015	2016	2017
Ratio	1,77	1,74	1,72	1,68	1,63	1,60	1,57	1,54	1,51

TABLE 5 – La moyenne du ratio par année

La moyenne diminue au fur et à mesure des années. Ceci s'explique par le fait que l'exposition augmente au fur des années. A la seule prise en compte de cette remarque, la conclusion serait que les taux bruts de Kaplan Meier et de Hoem converge vers le même objet (quand E_x tend vers infini).

Pour aller plus loin, une alternative est d'étudier la convergence des deux méthodes d'estimation. En effet, nous savons que la méthode de Hoem est sans biais. De ce fait, si l'estimateur de Kaplan Meier est sans biais alors les deux estimateurs doivent converger vers même objet.

Donc nous allons comparer l'intervalle de confiance de Hoem avec l'estimateur de Kaplan Meier.

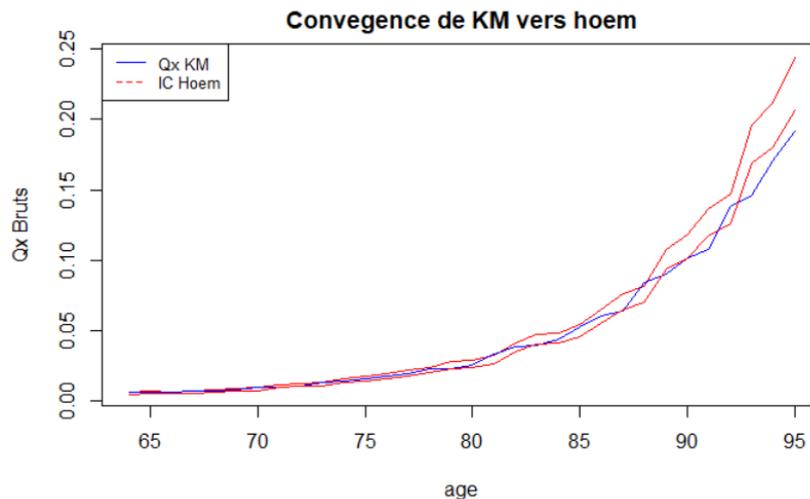


FIGURE 67 – Les taux bruts de Kaplan Meier dans l'intervalle de confiance de Hoem

L'estimateur de Kaplan-Meier est compris dans l'intervalle de confiance à 95% de l'estimateur de Hoem entre la plage d'âge 64 et 88 ans, de plus, l'exposition est plus grande dans cette plage. Au delà, l'estimateur n'est pas dans le corridor de l'intervalle. Puisque l'intervalle de confiance de Hoem converge quand l'exposition est grande, nous pouvons donc conclure que les deux estimateurs convergent vers le même objet. Ceci est rassurant car les 2 approches doivent en théorie modéliser la même mortalité sous-jacente.

Au vu de cette analyse ci-dessus, les taux bruts estimés par la méthode de Kaplan-Meier sont retenus pour la suite de l'étude.

Dès lors que les taux bruts de mortalité du portefeuille ont été déterminés, les modèles de lissage se proposent d'ajuster ces taux bruts à un modèle non paramétrique. Cet ajustement permet d'effectuer un lissage des fluctuations dues aux caractéristiques de l'échantillon de référence et de proposer une projection de ces taux dans le futur.

2.4 Lissage des taux bruts

En général, les taux annuels bruts de mortalité ont un caractère erratique dû aux fluctuations d'échantillonnage. Afin d'identifier les caractéristiques de la mortalité propres à la population étudiée, il est nécessaire de lisser ces taux. Les techniques de lissages reviennent à remplacer les taux de mortalité initiaux par des taux de mortalité plus réguliers, tout en gardant l'information des taux de mortalité bruts. Il est possible de faire appel à plusieurs méthodes de lissages non paramétriques. Cependant, nous allons nous contenter d'utiliser la méthode de Whittaker-Henderson et le lissage géométrique pour concilier ces deux éléments.

L'avantage énorme de lissage, au lieu des modèles d'ajustements paramétriques, c'est un outil non paramétrique qui nécessite très peu d'hypothèses pour converger vers le bon objet.

2.4.1 Lissage Whittaker-Henderson

Il s'agit d'une méthode non paramétrique très utilisée pour le lissage des taux bruts. Elle doit son nom à E. T. Whittaker et à R. Henderson. Elle cherche à concilier deux principaux critères : le critère de fidélité et le critère de régularité. La fidélité consiste par le fait que les taux estimés par âge doivent être proches des taux bruts. La régularité repose sur l'hypothèse que la courbe des taux estimés doit être aussi régulière que possible. L'objectif étant donc de trouver les taux qui maximisent ces deux critères. Pour parvenir à concilier ces deux critères, nous définissons d'abord un opérateur de différentiation pour toute fonction u par :

$$\forall x, \Delta u(x) = u(x+1) - u(x)$$

Ensuite, on définit la fonction qui définit le critère de fidélité. C'est la distance au carré pondéré par les poids w_x :

$$F(c) = \sum_{x=1}^{X_{max}} w_x (c_x - \hat{q}_x)^2$$

Où w correspond à une matrice de poids. Dans notre cas, w correspondrait à une matrice de nombre de personnes à risques à tout instant (l'exposition donc) calculée proportionnellement à la courbe de survie. D'autre part, on définit la fonction caractérisant le critère de régularité. Elle est égale à la somme des carrés des dérivés d'ordre z . Nous définissons la dérivés par une variation entre deux âges comme l'opérateur de différentiation ci-dessus. Nous avons donc :

$$S(c) = \sum_{x=1}^{X_{max}-z} ((\Delta^z c)_x)^2$$

Où z correspond au degré de polynôme, de déviation.

Enfin, la courbe des taux lissés par la méthode de Whittaker Henderson est alors celle qui minimise :

$$WH(c)_{z,h} = F(c) + hS(c)$$

Une fois le lissage effectué, il convient d'en valider la qualité. En effet, il s'agit de vérifier l'adéquation des taux ajustés aux taux bruts estimés.

Nous avons effectué ces lissages sur notre portefeuille et sur la base des qx donnés par la méthode de Kaplan - Meier. Les résultats sont résumés dans le graphique qui suit.

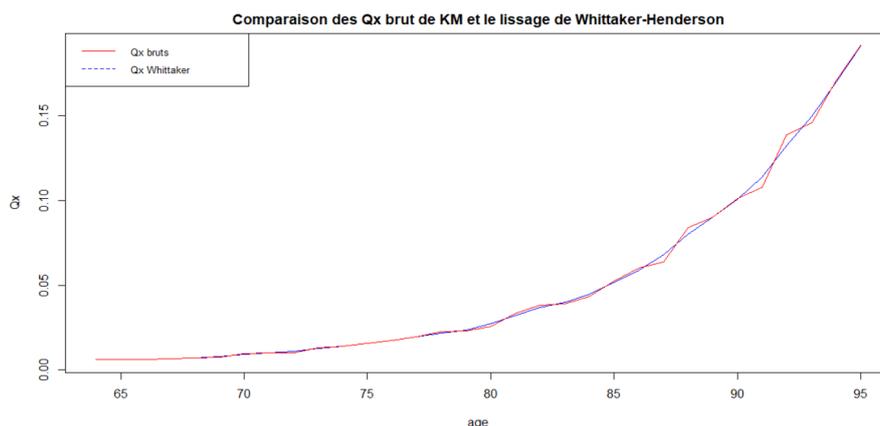


FIGURE 68 – Les taux unisexe lissés par la méthode de Whittaker-Henderson

La méthode de Whittaker lisse bien nos taux et est presque confondue avec les Qx de Kaplan-Meier jusqu'à l'âge de 88 ans, au delà les taux lissés s'écartent un peu des taux bruts. Ceci s'explique par la forte fluctuation aux grands âges. La méthode joue entre le critère de fidélité et de régularité par le biais du coefficient h . Avant toute conclusion, nous allons étudier un lissage plus simple appelé lissage géométrique.

2.4.2 Moyenne Géométrique

Le deuxième lissage que nous allons effectuer est un lissage simple appelé lissage géométrique. Il s'agit de remplacer chaque valeur par une moyenne géométrique sur intervalle d'âges. Nous avons choisi une intervalle de 5ans avec les données des deux années précédentes et des 2 suivantes. Cette méthode ressemble à la méthode de Moyenne Mobile, la différence est, nous avons une moyenne géométrique au lieu d'une moyenne arithmétique. La formule est la suivante :

$$q_x = \sqrt[5]{q_{x-2} * q_{x-1} * q_x * q_{x+1} * q_{x+2}}$$

L'avantage de cette opération est de lisser les \hat{q}_x avec une méthode très facile à mettre en place mais un des désavantages de cette méthode est que si un des \hat{q}_x précédents ou suivants est nul alors le nouveau \hat{q}_x sera nul. On risque de perdre quelques années d'observation aux extrémités ou au centre si on a un « trou » d'observation. Nous avons lissé les q_x donnés par Kaplan-Meier.

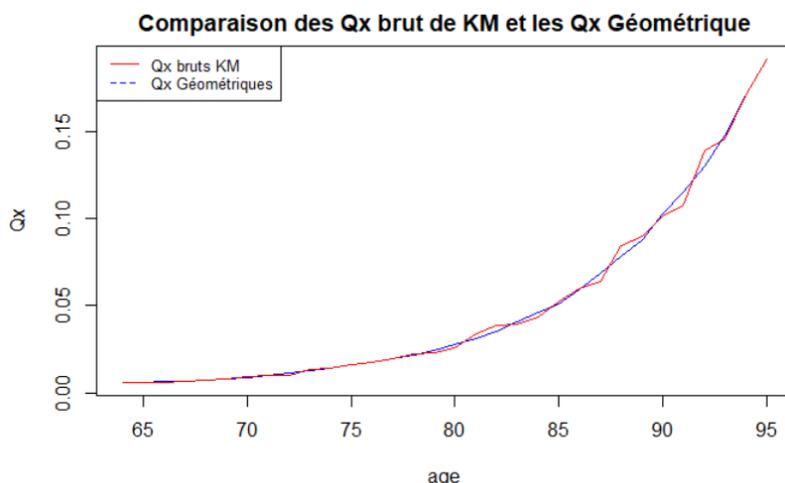


FIGURE 69 – Les taux lissés par le lissage Géométrique

Nous voyons bien que le lissage a permis de lisser les \hat{q}_x et on observe également une perte de données pour les âges supérieurs comme l'on s'y attendait. Plus les taux bruts sont volatils (faiblesse du nombre de données dans l'exposition) et moins le lissage semble bien représenter l'expérience. La figure suivante compare les q_x obtenus avec les deux méthodes.

Le lissage par Whittaker-Henderson permet d'obtenir une meilleure estimation des q_x là où l'exposition est grande car elle va lisser la courbe avec plus de liberté que la méthode géométrique mais le problème se pose quand on arrive aux q_x volatils. En effet, le critère de fidélité va faire « suivre » à Whittaker-Henderson sa volatilité et son non-sens aux âges élevés.

Cependant, les modèles de lissages ci-dessus ne nous permettent pas de corriger certaines incohérences liées à nos données car ces méthodes sont non paramétriques et très fidèles aux taux bruts. Par exemple, imaginons que nous allons utiliser les q_x lissés par la méthode Géométrique pour construire les provisions mathématiques de notre portefeuille retraite. Chaque q_x peut être vu comme la valeur actuelle probable (VAP) du versement de 1€ de

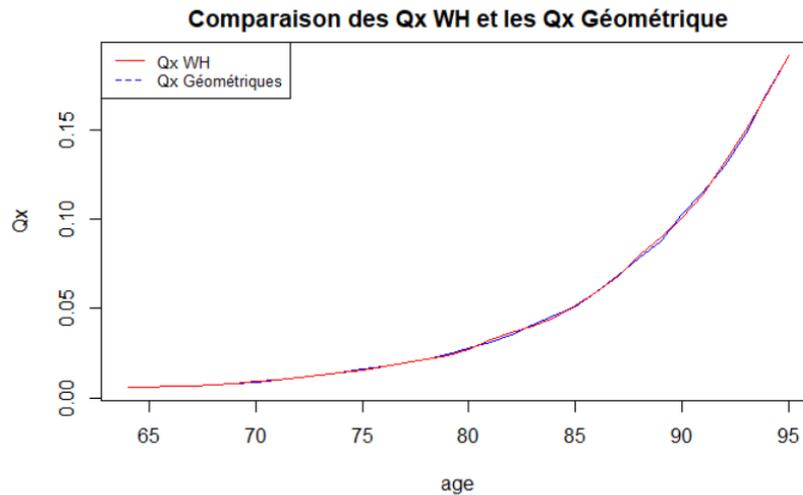


FIGURE 70 – Les taux lissés par le lissage Géométrique

rente annuelle pour un assuré d'âge x . Dans ce cas, Regardons les âges 65 et 66, les valeurs correspondantes sont $q_{65} = 0.006071089$ et $q_{66} = 0.006051393$. Nous constatons que, si nous arrêtons la simple applications de la méthode Géométrique, on provisionne plus cher pour une personne d'âge 66 qu'une personne d'âge 65 alors que cette dernière est plus jeune. Ceci, demande une explication que nous sommes pas en mesure de donner. Il est donc important, pour des questions de monotonie, d'étudier les modèles paramétriques.

2.5 Ajustements des taux par des modèles paramétriques

Ici, la modélisation donne la forme de la fonction de lissage. Nous allons présenter donc deux méthodes d'ajustements, le modèle de Makeham et le modèle de Tchatcher.

2.5.1 Ajustement de Makeham

Le modèle de Makeham repose sur l'hypothèse fondamentale que la fonction de risque instantané est de la forme :

$$\mu_x = \alpha + \beta * c^x$$

avec $\alpha > 0, \beta > 0, c > 1, .$ Cette formule peut s'interpréter comme une décomposition de la mortalité de la façon suivante :

- un paramètre α , indépendant de l'âge et faisant référence à la mortalité accidentelle.
- un facteur $\beta * c^x$, qui croît exponentiellement avec l'âge. Il s'agit du facteur modélisant le vieillissement de la population.

À partir de la formule précédente, nous obtenons :

$$p_x = \exp\left(-\int_x^{x+1} \mu_y * dy\right) = \exp(-\alpha) \exp\left(-\frac{b}{\ln c} * c^x (c-1)\right)$$

Avec p_x la probabilité de survie à l'âge x

En posant $s = \exp(-\alpha)$ et $g = \exp\left(-\frac{b}{\ln c}\right)$ alors la fonction pour estimer l'ajustement des taux s'écrit comme suit :

$$q_x = 1 - p_x = 1 - sg^{c^x(c-1)}$$

A partir de cette équation, nous pouvons écrire

$$\ln(1 - q_x) = \ln(s) + c^x * (c-1) \ln(g)$$

De plus, étant donné que les q_x sont petits, nous pouvons approximer $\ln(1 - q_x)$ par $-q_x$. On obtient finalement :

$$-q_x = \ln(s) + c^x * (c-1) \ln(g)$$

Ce qui conduit à :

$$\ln(q_{x+1} - q_x) = x * \ln(c) + \ln[-(c-1)^2 \ln(g)]$$

En faisant l'hypothèse que les données suivent une loi de Makeham, les points $(x, \ln(q_{x+1} - q_x))$ sont alignés sur la droite de pente $\ln(c)$.

On note $\theta = (s, g, c)$ le vecteur des paramètres à estimer et $q_x(\theta) = 1 - sg^{c^x(c-1)}$ fonction de Makeham à ajuster. On cherche alors $\hat{\theta}$, le vecteur de paramètres donnant la fonction qui ajuste le mieux la courbe des taux bruts observés.

Nous procédons par une régression non linéaire sur \mathbb{R} pour estimer les paramètres du modèle. La propriété d'alignement des points évoquée précédemment, permet de déterminer les paramètres initiaux de l'algorithme de Gauss Newton : Les paramètres g et c sont déterminés par l'ordonnée à l'origine et la pente de la droite. Quant au paramètre s , il est donné par la formule $\ln(s) = \ln p_x - c^x * (c-1) \ln(g)$. Par ailleurs, le paramètre α , qui fait référence au décès accidentel est calculé par $\alpha = -\ln(s)$.

Le graphique 25 présente l'ajustement par la loi de Makeham, effectué sur la tranche d'âge 64-95 ans :

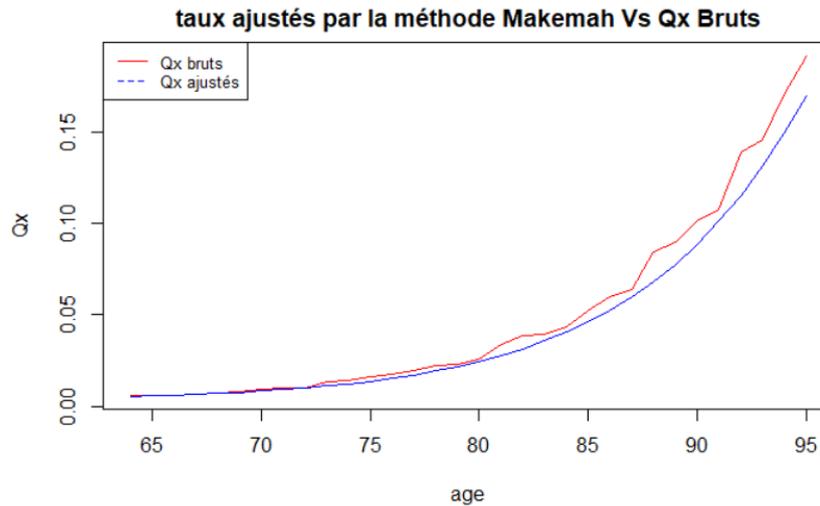


FIGURE 71 – Ajustement de Makeham

Nous constatons que le modèle de Makeham est bien régulier et respecte le critère de la monotonie. Cependant, les valeurs ajustées ne sont pas proches des q_x bruts. En effet, les taux de Makeham sont très faibles et touchent en quelques points la courbe des taux bruts entre es âges 64 et 72 ans. Ils sous-estiment donc, d’une façon exagéré, les taux bruts. Ceci s’explique par le fait que cette méthode suppose une forme du taux de risque instantané a priori, ce qui ne reflète pas la mortalité de notre portefeuille : En voulant être une courbe exponentielle, elle finit par éloigner des taux bruts d’expériences.

2.5.2 Ajustement de Tatcher

En pratique, le modèle de Makeham conduit à une sur-estimation des taux de décès conditionnels aux âges élevés. Afin de corriger cette surestimation TCHATCHER propose un modèle proche en posant

$$\mu_x = \frac{\alpha \exp(\beta x)}{1 + \alpha \exp(\beta x)} + \gamma.$$

Il reste à calculer $q_x = 1 - \exp\left(-\int_x^{x+1} \mu_y dy\right)$

qui conduit à :

$$q_x = 1 - \exp^{(-\gamma)} \left(\frac{\nu_{\alpha,\beta}(x+1)}{\nu_{\alpha,\beta}(x)} \right)^{-\frac{1}{\beta}}$$

avec $\nu_{\alpha,\beta}(u) = 1 + \alpha \exp(\beta u)$.

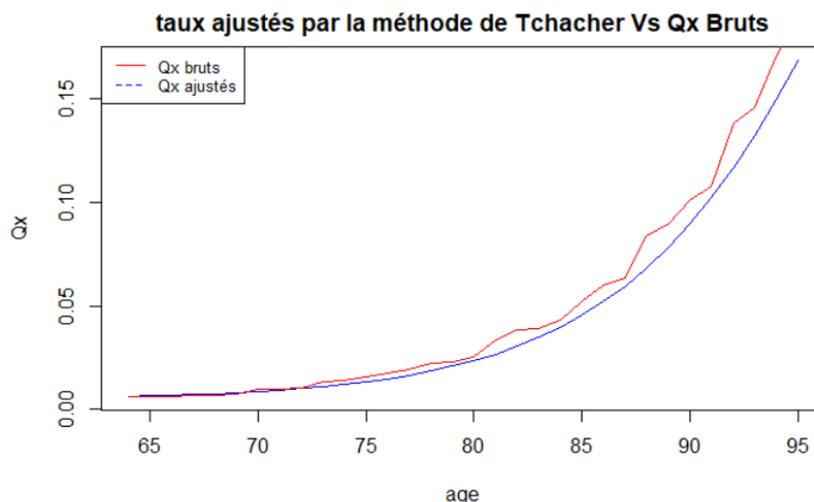


FIGURE 72 – Ajustement de Tchatcher

Nous obtenons des ajustements proches de ceux obtenus avec le modèle de Makeham, mais avec des taux légèrement plus faibles.

Comme la méthode de Makeham, Tchatcher sous-estime les taux bruts sauf que Tchatcher est faiblement plus proche des taux bruts.

Une fois le lissage effectué, il convient d'en valider la qualité. En effet, il s'agit de vérifier l'adéquation des taux ajustés aux taux bruts estimés. La section qui suit va se consacrer à la définition d'indicateurs statistiques capables d'apporter des réponses à cette problématique.

2.6 Choix du meilleur modèle

Nous avons mis en oeuvre plusieurs méthodes pour éliminer les irrégularités observées sur la courbe des taux bruts estimés. La question est de savoir à présent, laquelle de ces méthodes est la mieux adaptée pour nos données. Nous allons faire une synthèse des méthodes abordées afin d'en retenir une et, présenter par la suite des tests sur lesquels nous allons nous baser pour valider lissage adéquat.

— Standardized Mortality Ratio (*SMR*) par âge SMR_x

Il est nécessaire de choisir quel modèle est le plus approprié pour expliquer la mortalité de notre portefeuille. Un critère intuitif serait de comparer le nombre de décès par âge sur notre portefeuille, les décès réalisés donc, et les décès estimés selon les taux ajustés avec les modèles établis dans la section précédente. On retiendrait alors le(s) modèle(s) qui

prévoit(ent) un nombre de décès le plus proche du nombre de décès observés.

Une méthode intuitive consiste à comparer les rapports des nombres de décès. Ce ratio est appelé SMR_x Standardized Mortality Ratio par âge x .

A chaque âge x nous avons donc :

$$SMR_x = \frac{\hat{d}_x}{d_x} \quad \forall x \in [64 : 95]$$

avec \hat{d}_x les décès prédits avec les taux lissés q_x .

Naturellement, plus la moyenne de ce ratio est proche de 1, plus les décès calculés avec les taux ajustés sont proches des décès observés et donc plus la modélisation est proche des données brutes, en terme de nombre de décès. Si le SMR est supérieur à 1, notre modèle sous-estime le niveau de mortalité dans ce cas aussi la modélisation est satisfaisante s'il n'est pas éloigné de 1. En effet, ceci signifierait alors que nous observons plus de décès réels que de décès prédits, et donc que la modélisation est prudente.

D'autre part, nous allons étudier le nombre de fois où les décès prédits se trouvent dans l'intervalle de confiance à 95% des taux bruts de mortalités. Pour calculer cet intervalle de confiance, nous allons faire l'hypothèse que les décès suivent une loi binomiale (cf. Planchet et Thérond [2006]). Autrement dit, la variable aléatoire des taux décès notée Q_x suit une loi binomiale $B(L_x, q_x)$ pour tout âge $x \in [64 : 95]$, c'est à dire pour tout âge x vérifiant le critère de Cochran. On obtient donc que.

$$Q_x \hookrightarrow \mathcal{N}(q_x, \sigma^2 = \sqrt{\frac{q_x(1-q_x)}{L_x}})$$

Ce qui permet de définir l'intervalle de confiance :

$$IC_{95\%} = [q_x - u_\alpha \sqrt{\frac{q_x(1-q_x)}{L_x}}; q_x + u_\alpha \sqrt{\frac{q_x(1-q_x)}{L_x}}]$$

Où u_α est le quantile d'ordre α d'une gaussienne centrée réduite à un niveau de confiance égal à 95%. C'est à dire $u_{\frac{\alpha}{2}} = 1.96$.

Le tableau suivant résume les résultats pour chacun des modèles.

	Whittaker-Henderson	Makeham	Tchatcher	Lissage Géométrique
SMR moyen	102.9%	115.8%	114.8%	102.8%
σ moyen	0.021	0.21	0.03	0.020
Nb dans l'IC	30	14	17	31

TABLE 6 – SMR, σ , Nombre de fois dans l'IC pour chaque modèle

Tout d'abord, nous constatons que la moyenne des SMR est supérieure 1 pour chaque modèle, ils sous-estiment donc les taux bruts. D'autre part, les modèles de Whittaker-Henderson et Géométrique présentent les meilleurs SMR et plus de valeurs dans l'intervalle de confiance, contrairement aux modèles paramétriques qui sont très éloignés de la réalité. En effet, les méthodes de lissages ne supposent pas de loi a priori afin d'être le plus proche possible des données.

A la seule prise en compte de ces moyennes, la conclusion serait que la modélisation Géométrique est la plus proche de la mortalité réelle que Whittaker-Henderson. Ce qui n'est pas le cas. En effet, pour modéliser cette méthode, il a fallu prendre un intervalle de cinq années, et calculer la moyenne géométrique pour chaque âge, cela suppose donc les q_x des deux premiers âges (q_{64}, q_{65}) et les deux derniers (q_{94}, q_{95}) sont égaux aux q_x bruts correspondants. De ce fait, plus l'intervalle d'âge est grand plus le SMR est proche de 1 et donc plus la courbe est irrégulière. De plus, ne disposant que 32 âges, un ordre de 5 implique prendre en compte un nombre non négligeable d'observations.

En ce qui concerne l'ajustement par une loi de Makeham ou Tchatcher, nous remarquons que l'analyse en terme de nombre de présence dans l'intervalle de confiance n'est pas du tout satisfaisante. Par ailleurs, l'écart-type a une valeur très élevée comparé aux autres méthodes.

Les lissages obtenus par la méthode Géométrique et Whittaker-Henderson présentent des résultats similaires. Nous retiendrons le lissage par la méthode de Whittaker-Henderson. En effet, cette dernière, en plus d'avoir une SMR moyen satisfaisant en même temps elle est plus régulière, présente le plus faible écart type de SMR. Nous pouvons ajouter que retenir un lissage non paramétrique, nous évite d'être dépendant des paramètres de modèle (comme dans les méthodes paramétriques) et d'avoir un modèle sensible à leur calibrage.

2.6.1 test de changements des signes

Il s'agit d'un test non-paramétrique qui examine la fréquence de changements des signes de la différence entre les forces de mortalité observées et ajustées. Il est basé sur le signe que prend la différence : $d_x = \hat{q}_x - q_x$ où \hat{q}_x est le taux brut et q_x le taux lissé à l'âge x . Si les âges sont indépendants, la probabilité que d_x et d_{x+1} ne soient pas de même signe est égale à 1/2. On en déduit donc que la distribution de changements des signes suit une loi binomiale de paramètres $(p-1; 1/2)$ où p désigne le nombre d'observations sur la plage d'étude.

On définit donc le test suivant :

$$\begin{cases} H_0 : P(q_x - \hat{q}_x > 0) = P(q_x - \hat{q}_x < 0) = \frac{1}{2} \\ H_1 : P(q_x - \hat{q}_x > 0) \neq \frac{1}{2} \end{cases}$$

Autrement

(H0) : les d_x suivent une binomiale contre

(H1) : La loi des d_x est différente d'une binomiale, nous obtenons les résultats

	Whittaker-Henderson	Lissage Géométrique
p-value	100%	99.8%
Décision	Retient H_0	Retient H_0

TABLE 7 – Résultat test de changements des signes

Le test de changements des signes valide l'hypothèse H_0 pour les deux méthodes, avec une précision à 100% pour Whittaker-Henderson.

2.6.2 Calcul du SMR (Standardized Mortality Ratio)

Nous nous intéressons ici au SMR global en nombre. On cherche à savoir si la mortalité prédite est identique à la mortalité réellement observée. Pour cela nous calculons :

$$SMR = \frac{\sum_{i=x_{min}}^{x_{max}} D_i}{\sum_{i=x_{min}}^{x_{max}} q_i L_i}$$

où

- D_i le nombre de décès observés à l'âge i .
- q_i taux de décès à l'âge i lissé
- L_i Le nombre de personne sous risque

D_i le nombre de décès observés à l'âge i .

Nous obtenons pour Whittaker-Henderson les valeurs suivantes de SMR :

Méthode de Lissage	Whittaker-Henderson	Lissage Géométrique
SMR	102,7%	101,3%

TABLE 8 – Résultat test du SMR Globale

La valeur du SMR permet de dire que les taux lissés sont fidèles aux taux bruts estimés, pour la méthode de lissage de Whittaker-Henderson.

Conclusion

Le test des signes valide les deux méthodes de lissage. Le SMR globale est proche de 1 égale à 102,7%, donc cette méthode est prudente car elle sous-estime le niveau de mortalité réel.

Pour les raisons énoncées dans la section précédente, nous retenons le lissage par Whittaker-Henderson. Dans la suite les taux lissés considérés sont donc ceux obtenus par cette méthode. Nous avons à présent les taux bruts estimés, Il reste donc à compléter notre table de mortalité en construisant des taux bruts de mortalité aux grands âges.

2.7 Fermeture de la table

Les taux bruts estimés aux âges élevés ne sont pas fiables. Et pour cause, le faible effectif à ces âges, induit un taux de mortalité très élevé pour un seul décès observé. Pour compléter la table construite en dehors de la plage d'âge qui a été retenue pour la modélisation, nous allons procéder à une prolongation aux grands âges, à partir des taux lissés retenus. Il existe plusieurs méthodes pour extrapoler les taux de mortalité.

La plus utilisée est celle de Denuit et Goderniaux, car elle est à la fois simple à mettre en oeuvre et a des résultats satisfaisants. Nous allons, après avoir présenté la théorie et les hypothèses du modèle, en évaluer la performance en le testant sur les données réelles observées.

— Méthode de Denuit & Goderniaux

Cette approche utilise les logarithmes des taux de mortalité des âges les plus élevés et, repose sur l'ajustement par les moindres carrés ordinaires (MCO) le modèle suivant.

$\ln \hat{q}_x = a + bx + cx^2 + \epsilon_x$ avec ϵ_x variables aléatoires $iid \sim N(0, 1)$.

a, b, c sont donc des paramètres que nous allons déterminer.

Ce modèle pose deux contraintes fondamentales de fermeture de table :

$$q_{130} = 1$$

Cette première contrainte signifie que tout le monde est considéré décédé au delà de cet âge.

$$q'_{130} = 0$$

Avec q'_x la dérivée première de q_x par rapport à l'âge. Cette contrainte permet d'empêcher une éventuelle décroissance des taux de mortalité aux grands âges. En appliquant ces deux contraintes au modèle de base, nous arrivons aisément à la relation suivante :

$$a + bx + cx^2 = c(130^2 - 2 * 130x + x^2)$$

Nous avons donc

$$\ln(\hat{q}_x) = c(130^2 - 2 * 130x + x^2) + \epsilon_x$$

Les deux contraintes traduisent la concavité de la courbe des taux de mortalité aux grands âges.

Nous effectuons une régression linéaire par MCO de la dernière équation sur la tranche [64,95] ans, qui correspond au segment d'âges les plus élevés avec des taux fiables. Nous déterminons ainsi la valeur du paramètre c et nous en déduisons les valeur de a et b grâce à la relation $(a; b) = (c\hat{x}^2; -2c\hat{x})$ où \hat{x} désigne l'âge de clôture de la table qui est égal à 130 ans dans notre cas. On obtient les résultats suivants en ce qui concerne la régression :

\hat{x}	R^2_{ajust}	Valeur de c
130	99.73%	-0.001151782

TABLE 9 – Résultat de la regression linéaire

Nous avons déterminé précédemment la plage d'âge pour laquelle nos données étaient suffisantes et fiables. En pratique, nous conservons les taux lissés sur cette plage c'est à dire entre 64 et 95 ans. Ensuite de 96 à 130 ans, nous prenons les taux extrapolés.

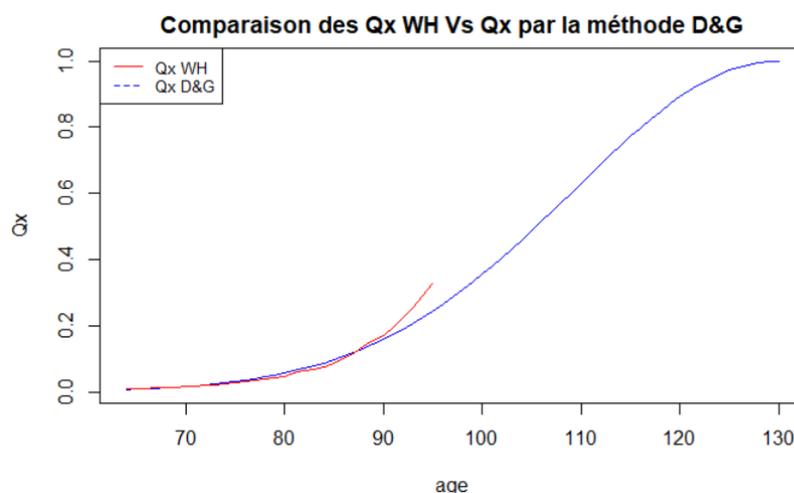


FIGURE 73 – les Qx lissés dans la plage d'âge 64 et 130 ans

Le SMR aux grands âges est calculé, en faisant le rapport entre les décès observés et les décès prédits par le modèle avec extrapolation, nous obtenons la valeur 102,7%. Nous

constatons alors que la prédiction est très proche de la réalité. On valide par ailleurs l'âge de clôture à 130 ans, ainsi que l'extrapolation par la méthode de Denuit & Goderniaux. Après avoir construit la table, la dernière étape consiste la validation globale de celle-ci.

2.8 Validation de la table globale

Afin de valider la table complète, nous allons étudier le positionnement de la table construite par rapport aux tables règlementaires. Le choix d'une table de référence n'est pas facile et nous avons deux possibilités. La première était d'utiliser une table certifiée en 2005 sur le portefeuille mais compte tenu de l'ancienneté de la certification nous préférons construire nous même une table de référence plus prudente. Nous avons donc choisi une table de référence en pondérant par le sexe-ratio les tables TGH TGF 05. . Ce que nous appelons sexe-ratio par âge est le rapport $\frac{Homme}{Homme+Femme}$ pour un âge donné en exposition ou en nombre selon notre choix. La construction de la table de référence est donc basée sur la formule.

$$tab_x^{ref} = TGH - 05_x * SexeRatio_x + TGF - 05_x * (1 - SexeRatio_x)$$

On remarquera pour l'instant qu'aucun abattement n'est appliqué. Une fois la table de référence construite, on va effectuer une régression linéaire sur les logit. Rappelons d'abord la définition du logit et la régression linéaire que nous allons effectuer.

$$logit(x) = \ln\left(\frac{x}{x+1}\right)$$

Nous faisons l'hypothèse qu'il existe une relation affine entre les logits de notre table et la table de référence. En résumé il va falloir estimer les coefficients a et b qui minimise l'erreur ϵ_x de l'équation du moindre carré suivante.

$$logit(q_x) = a * logit(q_x^{ref}) + b + \epsilon_x$$

Une fois que nous avons effectué la régression linéaire et ainsi obtenu les coefficients a et b , nous pouvons calculer les nouveaux q_x .

$$logit(q_x) = a * logit(q_x^{ref}) + b$$

Nous avons donc,

$$\begin{aligned} \frac{q_x}{1-q_x} &= \exp(b) * \left(\frac{q_x^{ref}}{1-q_x^{ref}}\right)^a \\ q_x * \left(1 + \exp(b) * \left(\frac{q_x^{ref}}{1-q_x^{ref}}\right)^a\right) &= \exp(b) * \left(\frac{q_x^{ref}}{1-q_x^{ref}}\right)^a \\ q_x &= \frac{\exp(b) * \left(\frac{q_x^{ref}}{1-q_x^{ref}}\right)^a}{1 + \exp(b) * \left(\frac{q_x^{ref}}{1-q_x^{ref}}\right)^a} \end{aligned}$$

La table suivante résume les résultats de cette régression.

	a	b	R^2 Ajusts
Valeur	0.99442	-0.55182	99,64%
P-value	< 2e-16	4.89e-16	< 2.2e-16

Les résultats présentés dans cette table nous permettent de conclure que tous les coefficients du modèle sont significatifs à 1%. Enfin, nous pouvons dire que nous avons une très bonne adéquation des logits d'expérience aux logits de référence compte tenue de la valeur élevée du coefficient de détermination R^2 ajusté.

Conclusion

Au vu de ces différents analyse ci-dessus, avec les différents indicateurs statistiques (calcul du SMR, test de changement de signe et le positionnement par rapport à la table réglemn-taire), nous avons pu construire une table du moment en réalisant plusieurs étapes. En premier lieu, il a été question d'estimer les taux de mortalité, ensuite il a fallu les fiabiliser par une méthode de lissage. Enfin, une méthode d'extrapolation a permis de reconstituer des taux significatifs aux âges élevés. Nous disposons à présent d'une table quasi complète avec des taux de mortalité pour une plage de [64 :130] ans.

Cependant, La table que nous venons de construire est une table statique, qui ne tient pas compte de l'évolution future de la mortalité. Pour prendre en compte une éventuelle dérive de mortalité dans le temps, nous allons construire dans la partie suivante une table prospective.

Quatrième partie

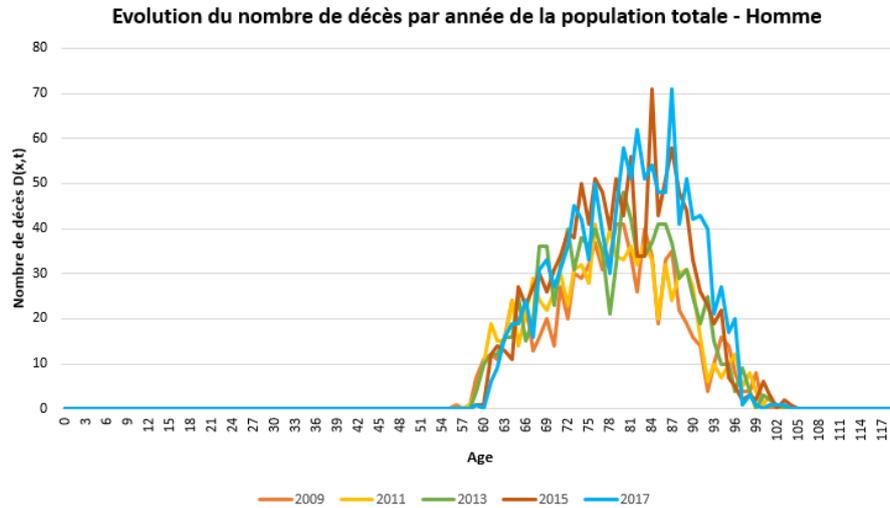
Construction d'une table de mortalité prospective

L'objectif des tables de mortalité prospectives est de tenir compte des évolutions futures de la mortalité, ce qui est la limite principale d'une table du moment. En effet, la table du moment obtenue dans la partie précédente a été construite en se basant sur un historique de 9 ans allant de 2009 à 2017. L'hypothèse sous-jacente était que sur cette période, nous avons une même génération d'individus. Autrement dit, dans notre modèle de la table du moment, la probabilité de décès pour une personne âgée de 70 ans en 2012, est exactement la même que pour une personne âgée de 70 ans en 2017. En réalité, une telle hypothèse n'est pas toujours vérifiée. Les méthodes usuelles cherchent tout d'abord à ajuster les tendances passées, puis à les extrapoler à l'avenir.

Idéalement, afin d'étudier la mortalité de façon précise, il faudrait pouvoir observer chaque individu sur la période allant de sa date de liquidation jusqu'à son décès. En pratique, cela n'est pas possible dans la mesure où le portefeuille n'est pas éteint. L'étude est donc réalisée sur une période d'observation définie. Sur cette période, il est nécessaire de définir les taux de mortalité observés qui s'appuient d'une part sur l'estimation de la population sous risque et d'autre part le niveau des décès. Il convient alors d'expliquer les données du portefeuille ainsi de définir les notions d'exposition sous risques et le nombre de décès.

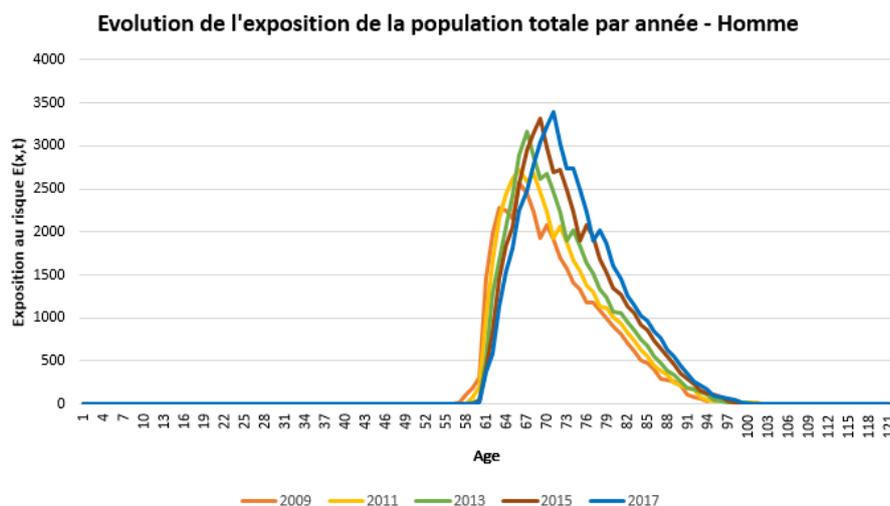
Dans la suite, nous disposons des données d'un portefeuille de rentiers sur un historique de 2009 à 2017, soit 9 ans. En outre, en suivant la même démarche qu'au chapitre précédent, nous arrivons à une plage d'âge [63,94] ans pour laquelle nous estimons que les taux bruts par âge pour chaque année calendaire considérée.

Le nombre de décès au cours d'une année calendaire t de personnes qui ont fêté leur x^{eme} anniversaire ainsi défini est noté $D(x, t)$. Nous faisons hypothèse classique (DELWARDE [2004]) que les décès sont répartis de façon uniforme entre deux âges entiers, c'est à dire qu'ils interviennent en moyenne au milieu de l'année.



L'estimateur du nombre moyen de personnes au cours de l'année t dont le dernier anniversaire est le x^{eme} est notée $E(x, t)$. Elle représente l'exposition au risque, et peut être définie en fonction du nombre de personnes d'âge x survivant en début de période, noté $L(x, t)$, qui est la statistique la plus généralement à disposition des actuaires. L'hypothèse d'uniformité des taux de décès permet d'écrire que :

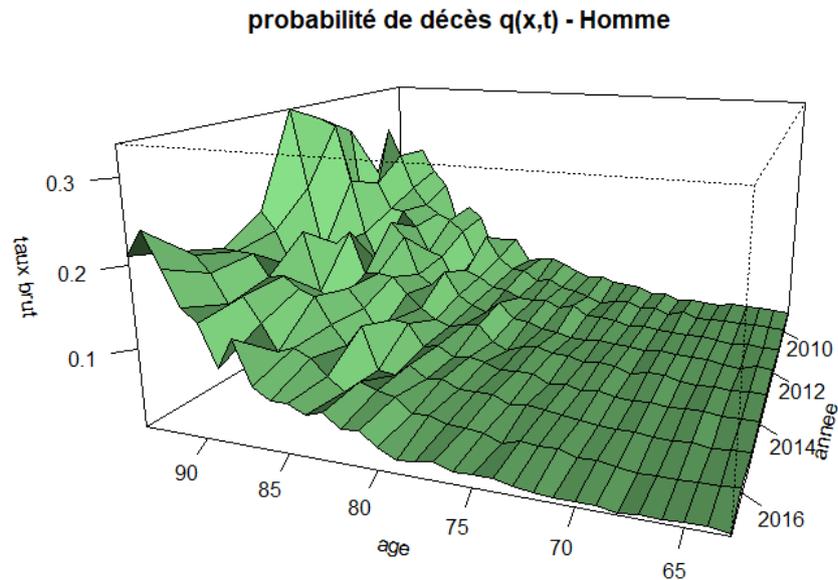
$$E(t, x) = \frac{L(t, x) + L(t + 1, x + 1)}{2} \quad (9)$$



On définit ensuite la probabilité pour un individu d'âge x de décéder durant l'année calendaire t , noté est définie par l'estimateur de Hoem. Il est noté :

$$\hat{q}_x(t) = \frac{D_{x,t}}{E_{x,t}} \quad (10)$$

la valeur $\hat{p}_x(t) = 1 - \hat{q}_x(t)$ est la probabilité de survie correspondante.



Les trois graphiques mettent en évidence des irrégularités en fonction des années mais également en fonction des âges. Aux grands âges, la volatilité des taux de mortalité augmente. En effet, La population sous risque diminue et en même temps le nombre de décès augmente ce qui provoque des fluctuations d'échantillonnage importantes. De plus, d'une année à l'autre, et pour les âges les plus élevés, l'exposition au risque peut être significativement différente.

L'objectif de ce chapitre est donc de faire une analyse prospective de la mortalité afin de pouvoir anticiper une éventuelle dérive dans le temps. Pour cela, nous allons dans un premier temps, mettre en oeuvre plusieurs méthodes d'ajustement prospective, ensuite nous ferons une analyse comparative de ces différentes méthodes afin de ne retenir que celle qui correspond statistiquement le mieux, à nos données réelles.

1 Ajustement par Lee Carter

Il s'agit d'une méthode d'extrapolation des tendances passées initialement utilisée sur des données américaines, qui est devenue rapidement un standard (voir l'article original LEE et CARTER [1992]).⁷ Ce modèle est la référence standard en ce qui concerne la construction des tables prospectives de mortalité. Plusieurs modèles modernes, comme

7. Pour plus détaillé, nous invitons au lecteur de consulter *freakonometrics*

nous allons le voir dans la suite, sont en réalité des déclinaisons de ce modèle de base.

L'idée fondamentale de ce modèle est d'ajuster la série des logarithmes des taux instantanés de décès par une fonction paramétrique à laquelle se rajoute un phénomène aléatoire non capté par le modèle. La modélisation retenue pour le taux instantané de mortalité est la suivante :

$$\ln \mu_{xt} = \alpha_x + \beta_x k_t, \quad (11)$$

ou, de manière équivalente,

$$\ln \mu_{xt}^* = \alpha_x + \beta_x k_t + \varepsilon_{xt} \quad (12)$$

Avec l'hypothèse $\varepsilon_{x,t} \sim N(0, \sigma^2)$ i.i.d.

L'équation 11 peut s'interpréter de la manière suivante :

- α_x est la composante liée à l'âge. Il s'interprète comme la valeur moyenne des $\ln(\hat{\mu}_{x,t})$ au cours du temps.
- De l'équation 11, il est possible de vérifier aisément que $\frac{d \ln \hat{\mu}_{x,t}}{dt} = \beta_x \frac{d \kappa_t}{dt} \therefore$ On peut déduire alors que le paramètre β_x traduit comme la sensibilité de la mortalité instantanée à l'âge x , par rapport à l'évolution générale de la mortalité κ_t . On aura en particulier, pour les âges qui présentent un β_x élevé, une sensibilité plus grande à l'évolution de la mortalité.
- Le paramètre κ_t est la composante temporelle, il traduit le vieillissement de la population. Son analyse permet ainsi de détecter une éventuelle dérive de mortalité.

En particulier, le modèle de Lee-Carter suppose la constance au cours du temps de cette sensibilité. Cette contrainte du modèle peut apparaître relativement forte :

- Pour tout âge x les quotients des variations relatives des taux de mortalité à des dates différentes ne dépendent pas de l'âge x . Si la variation relative du taux de mortalité à 60 ans en 2000 était 70% de ce qu'elle était en 1990 ce coefficient de 70% est retenu pour tous les âges ;
- Pour une même date t les quotients des variations relatives des taux de mortalité à des âges différents ne dépendent pas de la date t . Si en 2000 la variation relative du taux de mortalité à 35 ans est 50% de la variation relative du taux à 50 ans ce coefficient de 50% s'appliquera à toute date future ou passée.

Enfin, on peut remarquer que la forme du modèle implique l'homoscédasticité des taux de mortalité, ce qui est manifestement faux en pratique. Cet inconvénient sera examiné plus en détails lors de la phase de validation de notre modèle.

De plus, nous pouvons remarquer que le modèle n'est pas identifiable. En effet, en remplaçant par exemple α_x par $\alpha_x - \beta_x$, et κ_t par $1 + \kappa_t$ dans l'équation 12, l'équation reste inchangée et on obtient ainsi exactement le même modèle. D'une manière générale le modèle est invariant par les transformations suivantes (c constante non nulle).

$$\begin{aligned} (\alpha_x, \beta_x, k_t) &\rightarrow \left(\alpha_x, \frac{\beta_x}{c}, c \times k_t\right) \\ (\alpha_x, \beta_x, k_t) &\rightarrow (\alpha_x - c\beta_x, \beta_x, k_t + c) \end{aligned} \quad (13)$$

Afin de rendre le modèle identifiable, il convient alors d'ajouter des contraintes sur les paramètres.

$$\begin{aligned} \sum_{x=x_m}^{x_M} \beta_x &= 1 \\ \sum_{t=t_m}^{t_M} \kappa_t &= 0 \end{aligned} \quad (14)$$

Une fois le modèle défini, le critère d'optimisation est la maximisation de la variance expliquée du modèle. Ce qui équivaut à minimiser la variance des erreurs du modèle. Par la méthode des moindres carrés non linéaires, le problème d'optimisation à résoudre sous les contraintes d'identifiabilité est donc :

$$\left(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\right) = \arg \min_{\alpha, \beta, \kappa} \sum_{x,t} (\ln \hat{\mu}_{x,t} - \alpha_x - \beta_x \kappa_t)^2 \quad (15)$$

Il convient donc de résoudre ce programme d'optimisation, sous les contraintes d'identifiabilité. Le nombre de paramètres à estimer est élevé, il est égal à $2 \times (x_M - x_m + 1) + t_M - t_m - 1$

1.1 Estimation des paramètres

En dérivant l'équation 15 par rapport à α_x d'une part, et en tenant compte de la contrainte d'identifiabilité 14 d'autre part, nous obtenons que :

$$\frac{\partial}{\partial \alpha_x} \sum_{y,t} (\ln \mu_{yt}^* - \alpha_y - \beta_y k_t)^2 = 2(t_M - t_m + 1) \alpha_x - 2 \sum_{t=t_m}^{t_M} (\ln \mu_{xt}^* - \beta_x k_t), \quad (16)$$

on trouve en tenant compte de la contrainte $\sum_{t=t_m}^{t_M} k_t = 0$ que :

$$\hat{\alpha}_x = \frac{1}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \ln \mu_{xt}^* \quad (17)$$

En d'autres termes, $\hat{\alpha}_x$ est la moyenne temporelle, à l'âge x , des taux instantanés de décès (sur l'échelle logarithmique). On considère alors la matrice $Z = (z_{xt})$ des taux centrés par rapport à la dimension temporelle :

$$z_{xt} = \ln \mu_{xt}^* - \hat{\alpha}_x \quad (18)$$

Z est une matrice de dimension $(x_M - x_m + 1, t_M - t_m + 1)$. La forme du modèle revient à chercher une approximation de Z en produit de 2 vecteurs de la forme $Z \approx \hat{\beta} \hat{k}'$, de sorte que la décomposition soit optimale au sens du critère des moindres carrés, i.e. explique la plus grande part possible de la variance totale. La résolution de ce problème passe par la décomposition en valeurs propres de la matrice Z , que l'on met sous la forme :

$$Z = \sum_{i \geq 1} \sqrt{\lambda_i} v_i u_i' \quad (19)$$

avec $\lambda_1 \geq \lambda_2 \dots \geq 0$ les valeurs propres de $Z'Z$, u_i le vecteur propre normé de ZZ' associé à λ_i , et v_i le vecteur propre associé à la même valeur propre pour ZZ' . Ceci est justifié par le fait que, comme $Z'Z u_i = \lambda_i u_i$, on a $ZZ' (Z u_i) = \lambda_i Z u_i$, et donc les deux matrices transposées ont les mêmes valeurs propres avec des ordres de multiplicité identiques. De plus, si u_i est un vecteur propre de $Z'Z$ alors $Z u_i$ est un vecteur propre de ZZ' associé à la même valeur propre. Si on pose $v_i = \frac{1}{\sqrt{\lambda_i}} Z u_i$, on voit que

$Z u_i u_i' = \sqrt{\lambda_i} v_i u_i'$, ce qui en sommant et en tenant compte de l'orthonormalité des vecteurs propres, conduit à la décomposition de Z . On est ainsi conduit à proposer comme approximation $Z \approx \sqrt{\lambda_1} v_1 u_1'$, avec comme mesure de la qualité de cette approximation la part d'inertie expliquée, $\frac{\lambda_1}{\sum \lambda_i}$. On obtient finalement les estimateurs de β et k suivants :

$$\hat{\beta} = \frac{1}{\sum v_{1j}} v_1 \quad (20)$$

et

$$\hat{k} = \sqrt{\lambda_1} \sum v_{1j} u_1 \quad (21)$$

L'objectif est d'utiliser les résultats de cet ajustement pour extrapoler les taux de mortalité pour $t > t_M$; l'idée est d'analyser la série des (\hat{k}_t) , qui capture l'information sur l'évolution temporelle de ces taux pour lui ajuster un modèle de type ARIMA.

À ce stade on dispose d'une première estimation des paramètres du modèle; toutefois, si on se trouve dans une situation dans laquelle l'effectif soumis au risque est très important, on peut se dire que les fluctuations d'échantillonnage sur le nombre total de décès par âge

doivent être très faibles. Il apparaît alors souhaitable d'ajuster auparavant les paramètres du modèle pour que le nombre de décès prévus par le modèle chaque année soit égal au nombre de décès observés. Comme l'exposition au risque est $L_{xt} = \frac{D_{xt}}{\mu_{xt}^*}$

avec $D_{xt} = D_{xt}(t-x) + D_{xt}(t-x-1)$, cette contrainte s'exprime par :

$$\sum_{x=x_m}^{x_M} D_{xt} = \sum_{x=x_m}^{x_M} L_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{k}_t) \quad (22)$$

la variable étant \hat{k}_t . On introduit la fonction $F(k) = \sum_{x=x_m}^{x_M} L_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x k) - \sum_{x=x_m}^{x_M} D_{xt}$, de sorte que la contrainte ci-dessus s'exprime par $F(k) = 0$. La forme de la fonction F assure l'unicité de la racine si elle existe. La recherche de la racine peut se faire par un algorithme de type Newton-Raphson, en posant :

$$k_{i+1} = k_i - \frac{F(k_i)}{F'(k_i)} \quad (23)$$

avec la valeur initiale $k_0 = \hat{k}_t$ et le critère d'arrêt $\frac{k_{i+1} - k_i}{k_i} \leq \varepsilon$ en prenant par exemple $\varepsilon = 10^{-7}$. On obtient ainsi un nouvel estimateur \hat{k}_t . Mais la série (\hat{k}_t) ainsi ajustée doit être corrigée pour respecter la contrainte d'identifiabilité $\sum_{t=t_m}^{t_M} k_t = 0$, ce qui conduit à poser :

$$k_t^* = \hat{k}_t - \frac{1}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \hat{k}_t \quad (24)$$

L'allure « typique » des paramètres obtenus est la suivante, tout d'abord pour les paramètres en fonction de l'âge⁸ :

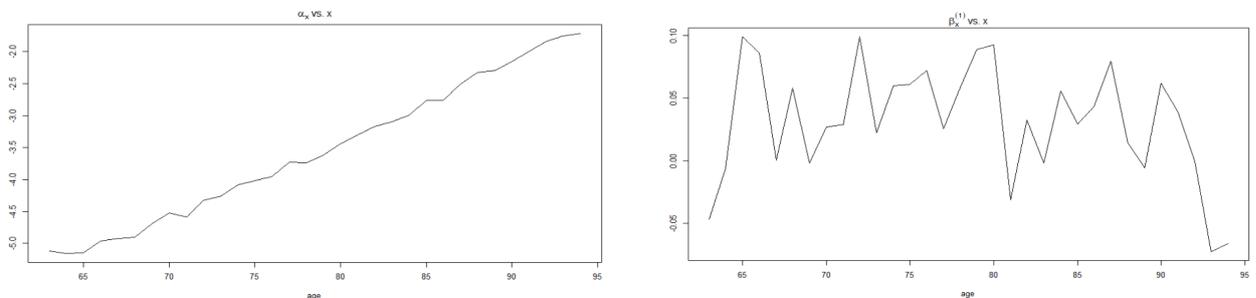


FIGURE 74 – Paramètres du modèle Lee Carter : coefficient α_x à gauche et coefficient α_x à droite

8. Nous avons utilisé le package StMomo pour estimer les paramètres : voir *stmomo*

puis la composante temporelle :

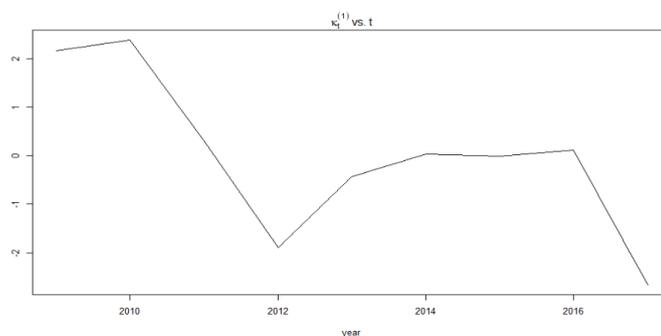


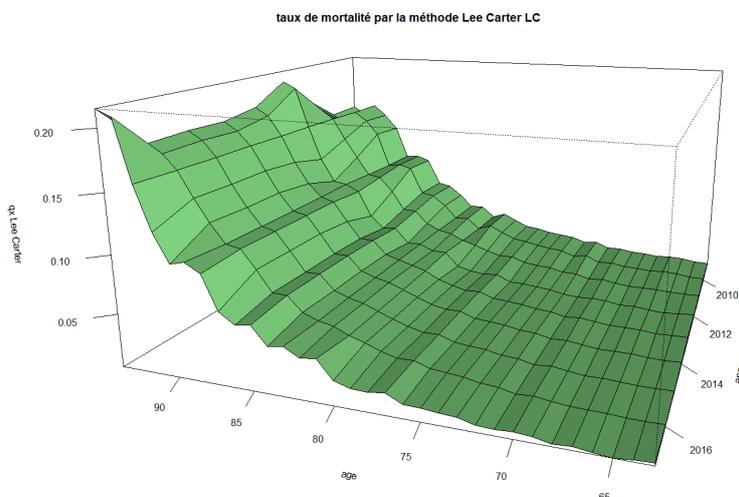
FIGURE 75 – Evolution de la tendance temporelle

1.2 Résultat du modèle Lee Carter

Rappelons que l'objectif est résoudre l'équation du moindre carré afin de réduire le bruit blanc ϵ_t . ainsi nous obtenons un taux instantané $\hat{\mu}_{xt}$:

$$\ln \hat{\mu}_{xt} = \hat{\alpha}_x + \hat{\beta}_x \hat{k}_t, \quad (25)$$

Avec les paramètres estimés (ci-dessus) et la série (κ_t) modélisée, nous obtenons les résultats de notre modélisation sur la plage [63; 94] ans. Les taux obtenus peuvent être alors projetés sur un horizon $h = 40$ ans par exemple. .



Rappelons que nous sommes dans un contexte où les effectifs (environ 100000 lignes) sont relativement faibles, il subsiste donc des fluctuations d'échantillonnage, qui conduisent

à des irrégularités sur les résultats. La surface ci-dessus est irrégulière dans la dimension temporelle et dans la dimension âge. En effet, nous avons modélisé le paramètre (κ_t), par un processus ARIMA, pour projeter la mortalité dans le temps. Cependant, dans notre étude, il s'avère que nous ne disposons pas d'un historique conséquent pour mettre en oeuvre un ARIMA. De plus, nous remarquons d'après les figures 58 et 59 une allure irrégulière des courbes α_x et β_x .

Afin de remédier cela, nous proposons alors de diminuer le nombre de paramètres en recalibrant le modèle.

1.3 Recalibrage des coefficients du modèle

On propose de faire une régression linéaire en ce qui concerne les α_x et une régression d'ordre 3 pour les β_x .⁹

$$\alpha_x^* = ax + b \quad (26)$$

$$\beta_x = a_1 + b_1x + c_1x^2 + d_1x^3. \quad (27)$$

La table suivante présente les résultats obtenus en modélisant la série (α_x) par une régression linéaire :

	a	b	Modèle
Valeur	0.1183	-12.8428	R^2 ajusté = 99.35%
P-value	2.2×10^{-16}	2.2×10^{-16}	2.2×10^{-16}

On obtient une modélisation satisfaisante de $\alpha(x)$. En effet, la valeur du R^2 ajusté est assez proche de 1 et les paramètres sont significatifs à 1%. De plus, nous obtenons des p-value très faibles pour les paramètres et le modèle global. Par contre la modélisation de β_x est assez satisfaisante avec un R^2 ajusté d'environ de 63,5.

Il reste alors à modéliser la série (k_t^*) pour extrapoler les taux futurs, pour cela, nous avons utilisé ci-dessous un modèle ARIMA, dans notre étude, il s'avère que nous ne disposons pas d'un historique conséquent pour mettre en oeuvre un ARIMA. Toutefois, compte

9. Pour revoir le reparamétrage des coefficients de Lee Carter, vous pouvez consulter le site de <http://www.planchet.net>

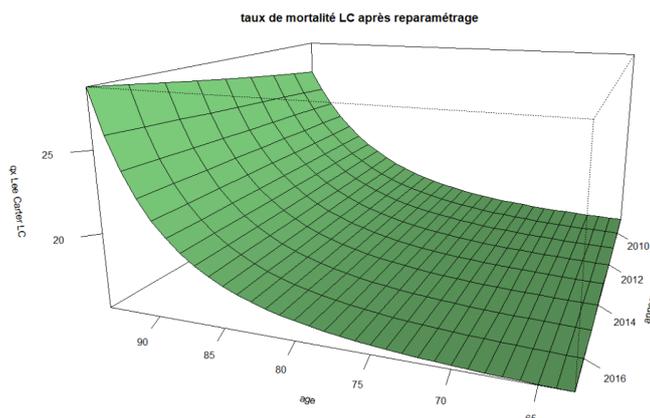
tenu de l'allure du graphe ci-dessus, la modélisation la plus simple que l'on puisse imaginer est par exemple une régression linéaire de ces coefficients :

$$k_t^* = at + b + \varepsilon_t \quad (28)$$

avec $\varepsilon_t \sim N(0, \sigma^2)$ *i.i.d.*

On obtient une modélisation assez satisfaisante de κ_t . En effet, la valeur du R^2 ajusté est de 70%. La significativité de a nous indique que le paramètre κ_t décroît avec le temps. On en déduit une dérive de mortalité dans notre portefeuille.

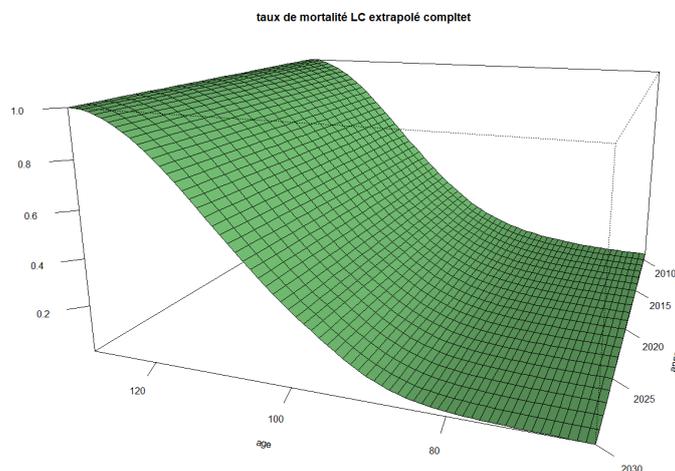
Une fois les coefficient ($\kappa_t, \alpha_x, \beta x$) recalibrés, nous obtenons les résultat de notre modélisation sur la plage [63; 94] ans.



1.4 Fermeture de la table

L'estimation des paramètres du modèle de Lee-Carter nécessite que l'on dispose d'une matrice rectangulaire complète de taux de décès (μ_{xt}^*). En pratique, les valeurs brutes estimées présentent une grande instabilité aux âges élevés, du fait du faible exposition au risque mais, il peut arriver aussi que les données ne soient plus disponibles au-delà d'un âge limite. Plusieurs méthodes existent pour compléter la table obtenue en extrapolant les taux de mortalité aux grands âges. Pour ce faire, il suffit pour chaque année calendaire t , d'exécuter la méthode de Denuit & Goderniaux vue dans la partie précédente. Cette méthodologie ayant été étudiée lors de la construction de la table du moment, nous faisons le choix de la reprendre pour chaque ajustement prospectif mis en œuvre.

Ainsi, nous abtenons les résultat de notre modélisation sur la plage [63; 130] ans et projetés jusqu'à 2030.



1.5 Validation du modèle

Nous allons dans cette partie, analyser la performance du modèle de Lee-Carter appliqué à nos données. Dans un premier temps, nous allons faire une analyse statistique de la qualité de l'ajustement obtenu et ensuite nous allons mettre en relief les limites du modèle.

1.5.1 Analyse par espérance de vie générationnelle

Le principe ici est d'observer les espérances de vie générationnelles durant 40 ans pour des individus d'âge x en 2017. Ceci permet d'observer l'évolution de la tendance de mortalité entre les trois modèles. L'espérance de vie générationnelle s'appuie sur la lecture longitudinale de la table construite et s'exprime de la façon suivante :

$$\omega e_x^{\rightarrow}(t) = \sum_{x=1}^{\omega} \prod_{j=0}^{x-1} (1 - q_{x+t}(t+j)) \quad (29)$$

Il s'agit de l'espérance de vie partielle sur ω années d'un individu âgé de x années à la date t .

Nous avons calculé les espérances de vie générationnelles deux fois : en 2015 et en 2017, ensuite, pour chacune de ces dates nous avons, l'espérance de vie générationnelle sur les taux brutes et sur les taux ajustés par LC.

Sur les graphiques précédents nous remarquons que les tendances entre les espérance de vie générationnelle du modèle de Lee Carter et les données brutes se ressemblent. Le modèle de LC est inférieur sur toute la première partie, puis il y a croisement pour $x = 75$

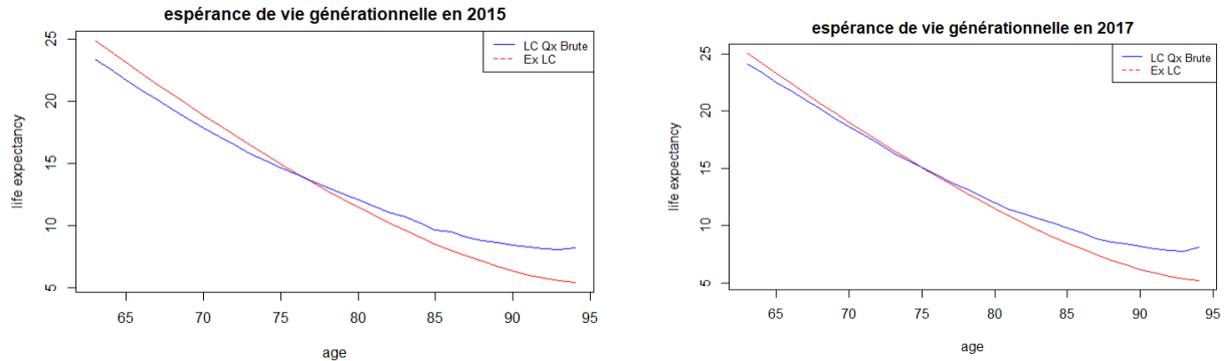


FIGURE 76 – espérance de vie générationnelle calculées en 2015 et 2017 avec les données brutes et les données ajustées selon Lee-carter

ans. L'espérance de vie a tendance à diminuer avec l'âge.

Toutefois, il est difficile de conclure après une analyse graphique des espérances de vie générationnelles calculées. En se basant sur ce critère, nous ne pouvons donc pas conclure à une fidélité rigoureuse du modèle aux données réelles. Elle permet de mesurer la mortalité par âge en prenant en compte l'évolution temporelle. Cependant, il faut avoir une table finie pour pouvoir la calculer. Il faut que la largeur de la table soit d'au moins $\omega - x_{\min} + 1$. Les données brutes dont nous disposons ne sont observables que sur la période 2009 – 2017, il est donc impossible de calculer une espérance de vie générationnelle « complète ».

Pour obtenir une indication pour la validation du modèle, l'idée pourrait être de calculer ce que l'on appelle une espérance de vie générationnelle partielle « maximale »¹⁰, noté EVGP, c'est-à-dire sur 9 années (car $2017 - 2009 + 1 = 9$), pour tout âge x acceptable. Nous allons donc calculer le EVGP sur la plage d'âge $[63, 82]$, à l'aide de la formule suivante.

$$EVGP_x = \sum_{h=1}^{2017-2009+1} \prod_{j=0}^{h-1} (1 - q_{x+j,2009+j}) = \sum_{h=1}^9 \prod_{j=0}^{h-1} (1 - q_{x+j,2009+j}) \text{ pour } x \in [63; 82] \quad (30)$$

ainsi, nous obtenons sur la plage d'âge $[63, 94]$ le graphique suivant.

On remarque toujours que les tendances entre les EVGP de Lee Carter des données brutes se ressemblent. Contrairement à l'espérance de vie générationnelle, avec l'EVGP on constate que la différence est beaucoup plus pointue.

10. Pour plus de détails : nous invitons au lecteur de consulter le site de ressources actuarielles

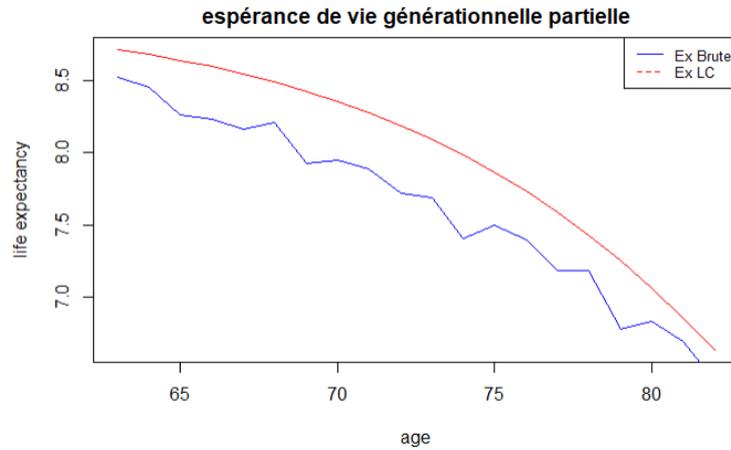


FIGURE 77 – Espérances de vie générationnelles partielles des taux observés et des taux estimées calculées sur la période d’observation 2009 - 2017

Toutefois, il semble être difficile de conclure après une analyse graphique des EVGP calculées. Il est donc indispensable de trouver un autre critère pour conclure rigoureusement concernant la fidélité du modèle aux données réelle.

Nous regardons alors l’erreur relative entre les probabilités de décès observées \hat{q}_x avec celles estimées par LC q_x pour chaque année t . l’erreur relative est calculée comme suit

$$\xi_x = \frac{\hat{q}_x - q_x}{\hat{q}_x} \quad (31)$$

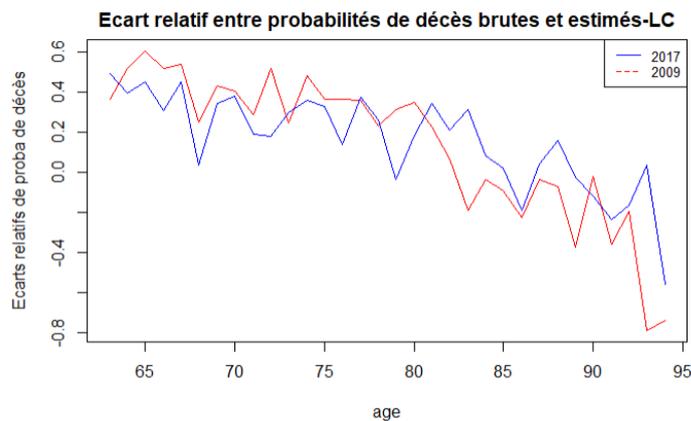


FIGURE 78 – Erreur relative entre probabilités de décès observées et ajustées entre 63 et 94 ans en 2009 et 2017

Les écarts relatifs observés sont assez volatils et nous remarquons qu’ils sont moins

importants avec les probabilités de décès ajustées. En se fixant un seuil de 5%, nous ne pouvons pas valider le modèle au regard des erreurs relatives des taux de mortalité. Cependant, ce résultat est à considérer avec prudence car nous sommes sur une plage d'âge réduite et de fait, nous utilisons peu de données pour ce calcul.

1.6 Intervalle de confiance

Nous avons vu en debut de chapitre que sur la classe d'âge [63, 94] le nombre de données utilisées est satisfaisant étant donné que :

$$E_{x,t} * \hat{q}_{x,t}(t) > 5 \text{ et } E_{x,t} * (1 - \hat{q}_x(t)) > 5 \quad (32)$$

L'exposition au risque est grand alors, nous pouvons ainsi approximer les taux de mortalité par une loi normale de paramètres :

$$Q_x \sim N(q_x; \sigma_x) \quad (33)$$

où $\sigma_x = \sqrt{\frac{q_x(1-q_x)}{E_x}}$ et obtenons l'intervalle de confiance suivant

$$\hat{q}_x(t) \pm u_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x(t) (1 - \hat{q}_x(t))}{E_{x,t}}} \quad (34)$$

Ou plutôt avec les décès prédit, nous obtenons.

$$IC_{\alpha} = E_{x,t} \cdot \left[\hat{q}_x \pm u_{\frac{\alpha}{2}} \frac{\sqrt{\hat{q}_x (1 - \hat{q}_x)}}{\sqrt{E_{x,t}}} \right] \quad (35)$$

où $u_{\frac{\alpha}{2}}$ désigne le quantile $\alpha/2$ de la loi Normale centrée réduite et $1 - \alpha$ désigne le niveau de confiance.

Avec un niveau de confiance de 95%, nous obtenons les intervalles de confiances pour les années 2009 à 2017 : $u_{\frac{\alpha}{2}} = 1.96$. Pour plus de lisibilité des résultats, nous nous concentrons seulement sur l'années 2009.

Pour avoir plus de visibilités, nous avons représenté les décès prédits sur la plage d'âge [63, 94]. Nous remarquons que les décès observés sont quasiment tous dans l'intervalle de confiance à partir des taux estimés par le modèle. Nous remarquons, par ailleurs, les grands âges sont hors de l'intervalle de confiance à cause du faible effectifs.

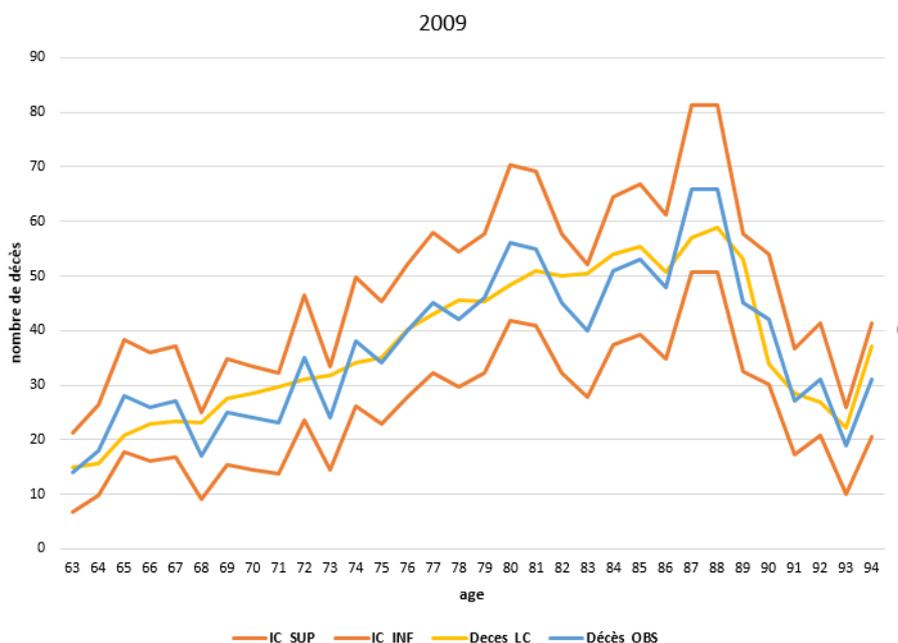


FIGURE 79 – Décès prédit par LC et décès observés sur notre portefeuille et Intervalles de confiance - en année 2009

On peut conclure que avec la première approche (sans reparamétrage des coefficients), nous remarquons que le lissage réalisé par la méthode de Lee-Carter n'est pas optimal. Ceci est dû à la fluctuation d'échantillonnage et à l'hypothèse d'homoscédasticité des erreurs, qui ne correspondent pas aux observations. En effet, suite au peu de données aux âges élevés, la variance des taux de mortalité observés augmente avec l'âge.

Nous remarquons alors une amélioration du modèle de Lee Carter par le reparamétrages des paramètres du modèle dans le sens où nous observons moins de fluctuations des taux estimés. Ainsi, après l'analyse des intervalles de confiances du nombre de décès, les décès prédits par le modèle sont proches des décès observés. Ce qui rend le modèle satisfaisant globalement.

Au regard des résidus (partie où nous allons pas aborder ici), nous constatons que le modèle de Lee Carter présente plusieurs limites. En effet, ces modèles sont conçus à la base sur de très grand volume de données. Ce qui n'est pas notre cas, rappelons nous sommes dans une portefeuille de 98 313 rentiers avec une profondeur historique de 9 années. De plus, On remarque par ailleurs que le modèle de Lee-Carter ne tient pas compte de la dimension générationnelle. Un modèle développé par Renshaw et Haberman [2005] permet de prendre en compte cette dimension avec un modèle type Age-Période-Cohorte.

2 Ajustement des taux brutes par des méthodes de positionnement

Au vu de la taille de notre portefeuille un peu limité environ 98313 rentiers pour 9 ans d'observations, il semble difficile d'envisager uniquement une construction endogène des tables d'expériences comme nous l'avons commencé dans la section précédente. Nous allons alors à présent construire une table de mortalité avec notre portefeuille en se référant à une table externe.

En 2014, grâce aux travaux de Tomas et Planchet et Wassim YOUSSEF sur la longévité, un package R dénommé ELT, (Experience Life Table) a été mis en place afin de permettre aux organismes d'assurances de construire une table de mortalité reflétant leurs expériences par rapport à des références validées pour suivre le risque de longévité. Des notes techniques sont disponibles sur le site internet de l'Institut des Actuaire et le site de Monsieur Frédéric PLANCHET¹¹. Ces notes expliquent les différentes méthodes proposées au sein du package ainsi que les critères de validation et la méthode de fermeture des tables. Nous renvoyons le lecteur à ces notes pour plus de détails techniques et nous proposons au sein de ce mémoire un exemple d'utilisation du package et l'interprétation des résultats obtenus¹².

Ce package permet de tester quatre méthodes de positionnement, d'ordre de complexité croissant, et des critères de validations afin de comparer les méthodes entre elles et sélectionner la plus adaptée aux données. Il contient un ensemble de fonctions permettant d'implémenter les quatre méthodes de positionnement qui ont été retenues dans ce cadre. voici le schemas proposés par le package pour construire une table d'expérience.

1. Calcul du nombre de décès, d'individus, de l'exposition sur les données ligne à ligne (utilisation de la méthode de Hoem) ;
2. Importation des tables de références ;
3. Exécution de la méthode de positionnement choisie ;
4. Validation de la table obtenue par les critères de proximités entre les observations et le modèle. S'ils ne sont pas validés, on passe à la méthode suivante ;
5. Validation de la table obtenue par les critères de régularité de l'ajustement ;
6. Fermeture de la table obtenue aux grands âges ;
7. Validation de la table finale par les critères de cohérence des tendances de l'évolution

11. <http://www.planchet.net>

12. <http://www.ressources-actuarielles.net/gtmortalite>

Les étapes 3 à 7 sont donc répétées pour chaque méthode choisie. Dans la suite de notre étude, nous allons décrire brièvement les méthodes utilisées au sein de l'outil et détaillerons les résultats obtenus pour la population masculine uniquement. Nous nous sommes concentrés sur les trois premières méthodes proposées. Nous ne proposons pas l'application de la quatrième méthode disponible car celle-ci repose sur un lissage non paramétrique de la table du moment suivi de l'application d'une dérive de mortalité de la référence choisie. Les fluctuations d'échantillonnage de nos données étant très importantes, le choix par avis d'expert des paramètres de lissage s'est vu très délicat et les résultats obtenus incohérents.

Les études faites par Tomas et Planchet [2014 a] proposent deux tables de références : la table INSEE 2060 et la table IA 2013. Dans la suite, nous considérons les tables de l'Institut des Actuariers (IA) comme tables de référence. Cette table reprend les tables réglementaires THG TGF 05. Ces tables générationnelles couvrent la période 1900 à 2005. Elles ont donc été retraitées afin de couvrir la période 2007 à 2035 puis ajustées et extrapolées jusqu'en 2060 à l'aide d'un modèle non paramétrique de p-splines.

2.1 Méthode 1 : Approche à un paramètre avec le SMR

Il s'agit de l'approche la plus aisée à mettre en oeuvre : Elle consiste en l'application d'un coefficient de majoration/minoration aux taux de mortalité de la table de référence pour obtenir les taux de mortalité d'expérience de la population considérée. C'est donc une approche à un paramètre. En pratique ce coefficient correspond au SMR de la population. En reprenant les notations du chapitre précédent, nous le définissons par :

$$SMR = \frac{\sum_{(x^*, t^*)} D_{x^*, t^*}}{\sum_{(x^*, t^*)} E_{(x^*, t^*)} \times q_{x^*}^{ref}(t^*)} \quad (36)$$

où (x^*, t^*) respectivement la plage d'âge et la période d'observation communes au portefeuille de données et à la table de référence. D_{x^*, t^*} et E_{x^*, t^*} correspondent respectivement aux décès observés et à l'effectif sous risque, et $q_{x^*}^{ref}(t)$ est la probabilité de décès à l'âge x durant l'année t .

Le SMR représente alors la dérive de mortalité observée entre le portefeuille et ceux de la référence. Il est important de noter que la table ainsi obtenue, est valide uniquement sur la plage d'âges retenue pour le calcul du SMR. Autrement dit, le choix de la plage d'âges x^* revêt une importance non négligeable car le SMR varie beaucoup. Nous avons alors testé plusieurs plages entre 63 et 94 ans (correspondant à la limite d'âges où les observations sont suffisantes).

	[63, 90]	[63, 94]	[70, 90]	[7, 94]
SMR	0.1054	0.1076	0.1173	0.1101

Nous choisissons de travailler avec la plage d'âges [63 , 94] ans afin de trouver un compromis entre une exposition au risque large et un SMR représentatif des données estimées. Par ailleurs, Ce choix est motivé d'une part par le fait qu'il s'agit de la plage pour laquelle nous avons des données fiables, et d'autre part pour rester cohérent avec les méthodes précédentes et permettre ainsi une meilleure comparaison. Nous remarquons, cependant, que pour chaque plage d'âges testée, le SMR obtenu est toujours supérieur à 1, les références sous estiment les taux de décès.

Les probabilités de décès d'expérience à l'âge x et pour chaque année calendaire t s'obtiennent alors par la relation simple :

$$\tilde{q}_x(t) = SMR \times q_x^{ref}(t) \quad (37)$$

Une fois les taux de décès obtenus, nous pouvons ensuite procéder à une fermeture de la table pour obtenir une table complète.

2.2 Méthode 2 : Approche à deux paramètres avec un modèle relationnel

La deuxième méthode utilisée positionne les données via une régression logistique des probabilités conditionnelles de décès par rapport à celles de la table de référence, soit

$$\text{logit}(\tilde{q}_{x^*}(t^*)) = \alpha + \beta \text{logit}\left(q_{x^*}^{ref}(t^*)\right) + \epsilon_{x^*,t^*} \quad (38)$$

où la fonction logit se définit comme $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ et ϵ_{x^*,t^*} représente le terme d'erreur.

Cette fois-ci, la plage d'âges x^* est choisie afin d'avoir une exposition au risque non nulle, ce qui est le cas pour les rentiers entre [63, 94] ans. α est un indicateur de mortalité qui affecte l'ensemble des âges tandis que β modifie cet effet avec l'âge. Ce modèle intègre ainsi l'estimation et l'extrapolation des tendances de mortalité. On réalise l'estimation en minimisant la distance pondérée par l'exposition entre les probabilités de décès observées $\hat{q}_{x^*}(t^*)$ et estimées $\tilde{q}_{x^*}(t^*)$:

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{(x^*, t^*)} | E_{x^*, t^*} * (\hat{q}_{x^*}(t^*) - \tilde{q}_{x^*}(t^*)) | \quad (39)$$

En choisissant une plage d'âges d'estimation x^* entre 63 et 94 ans et les années d'observations t^* , [2009 : 2017], les probabilités de décès sont obtenues par :

$$\tilde{q}_x(t) = \frac{\exp\left(\hat{\alpha} + \hat{\beta} \text{logit}\left(q_x^{ref}(t)\right)\right)}{1 + \exp\left(\hat{\alpha} + \hat{\beta} \text{logit}\left(q_x^{ref}(t)\right)\right)} \quad (40)$$

Les modèles utilisant le logit, qui est une fonction concave sur l'intervalle $]0 : 1/2]$. cette méthode a tendance alors à sous estimer les probabilités de décès lorsque $q_x(t) < 1/2$. Dans le cadre d'une étude sur les rentes viagères, cela n'est pas contraignant, bien au contraire, car la sous estimation des taux de décès nous conduira d'être prudent.

Nous obtenons les paramètres suivants selon la table de référence IA 2013.

	α	β
Coefficients	0.4546409	1.1176918

L'estimation des paramètres ne nous permettra pas de conclure sur la fidélité et la qualité du modèle, une partie est dédiée à l'analyse des résultats en se basant sur les critères de validation.

2.3 Méthode 3 : Modèle linéaire généralisé de Poisson incluant des interactions avec l'âge et l'année calendaire

Elle s'inspire du modèle linéaire généralisé de Poisson et prend en compte les interactions entre l'âge et l'année calendaire. Le modèle fait l'hypothèse que les décès observés suivent une loi de Poisson telle que :

$$D_{x^*,t^*} \sim P(E_{x^*,t^*} \mu_{x^*}(t^*)) \quad (41)$$

où

$$\mu_{x^*}(t^*) = \beta_0 + \beta_1 \log\left(q_{x^*}^{ref}(t^*)\right) + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^* \quad (42)$$

Sans interactions entre l'âge et l'année calendaire, nous observons des mouvements parallèles entre les forces de mortalité de références et celles estimées, ce qui n'est pas observable en pratique. C'est pourquoi il est plus judicieux d'intégrer des interactions âges \times années. L'intégration de ces interactions demandent cependant un historique de données commun avec la référence, supérieur à une dizaine d'années. Dans le cadre de notre étude, nous

disposons de 9 années communes et par conséquent nous ne prendrons pas en compte les interactions dans notre modélisation.

Les probabilités de décès (sans interactions) sont ensuite obtenues par :

$$\tilde{q}_x(t) = \exp \left(\hat{\beta}_0 + \hat{\beta}_1 \log (q_x^{ref}(t)) + \hat{\beta}_2 x \right) \quad (43)$$

Nous décidons de conserver les plages d'âges [63, 94] ans et les années calendaires [2009, 2017]. Nous obtenons les paramètres suivants selon les deux tables de références IA 2013.

	β_0	β_1	β_2
Coefficients	-0.101157	1.003539	0.003329
p-value	0.0487	3.26e-10	0.0664

Les p-values de l'estimation des paramètres sont acceptables. Donc l'estimation est satisfaisante. Maintenant nous allons faire une analyse globale des trois méthodes ci-après

2.4 Analyse des résultats obtenus

Afin de trouver la meilleure méthode qui ajuste nos données, nous utilisons des critères de validation proposés au sein de l'outil ELT. Ils se divisent en trois groupes. Si l'un n'est pas acceptable, l'ajustement du modèle est arrêté (correspond aux étapes 4, 5 et 7).

Le premier niveau de validation consiste à l'analyse des résultats obtenus en ce qui concerne les critères de proximité. Pour se faire le package ELT fournit de multiples tests et graphiques permettant l'appréciation de ces critères. Parmi ces critères, on peut citer :

- les graphiques sur l'ajustement avec la mortalité observée par année calendaire
- les graphiques sur les résidus de la réponse, de Pearson et de la déviance par année calendaire
- les graphiques sur les intervalles de confiance des décès ajustés et observés
- les tests du rapport de vraisemblance, du SMR, de Wilcoxon
- les quantités comme la déviance, le R^2 , le MAPE ou χ^2 .

la deuxième étape du processus de validation concernant la régularité de l'ajustement, voir Tomas and Planchet [2014b]. On retrouve les tests des runs et des signs.

la troisième étape du processus de validation concerne la cohérence des tendances d'évolution de la mortalité extrapolée. Nous pouvons ainsi étudier.

- les espérances de vie generationnelles partielles sur 5 ans,
- les esperances de vie periodiques residuelles
- Les surfaces des espérances de vie generationnelles partielles sur 5 ans par sexes
- Le rapport des espérances de vie generationnelles partielles sur 5 ans Femme / Homme
- Le rapport des espérances de vie périodiques résiduelles observées et ajustées par sexe
- Le rapport des probabilités de décès ajustées Femme / Homme
- Des graphiques illustrant la cohérence des espérances de vie périodiques résiduelles ajustées pour les âges 55, 70 et 85 par sexe

Nous présentons succinctement les résultats obtenus pour la population masculine, avec les trois méthodes utilisées. Rappelons que ici on utilise la table de référence IA 2013. La définition détaillée de chaque test est disponible au sein de la note de travail II 1291-14 v1.2.

Test de validation 1 :

Nous commençons tout d'abord par regarder les graphiques relatifs à l'ajustement de la mortalité observée par année calendaire. Nous projettons l'anayse en 2010 et en 2017.

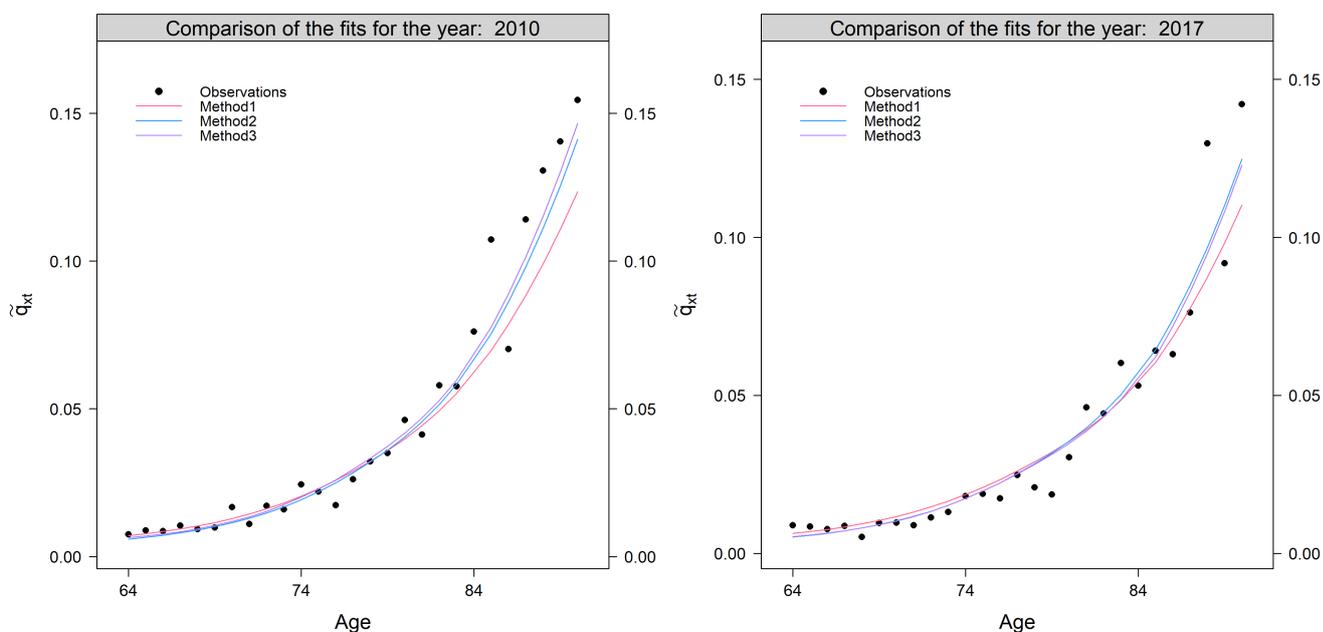


FIGURE 80 – Taux q_x ajustés en 2010 et 2017 via les 3 méthodes de positionnement

Nous remarquons qu'en date 2010 les trois méthodes sont très proches, la méthode 1

sousestime légèrement les taux de décès par rapport aux deux autres méthodes. Nous remarquons la même chose en 2017, mais la différence est beaucoup plus marquée en 2010. On constate également, les taux de décès estimés par la méthode 3 est plus grand que ceux par la méthode 2.

A première vue, il est difficile de dire quelles méthodes et quelles tables sont les plus appropriées. Nous remarquons cependant qu'au delà de 80 ans l'ajustement se fait plus difficile.

Ceci sera corrigé par la suite avec le processus de fermeture des tables et nous avons ainsi une première idée quant à l'âge pivot de fermeture.

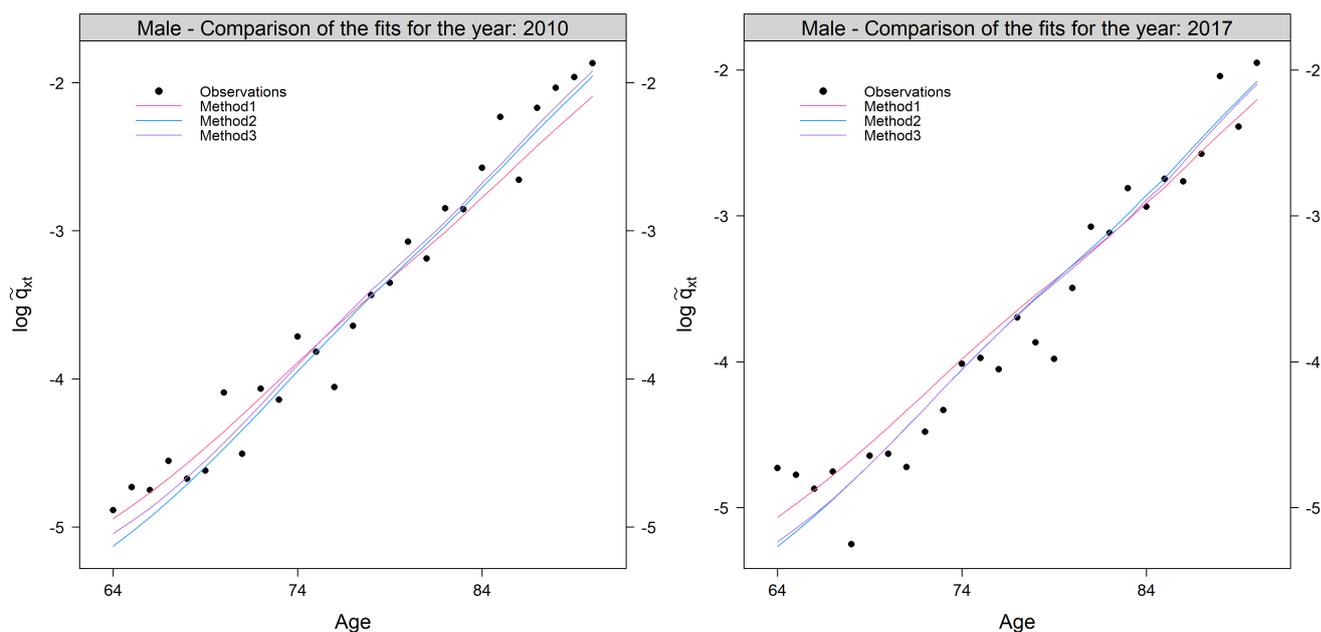


FIGURE 81 – Taux q_x ajustés en 2010 et 2017 via les 3 méthodes de positionnement

Nous constatons qu'en 2010 ou 2017, un point d'inflexion à 80 ans où les tendances estimées s'inversent. Ceci dans le sens où entre 63 et 80 ans, la méthode qui sur estime le plus les taux de décès est la méthode 1 tandis qu'après 80 ans, c'est la méthode 1 qui sous estime le plus les taux de décès. Ceci nous conforte sur l'idée d'un âge pivot de fermeture à 80 ans.

Il est une nouvelle fois difficile de définir la méthode la plus adaptée à nos données. Nous étudions alors les graphiques relatifs aux résidus de la réponse et de Pearson ainsi que les résidus de la déviance par année.

1. Les résidus de la réponse correspondent à la différence entre les observations et les valeurs ajustées : $r_{x,t} = \hat{q}_x(t) - \tilde{q}_x(t)$
2. Les résidus de Pearson se définissent comme : $r_{x,t} = \frac{D_{x,t} - E_{x,t}\tilde{q}_x(t)}{\sqrt{\text{Var}[E_{x,t}\tilde{q}_x(t)]}}$

Nous veillerons alors à ce que les résidus de la réponse et de Pearson ne présentent pas de tendance forte. S'ils sont de plus compris entre $[-2, 2]$, nous pourrions considérer que la méthode utilisée modélise correctement la variabilité des données.

1. Les résidus de la déviance sont définis par : $r_{x,t} = \text{signe}(D_{x,t} - E_{x,t}\tilde{q}_x(t)) * \sqrt{\text{Déviance}_{x,t}}$
2. La déviance est une mesure de qualité de l'ajustement. Lorsque nous supposons que les décès suivent une loi de Poisson $D_{x,t} \sim P(E_{x,t}q_x(t))$, la déviance se définit comme :

$$\text{Si } D_{x,t} > 0, \text{ Déviance}_{x,t} = 2 \left(D_{x,t} \ln \left(\frac{D_{x,t}}{E_{x,t}\tilde{q}_x(t)} - (D_{x,t} - E_{x,t}\tilde{q}_x(t)) \right) \right) \quad (44)$$

$$\text{Si } D_{x,t} = 0, \text{ Déviance}_{x,t} = 2E_{x,t}\tilde{q}_x(t) \quad (45)$$

Si plusieurs résidus de la déviance successifs sont de même signe, alors les données sont sur lissées. Nous présenterons ici les résultats en année 2010.

Male - Plot of the residuals - year: 2010

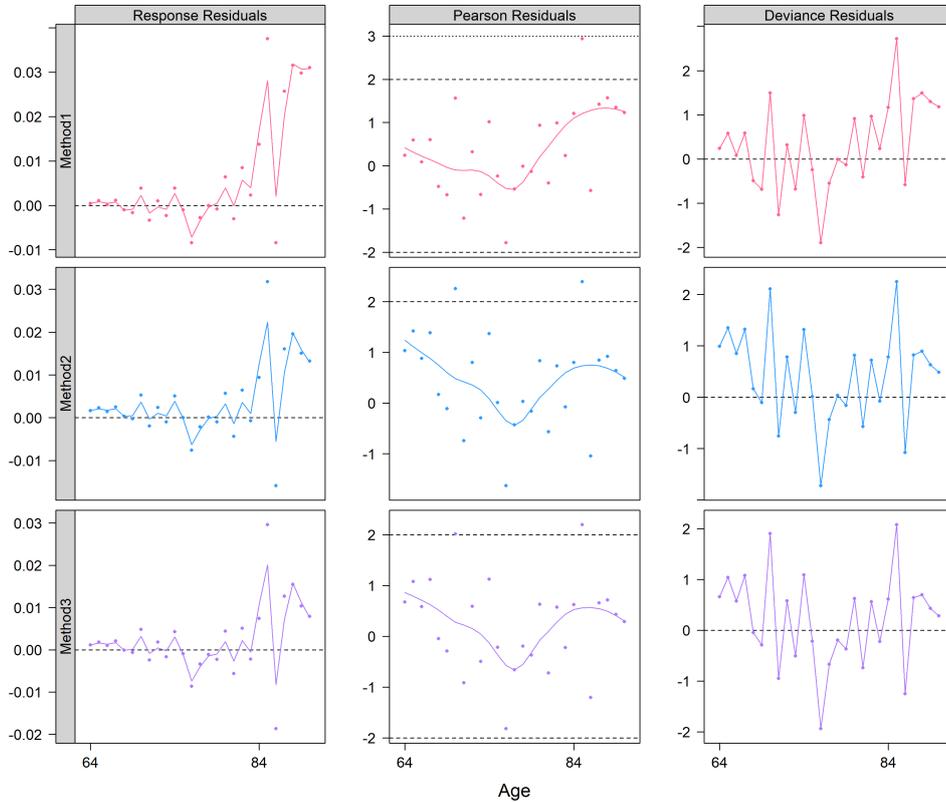


FIGURE 82 – Résidus relatifs à la table de référence IA, en $t = 2010$

Nous détectons une tendance forte entre 63 et 80 ans (tendances linéaires successives), indiquant un ajustement inapproprié pour ces âges. Nous retrouvons cette observation au travers des résidus de la déviance où une suite de résidus sont de même signe entre 63 et 70 ans. Nous constatons, cependant, qu'en général, les résidus de Pearson appartiennent à l'intervalle $[-2,2]$ ce qui indique que les modèles capturent correctement la variabilité des données.

Nous étudions également la qualité de l'ajustement via les intervalles de confiance des décès estimés. Nous supposons que les décès observés sont correctement représentés si l'ensemble des observations se situe dans l'intervalle de confiance à 95% défini comme suit :

Nous supposons que $D_{x,t} \sim N(E_{x,t}q_x(t), E_{x,t}q_x(t)(1 - q_x(t)))$. L'intervalle de confiance de niveau $(1 - \alpha)$ est donné par :

$$\left[E_{x,t}\tilde{q}_x(t) - u_{\alpha/2}\sqrt{E_{x,t}\tilde{q}_x(t)(1 - \tilde{q}_x(t))}; E_{x,t}\tilde{q}_x(t) + u_{\alpha/2}\sqrt{E_{x,t}\tilde{q}_x(t)(1 - \tilde{q}_x(t))} \right] \quad (46)$$

où $u_{\alpha/2}$ est le $(1 - \alpha/2)$ quantile de la distribution normale. Voici les résultats obtenus pour l'année 2010 :

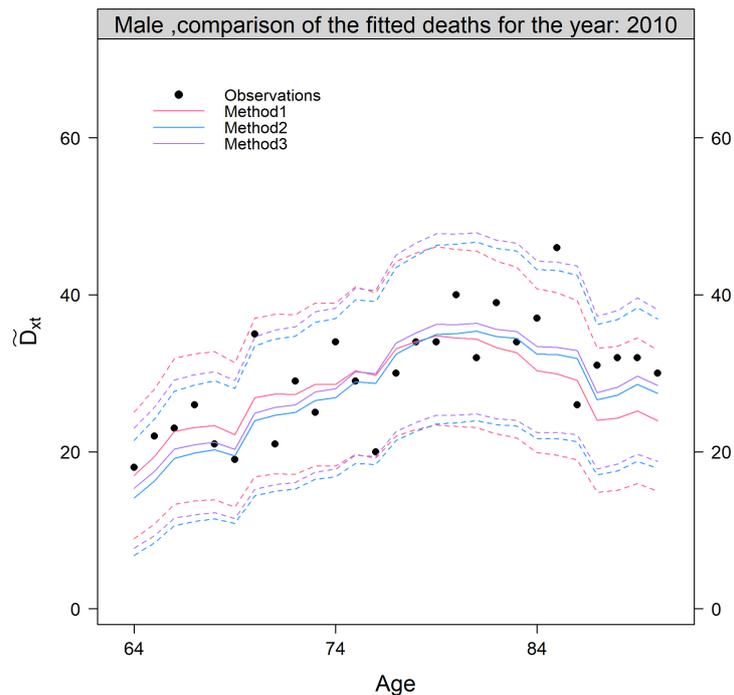


FIGURE 83 – les décès estimés, observés et les intervalles de confiance pour les trois méthodes en $t = 2010$

Nous remarquons que les décès observés sont (sauf quelques exceptions) bien dans l'intervalle de confiance à partir des taux estimés par le modèle. De plus, avec une analyse aux grands âges, on remarque plus de décès observés qui sont hors de l'intervalle de confiance.

Grâce aux études graphiques, nous avons une première idée de l'ajustement des taux de décès. Cependant, nous ne pouvons pas déterminer avec certitude la méthode la plus adéquate. Nous allons donc étudier les critères statistiques du groupe 1.

Nous détaillons succinctement les différents tests disponibles et rappelons que l'ensemble des critères de validation sont disponibles au sein de la note de travail de l'Institut des Actuariers *II1291 – 14, TomasetPlanchet[2014c]*.

Le test de rapport de vraisemblance

Nous cherchons à déterminer si l'ajustement correspond à la loi de mortalité sous-jacente

(correspond à l'hypothèse nulle). La statistique du test se définit comme :

$$\xi^{LR} = \sum_{(x,t)} D_{x,t} \ln \frac{D_{x,t}}{E_{x,t} \tilde{q}_x(t)} - (D_{x,t} - E_{x,t} \tilde{q}_x(t)) \quad (47)$$

Si l'hypothèse nulle est vraie, la statistique suit une loi de χ^2 avec un nombre de degrés de liberté égal au nombre d'observations. Nous rejettons l'hypothèse nulle si $\xi^{LR} > \chi_{1-\alpha}^2(n)$ où $\chi_{1-\alpha}^2(n)$ est le quantile à $(1 - \alpha)$ de la distribution de χ^2 avec n degrés de liberté. La p -value est donnée par

$$p - \text{value} = P [\chi_{1-\alpha}^2(n) > \xi^{LR}] = 1 - F_{\chi_{1-\alpha}^2(n)}(\xi^{LR}) \quad (48)$$

On cherchera alors à avoir une p -value proche de 1.

Le test du SMR

Il correspond au rapport entre le nombre de décès observés et ajustés. Sous l'hypothèse que les décès suivent une loi de Poisson, de paramètre $E_{x,t} \tilde{q}_x(t)$,

$$SMR = \frac{\sum_{x,t} D_{x,t}}{\sum_{x,t} E_{x,t} \tilde{q}_x(t)} \quad (49)$$

Le test cherche à savoir si le SMR ainsi déterminé est significativement différent de 1. Pour cela, nous calculons la statistique suivante :

$$\text{si } SMR > 1, \xi^{SMR} = 3 * D^{\frac{1}{2}} \left(1 - (9D)^{-1} - (D/E)^{\frac{1}{3}} \right) \quad (50)$$

$$\text{si } SMR < 1, \xi^{SMR} = 3 * D^{*\frac{1}{2}} \left((D * /E)^{\frac{1}{3}} + (9D*)^{-1} - 1 \right) \quad (51)$$

où $D = \sum_{x,t} D_{x,t}$, $D* = \sum_{x,t} D_{x,t} + 1$ et $E_{x,t} = \sum_{x,t} E_{x,t} \tilde{q}_x(t)$. Nous renvoyons le lecteur à l'étude du SMR et de son intervalle de confiance par LIDDELL [1984] pour plus de précisions.

Si le SMR est significativement différent de 1 (l'hypothèse nulle est vraie), la statistique suit une loi Normale centrée réduite. On rejette l'hypothèse nulle lorsque $\xi^{SMR} > N_{1-\alpha}(0, 1)$ où $N_{1-\alpha}(n)$ est le quantile à $(1 - \alpha)$ de la distribution Normale centrée réduite. La p -value est donnée par

$$p - \text{value} = 1 - F_{N(0,1)}(\xi^{SMR}) \quad (52)$$

On cherchera alors à avoir une $p - value$ proche de 1.

Le Test de Wilcoxon

Ce test prend en compte la direction et la grandeur des différences entre les paires composées des probabilités de décès observées et ajustées. Nous cherchons à tester si la médiane entre les différences de chaque paire est nulle (hypothèse nulle). On détermine pour cela les différences entre probabilités de décès observées et ajustées puis nous les classons par ordre croissant des valeurs absolues, sans les différences nulles. Nous attribuons à chaque différence ainsi classée son rang, et en cas d'ex-aequo, la moyenne. w_- , w_+ représente la somme des rangs des différences, respectivement strictement négatives et strictement positives, et w , le plus grand nombre entre les deux $w = \max(w_-; w_+)$. L'hypothèse nulle est vraie si w_+ et w_- sont à peu près égales. Ayant un nombre d'observation supérieur à 15, nous déterminons la statistique du test par :

$$\xi^{WIL} = \frac{(w - 1/2 - n(n + 1))/4}{\sqrt{n(n + 1)(2n + 1)/24}} \quad (53)$$

Si l'hypothèse nulle est vraie, alors la statistique suit une loi Normale centrée réduite $\xi^{WIL} \sim N(0; 1)$ et nous rejetons l'hypothèse nulle lorsque $|\xi^{WIL}| > N_{1-\alpha/2}(0, 1)$ où $N_{1-\alpha/2}(0, 1)$ est le quantile de niveau $(1 - \alpha/2)$ de la distribution Normale centrée réduite. La $p - value$ est donnée par

$$p - value = P [N_{1-\alpha/2}(n) > |\xi^{WIL}|] = 2 * (1 - F_{N(0,1)}(|\xi^{WIL}|)) \quad (54)$$

On cherchera alors à avoir une $p - value$ proche de 1.

Le test de χ^2

Cet indicateur mesure la qualité de l'ajustement du modèle. Nous le calculons comme :

$$\chi^2 = \sum_{x,t} \frac{(D_{x,t} - E_{x,t}\tilde{q}_x(t))^2}{E_{x,t}\tilde{q}_x(t)(1 - \tilde{q}_x(t))} \quad (55)$$

On cherchera à avoir χ^2 le plus faible.

Le coefficient de détermination R^2

Le R^2 ou coefficient de détermination, mesure l'adéquation entre le modèle et les données observées. On le définit comme la part de variance expliquée par rapport à la variance

totale :

$$R^2 = 1 - \left(\frac{\sum_{x,t} (D_{x,t}/E_{x,t} - \tilde{q}_x(t))^2}{\sum_{x,t} \left(D_{x,t}/E_{x,t} - \left(\sum_{x,t} (D_{x,t}/E_{x,t}) / n \right) \right)^2} \right) \quad (56)$$

où n correspond au nombre d'observations. On cherchera alors à avoir un R^2 proche de 1.

MAPE

le MAPE ou Mean Average Pourcentage *Error* mesure l'exactitude de l'ajustement par rapport aux observations. Il se détermine comme la moyenne des écarts en valeur absolue, des ajustements par rapport aux valeurs observées.

$$MAPE = \frac{\sum_{x,t} |(D_{x,t}/E_{x,t} - \tilde{q}_x(t)) / (D_{x,t}/E_{x,t})|}{\sum_{x,t} D_{x,t}} * 100 \quad (57)$$

Les observations doivent être non nulles.

La déviance Elle mesure la qualité de l'ajustement. Nous avons vu, qu'en supposant que les décès suivent une loi de Poissons $D_{x,t} \sim P(E_{x,t}q_x(t))$, nous définissons la déviance comme :

$$\text{Si } D_{x,t} > 0, \text{ Déviance }_{x,t} = 2 \left(D_{x,t} \ln \left(\frac{D_{x,t}}{E_{x,t}\tilde{q}_x(t)} - (D_{x,t} - E_{x,t}\tilde{q}_x(t)) \right) \right) \quad (58)$$

$$\text{Si } D_{x,t} = 0, \text{ Déviance }_{x,t} = 2E_{x,t}\tilde{q}_x(t) \quad (59)$$

La déviance totale est définie par $\sum_{x,t}$ Déviance $_{x,t}$.

Voici les résultats des tests décrit ci-dessus résumés dans le tableau.

Nous remarquons que la méthode 3 est la plus adéquate en termes de qualité de l'ajustement et d'écarts avec les observations. C'est notamment la seule méthode à être acceptée par le test de vraisemblance qui vérifie l'adéquation de l'ajustement.

Lorsque nous regardons les résultats obtenus pour le test du SMR, nous observons des p - *value* dans l'ensemble correctes excepté pour la méthode 2 où les décès observés sont sousestimés. Ce critère est primordial au vu de l'objectif de notre étude : nous souhaitons au mieux estimer la mortalité de notre portefeuille et ainsi déterminer de manière plus

		Method1	Method2	Method3
Test rapport de vraisemblance	Xi	330.98	392.19	277.15
	Hyp	H1	H1	H0
	p.val	0,0001	0	0,0653
Test du SMR	SMR	1	1.0135	1
	Xi	0.0074	1.1986	0.0037
	Hyp	H0	H0	H0
	p.val	0.497	0.1153	0.4985
Test de Wilcoxon	W	15485	16294	15139
	Xi	0.6031	1.3406	0.2876
	Hyp	H0	H0	H0
	p.val	0.5465	0.18	0.7736
Quantités	Chi2	346,43	298,6	286,97
	R2	0,8974	0,931	0,9328
	MAPE	17,74	16,37	16,42
	Deviance	661,96	784,38	554,29

FIGURE 84 – Résultats des tests statistiques du premier groupe de critère

précise que les tables réglementaires, le nombre de décès observés.

Pour le test de Wilcoxon, la $p - value$ associée à la méthode 3 est relativement bonne. Ce test est très important car il compare en grandeur et en direction les différences entre les probabilités de décès observées et ajustées.

En observant les quantités calculées, nous remarquons des R^2 très importantes proche de 1, notamment la méthode 2 et 3. Concernant les valeurs de MAPE, nous avons les pourcentages de décès observés en dehors de l'intervalles de confiance. Plus que c'est petit mieux est le modèle. En observant les MAPE, nous retrouvons les mêmes conclusions.

L'ensemble des tests, nous montre a priori la méthode 3 comme la meilleure méthode pour un ajustement via la table de référence IA 2013, il s'en suit la méthode 2. Ces observations sont opposées aux observations graphiques où les décès étaient surestimés par ces mêmes méthodes. Il est difficile une nouvelle fois de déterminer le meilleur ajustement à partir de ces indicateurs.

Pour aller plus loin dans l'analyse, nous allons maintenant voir les tests de qualité de l'ajustement et notamment avec les statistiques des signes et des runs, correspondant aux critères de validation de niveau 2. Ce niveau de validation permet également de voir si les données n'ont pas été sur où sous lissées.

Test de validation 2 :

Ci-dessus, nous allons d'abord présenter de manière condensée les différents tests réalisés :

Le test des signes

Nous examinons au travers de ce test non paramétrique, la fréquence des changements de signes de la différence entre les probabilités de décès observées et ajustées. L'hypothèse nulle est vraie si la médiane entre les signes positifs et négatifs de la différence est nulle. En posant n_+ et n_- , respectivement le nombre de signes positifs et négatifs et $n = n_+ + n_-$, la statistique du test s'écrit :

$$\xi^{SIG} = \frac{|n_+ - n_-| - 1}{\sqrt{n}} \quad (60)$$

Si l'hypothèse nulle est vraie, ξ^{SIG} suit une loi Normale centrée réduite et l'hypothèse nulle sera rejetée si $|\xi^{SIG}| > N_{1-\alpha/2}(0, 1)$ où $N_{1-\alpha/2}(0, 1)$ est le quantile de niveau $(1 - \alpha/2)$ de la distribution Normale centrée réduite. La p -value est donnée par

$$p - \text{value} = P [N_{1-\alpha/2}(n) > |\xi^{SIG}|] = 2 * (1 - F_{N(0,1)}(|\xi^{SIG}|)) \quad (61)$$

On cherchera à avoir une p -value proche de 1.

Le test des runs

Ce test non paramétrique détermine si les éléments d'une séquence sont mutuellement indépendants. Nous appelons *run*, un segment d'une séquence, formé par des éléments adjacents égaux. A titre d'exemple la séquence ci-dessous est composée de 18 éléments, 6 runs dont 3 avec des + et 3 avec des -.

$$(- - - + + + + + - - + - - + + +) \quad (62)$$

L'hypothèse nulle est vraie si une séquence de n éléments est une variable aléatoire dont la distribution conditionnelle sachant n_+ et n_- est approximativement Normale où :

$$\mu = \frac{2n_+n_-}{n_+ + n_-} \text{ et } \sigma^2 = \frac{2n_+n_- (2n_+n_- - (n_+ + n_-))}{(n_+ + n_-)^2 (n_+ + n_- - 1)} \quad (63)$$

La statistique est définie comme $\xi^{RUN} = \frac{\text{Nombre de runs} - \mu}{\sigma}$. Cette statistique suit une loi Normale centrée réduite si l'hypothèse nulle est vraie. Nous rejetons celle-ci si $|\xi^{RUN}| >$

$N_{1-\alpha/2}(0, 1)$ où $N_{1-\alpha/2}(0, 1)$ est le quantile de niveau $(1 - \alpha/2)$ de la distribution Normale centrée réduite. La $p - value$ est donnée par :

$$p - value = P [N_{1-\alpha/2}(n) > |\xi^{RUN}|] = 2 * (1 - F_{N(0,1)} (|\xi^{RUN}|)) \quad (64)$$

On cherchera à avoir une $p - value$ proche de 1.

En application avec non données, nous obtenons les résultats suivants.

		Method1	Method2	Method3
Test des runs	+(-)	117	133	131
	Xi (abs)	0, 7069	1, 4149	1, 0931
	Hyp	H0	H0	H0
	p.val	0.4797	0.1571	0.2744
Test des signes	Nombre	121(122)	129(114)	121(122)
	Xi	0	0, 8981	0
	Hyp	H0	H0	H0
	p.val	1	0, 3691	1

(65)

Pour le test des signes, la méthode 3 et la méthode 1 s'avèrent les meilleurs modèles d'ajustement avec des $p - value$ relativement élevées. Pour le test des runs, les p-value obtenues pour la méthode 1 et 3 via sont très bonnes. Ces deux tests sont non paramétriques et dépendent de la qualité des données. Or nous savons que nous avons de fortes fluctuations d'échantillonnage et la mesure de la régularité de l'ajustement est plus difficile.

Le choix entre l'une ou l'autre table de référence est difficile car les résultats des tests et l'étude graphique de l'ajustement donnent des résultats proches. Nous décidons de conserver pour le moment l'ensemble des méthodes utilisées.

Nous avons vu qu'aux grands âges, les ajustements étaient peu réalistes suites aux faibles expositions sur ces âges. Nous allons donc procéder à la fermeture des tables.

2.5 Fermeture des tables

La fermeture se fait en deux étapes. Il s'agit dans un premier temps de définir la tranche d'âge dans laquelle l'âge optimal de départ est déterminé pour le calcul des coefficients du modèle, cf. Tomas et Planchet (2014b), et dans un deuxième temps de valider la régularité

de la fermeture visuellement.

Le méthode proposée au sein du package ELT, et que nous conservons, est la méthode de Denuit et Goderniaux [2005]. Le lecteur intéressé pourra se référer à Quashie et Denuit [2005] pour la présentation et la comparaison de différents modèles de fermeture.

Nous n'allons pas développer la partie théorique ici car, la méthode de Denuit et Goderniaux [2005] est déjà détaillée succinctement dans la partie précédente. Nous contenterons d'analyser les résultats de la fermeture et la validation des tables via le troisième groupe de critère.

Pour définir la tranche d'âges à partir duquel on extrapole la table, nous appliquons les taux lissés plutôt que les taux ajustés, nous utilisons un âge commun pour chaque année t afin d'obtenir une surface de mortalité plus régulière que si nous ajustons pour chaque année les taux à des âges différents. L'âge est quant à lui déterminé à partir des résultats graphiques sur les ajustements des taux de décès estimés. Nous avons notifié lors de leurs études que l'ajustement montrait un point d'inflexion à 80 ans.

Nous obtenons les résultats suivants.

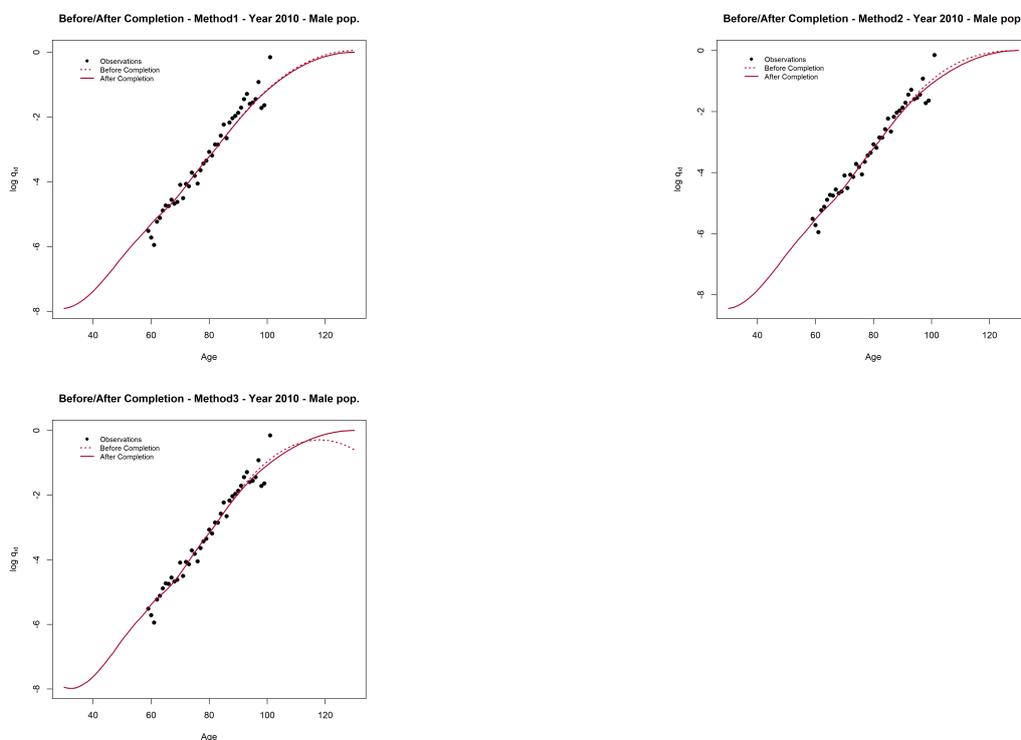


FIGURE 85 – Résultats de fermeture des tables en 2010

Nous ne constatons pas de cassures au point de raccordement, la fermeture est donc bonne. Lorsque nous regardons les taux de décès estimés avant et après fermeture entre 100 et 130 ans, nous remarquons pour la méthode 3 que la fermeture réajuste les taux où nous constatons une baisse des taux au grands âges, ce qui n'est pas observé en pratique. Concernant les âges entre 30 et 63 ans, nous utilisons les taux de décès observés par la référence de mortalité.

Pour valider la fermeture aux grands âges de nos données, nous réalisons l'étude des critères de validation du groupe 3.

Test de Validation 3

Après avoir fermé les tables, on dispose des tables complétées jusqu'à 130 ans. pour effectuer la troisième étape du processus de validation concernant la cohérence des tendances d'évolution de la mortalité extrapolée, voir Tomas et Planchet (2014b).

Primo, nous déterminons la fonction de survie à partir des tables fermées des cohortes d'un âge x , l'année t . On se ramène alors à travailler sur une diagonale du diagramme de Lexis et non une ligne horizontale. Nous indiquons par \uparrow l'étude par cohorte et \nearrow , l'étude

par période.

La fonction de survie est ainsi définie par :

$$S_{\tilde{x}}^{\nearrow} = \prod_{j=0}^{x-1} (1 - q_{\tilde{x}+1}(t + j)) \quad (66)$$

A partir de la fonction de survie, nous pouvons déterminer plusieurs marqueurs comme :

L'espérance de vie générationnelle partielle

Nous déterminons l'espérance de vie partielle sur w années pour un individu d'une cohorte g âgé de \tilde{x} l'année t par :

$$we_{\tilde{x}}^{\nearrow} = \int_1^w S_{\tilde{x}}^{\nearrow}(u) du \quad (67)$$

Avec l'hypothèse de constances par morceaux des forces de mortalités (c'est à dire, entre x et t), on obtient l'estimateur :

$$we_{\tilde{x}}^{\nearrow} = \sum_{x=1}^w \prod_{j=0}^{x-1} (1 - q_{\tilde{x}+j}(t + j)) \quad (68)$$

L'espérance de vie périodique résiduelle Comme nous l'avons brièvement défini précédemment, il s'agit cette fois ci de l'espérance de vie résiduelle sur w années d'un individu âgé de x années à la date t :

$$we_x^{\uparrow} = \sum_{\delta=1}^w \prod_{j=0}^{\delta-1} (1 - q_{x+j}(t)) \quad (69)$$

L'âge médian au décès L'âge médian au décès d'un individu de la cohorte g , âgé de \tilde{x} à la date t sur w années et noté $\text{Med} [{}_w T_{\tilde{x}}]$ correspond à la valeur médiane de la distribution de probabilité de la durée de via aléatoire $T_{\tilde{x}}$:

$$S_{\tilde{x}}^{\nearrow} (\text{Med} [{}_w T_{\tilde{x}}]) = 0,5 \quad (70)$$

Pour notre étude, nous nous concentrons sur l'étude de l'espérance de vie résiduelle périodique entre 63 et 94 ans afin d'établir des comparaisons cohérentes avec nos données brutes. Nous regardons dans un premier temps l'évolution de l'espérance de vie en 2009

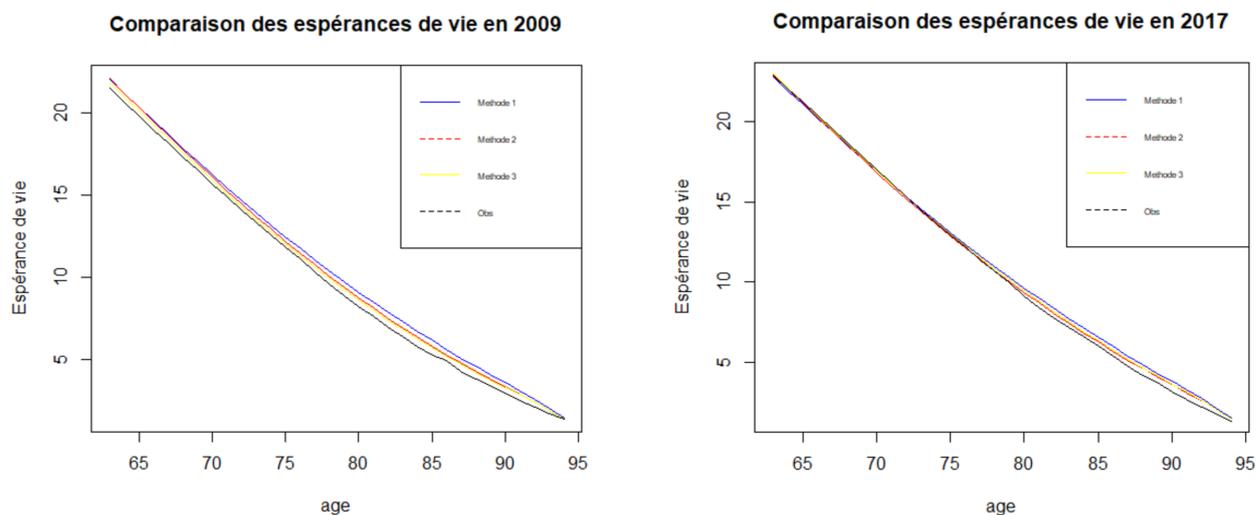


FIGURE 86 – Comparaisons des espérances de vie des trois méthodes en 2009 et 2017

de nos données (au sens de l'estimateur de Hoem des probabilités de décès) et des trois méthodes ajustées.

Nous remarquons que les espérances de vie estimées sont très proches de celles observées notamment pour l'année 2017. Nous constatons pour l'année 2009, une meilleure approximation de l'espérance de vie par la méthode 3 puis la méthode 2. De plus, en 2017, les espérances de vie résiduelles sont presque confondues avec les valeurs observées entre 63 et 75, puis elles s'éloignent légèrement entre 75 et 94 ans.

Nous regardons également les ratio entre l'espérance de vie des données observées par rapport aux données ajustées par les trois méthodes.

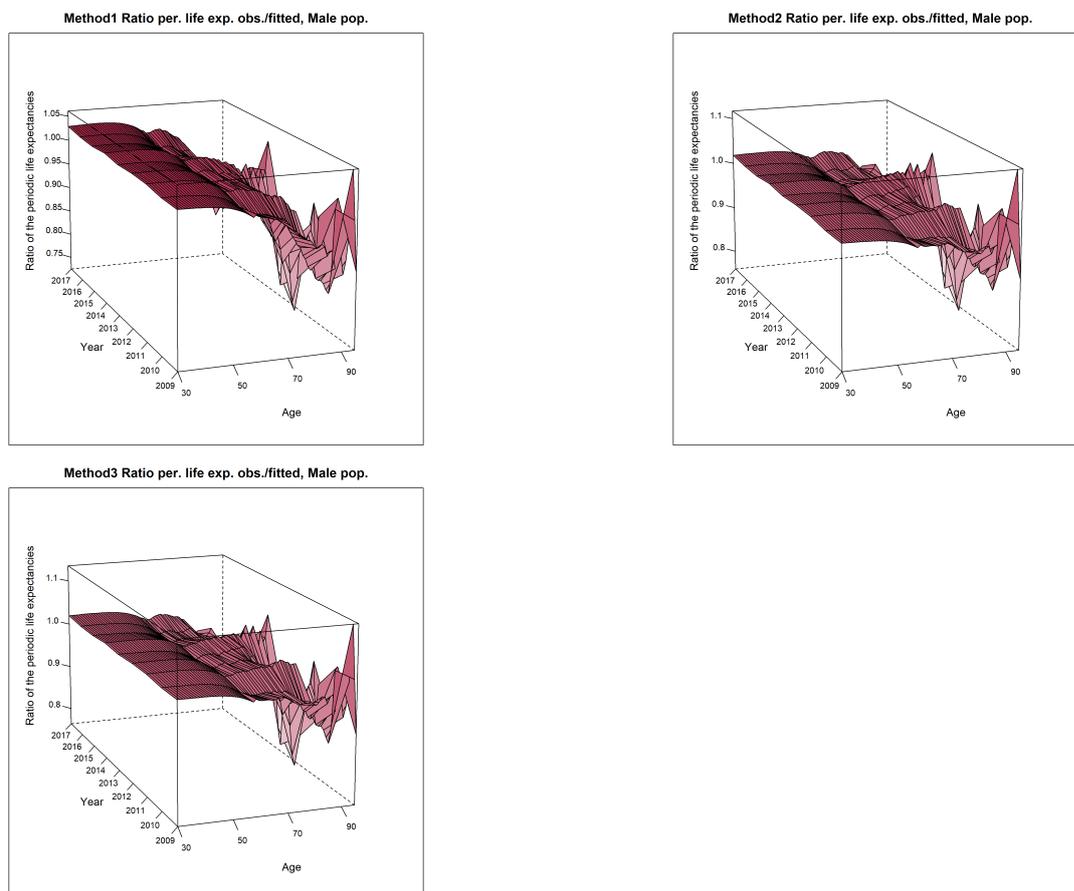


FIGURE 87 – Ratio des espérances de vie entre 2009 et 2017 et sur la plage d'âge [63, 94]

Nous constatons alors que les écarts sont plus importants aux âges élevées, dus au peu d'observations présentes à ces âges. Comme nous l'avons remarqué avec la figure 86, les estimations sont très proches et les écarts de mêmes ordres de grandeurs. Au vue de ces analyses, tous les critères de validations montrent que la méthode 3 est la plus adéquate à nos données. Par ailleurs, cette remarques peut être favorisé par le choix sur la table de référence, qui n'est pas sans risque. Il serait très important d'étudier le risque systématique d'avis d'expert portant sur un mauvais jugement lors de l'utilisation d'informations externes (à l'occurrence la table de référence IA 2013). Mais, nous avons fait le choix d'omettre cette partie à cause d'un manque de temps et de réserver cette partie pour une autre étude. Tout de même, une étude de notre portefeuille, montre que la table IA est adéquate avec nos données.

Toutesfois, Il est difficile de juger, à première vue, de la pertinence d'une table ou d'une

autre car les résultats obtenus sont très proches entre eux et de nos données. La fluctuations d'échantillonnage rend l'interprétation des tests statistiques difficiles et contradictoires.

Pour rappel, la méthode 1 correspond à la translation des taux de décès observés par rapport au taux de décès de référence, la méthode 2, une translation par rapport aux logit et la méthode 3 intègre une hypothèse de répartition des décès selon une loi de Poisson. Les méthodes sont donc simples à implémenter et comprendre mais sont très sensibles aux tables de références choisies.

L'objectif de notre étude est de construire une table d'expérience qui reflète la mortalité de notre portefeuille plus fidèlement que les tables réglementaires. En effet, l'étude réalisée ce jour souhaite étudier la dérive viagère induite par les tables de mortalité et ainsi établir un suivi technique de l'évolution de la mortalité de nos rentiers. Nous décidons de conserver les trois méthodes ainsi construites par sexe, et nous choisirons la plus appropriée au vu du SMR.

2.6 Comparaison des méthodes

Nous avons à présent analyser et calibrer différents modèles (modèle de Lee Carter et les trois modèles de positionnement via le package ELT). Il convient d'effectuer une comparaison de ces modèles afin de retenir celui qui permet de mieux capter la mortalité de la population de notre portefeuille de rentiers. Nous choisissons de baser notre comparaison essentiellement sur la valeur du SMR obtenue par chaque modèle et, sur la cohérence du caractère prospectif de chaque modèle au travers de l'espérance de vie résiduelle.

2.6.1 Comparaison du SMR

L'analyse du SMR nous permet d'apprécier la proximité entre les valeurs estimées par le modèle et les valeurs observées dans le portefeuille. Pour chacun de nos modèles nous obtenons les valeurs suivantes pour les années allant de 2009 à 2017 (période commune aux quatre méthodes) :

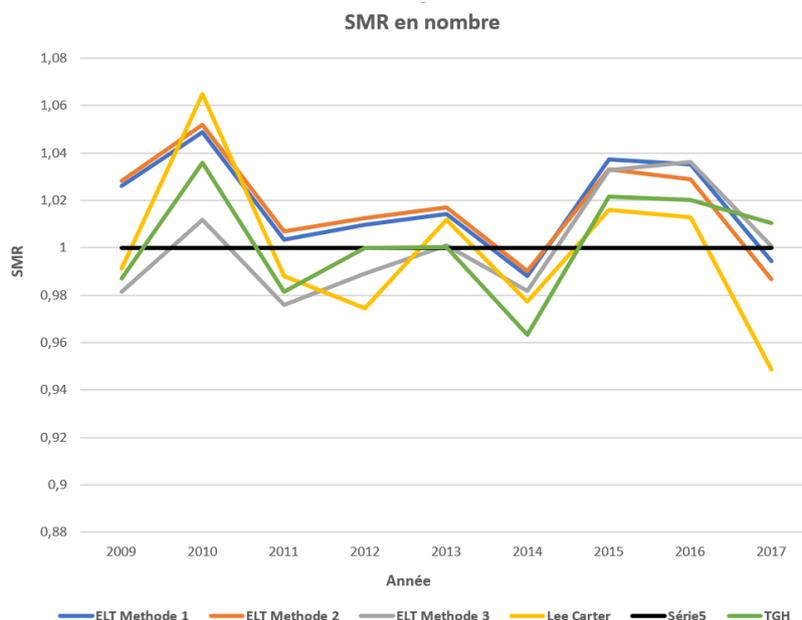


FIGURE 88 – SMR en nombre des modèles et la table TGH

Pour la population masculine, les SMR montrent que les tables d'expériences sont plus prudentes que la table réglementaire à l'exception de la méthode 3 qui a tendance à sous-estimer les taux de décès et, contrairement le modèle de Lee Carter. Ce résultat est satisfaisant pour notre étude car nous souhaitons approcher la tendance de mortalité de notre portefeuille de manière plus précise mais tout en gardant une marge de prudence afin de ne pas sous-estimer les probabilités de décès.

Nous remarquons, des mouvements similaires entre les tables d'expériences et les tables réglementaires. On retrouve ici les translations réalisées au travers de l'ajustement par positionnement des tables. De plus, le modèle de positionnement, la méthode 2, présente de meilleurs résultats en termes de SRM car, il est plus proche de la table TGH, tout en surestimant les probabilités de décès.

2.6.2 Comparaison des espérances de vie résiduelles périodiques

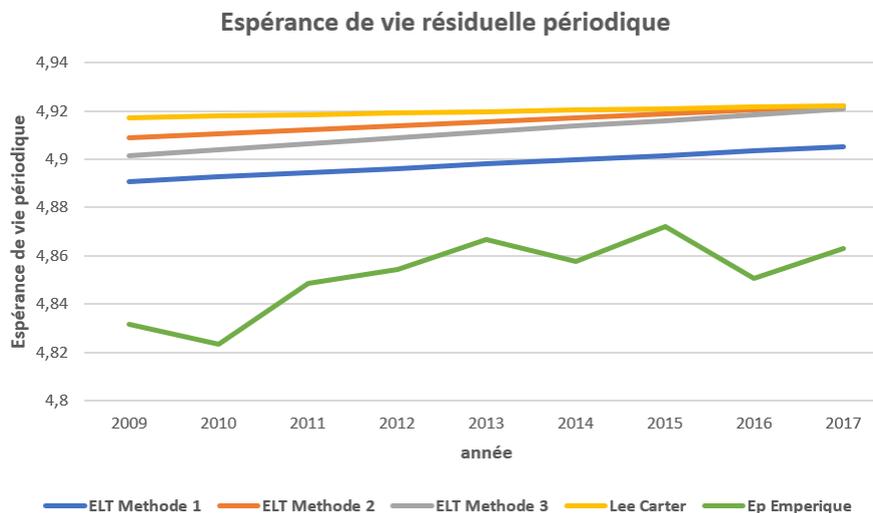


FIGURE 89 – Comparaison des espérances de vie sur 5 ans à l'âge de 63 ans

Nous remarquons que les méthodes implémentées prédisent une tendance espérance de vie générationnelle partielle relativement similaire, qui se situe à un niveau supérieur par rapport aux données brutes. tous les modèles ont un lissage satisfaisant et reste prudent par rapport aux données de portefeuille.

En considérant le SMR obtenu par chaque modèle, l'analyse de l'espérance de vie résiduelle ainsi que le backtesting via l'intervalle de confiance des décès prédits, nous retenons la méthodes 2 (Approche à deux paramètres avec un modèle relationnel) et la méthode 3 (modèle linéaire généralisé de Poisson) comme les meilleurs modèles pour nos données, ils présentent des résultats satisfaisants. Cependant entre les deux modèles, nous faisons le choix de retenir la méthode 2 (approche à deux paramètres avec un modèle relationnel). En effet, nous souhaitons conserver une marge de prudence dans nos estimations or au vu de l'analyse du SMR et de l'espérance de vie résiduel la méthode 3 sous-estime les probabilités de décès notamment pour les grands âges.

Cinquième partie

Application dans un portefeuille de rentes viagères

Dans les parties précédentes, nous avons vu comment mettre en place une table de mortalité d'expérience, de plus, nous avons comparé les modèles à l'aide du calcul des SMR. Toutefois, le choix du modèle qui explique rigoureusement nos données s'avère être un processus très délicat.

C'est pourquoi nous allons calculer les capitaux constitutifs avec les modèles utilisés dans la partie précédente, les données brutes et avec la table réglementaire de la population masculine afin de nous conforter sur notre choix du modèle adéquat.

1 Définition des capitaux constitutifs

A partir de la liquidation de sa retraite, l'assuré percevra, s'il est en vie, sa première prestation dès la fin de la première période. Par la suite, à la fin de chaque période, l'assureur effectuera des paiements, à condition que l'assuré soit en vie à chaque échéance.

Nous allons dorénavant estimer les capitaux à constituer pour une rente entre 2009 et 2017 à l'aide des tables de mortalité d'expérience, des tables réglementaires et de nos données.

Le coût du versement d'un euro de rente annuelle à un rentier à partir d'un âge x ans est défini comme :

$$C_x = \sum_{k=0}^w v^k {}_k p_x \quad (71)$$

où ${}_k p_x$ représente la probabilité pour un individu d'âge x d'être en vie dans k années, v^k , le facteur d'actualisation tel que $v = \frac{1}{1+i}$ où i , le taux d'actualisation.

2 Application

La maîtrise fine du risque de longévité permet à l'assureur d'anticiper toute déviation de la mortalité de son portefeuille de rentiers, mais surtout d'évaluer au plus près de la réalité les coûts de ses prestations futures.

Nous calculons le coût du versement d'une rente entre 63 et 94 ans afin de pouvoir comparer les résultats avec nos données. Nous utilisons comme taux d'actualisation, la moyenne pondérée par les montants des rentes des taux techniques observés sur l'ensemble du portefeuille. Voici les résultats obtenus pour la population masculine :

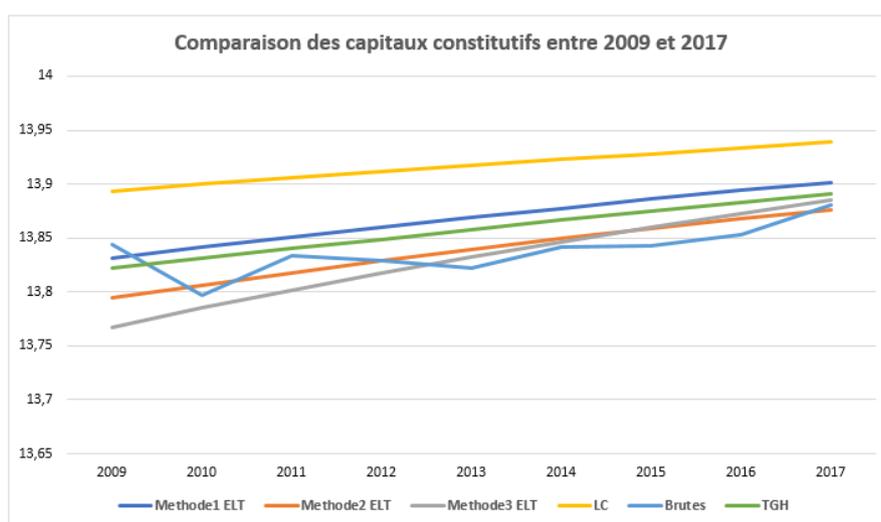


FIGURE 90 – Comparaison des capitaux constitutifs obtenus avec taux bruts, tables réglementaires et tables d'expérience

Nous remarquons l'augmentation des capitaux constitutifs au cours du temps ce qui est cohérent avec l'augmentation de la durée de vie des rentiers et ainsi de leurs probabilités de survie.

A l'exception du modèle Lee Carter, les résultats obtenus avec les tables d'expériences sont proches de nos données brutes. La méthode 2 a le double avantages d'être proche des données brutes mais aussi avec la table réglementaire, et avec une convergence vers les données brutes au cours du temps. Le coût estimé par la méthode 3 augmente d'une façon très significative au cours du temps (la pente est très forte).

Au vu de l'ensemble des résultats, nous conservons les tables d'expériences correspondant au deuxième modèle d'ajustement (Méthode 2 : modèle semi paramétrique de Brass)

avec pour référence de mortalité la table IA2013. Ce choix est motivé par deux raisons :

- Nous souhaitons conserver une marge de prudence dans nos estimations. Notre choix s’oriente alors vers le deuxième modèle (courbes vertes) qui passe par les maximums observés pour la population féminine et masculine. Cette prudence mesurée, se trouve également dans l’étude des SMR.
- cette méthode présente de meilleures estimations des capitaux constitutifs et est proche des coûts estimés par les données brutes.

En réalisant des tables de mortalité d’expérience prospectives, nous pouvons estimer le coût d’un euro d’une rente jusqu’en 2030. Voici les résultats obtenus avec les méthodes sélectionnées.

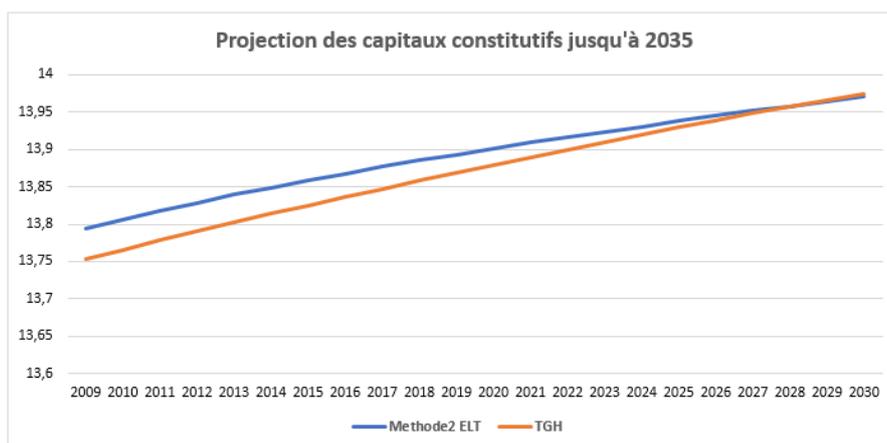


FIGURE 91 – Projection des capitaux constitutifs jusqu’en 2030 avec la méthode 2 et TGH

Nous constatons au travers de ces projections une amélioration constante des taux de décès. Nous remarquons également que les tables réglementaires utilisées dans le cadre du provisionnement et de l’estimation des engagements resteront prudentes à long terme. Les tables d’expérience ainsi construites permettent le suivi technique de longévité de notre population assurée. Nous remarquons également que même si le coût des capitaux obtenus en utilisant la table d’expérience construite présente un écart sensiblement différent de celui obtenu à partir des tables réglementaires, cet écart tend à diminuer fortement à horizon de 30 ans. L’évolution des taux de décès de notre population est à surveiller, en intégrant au fur et à mesure les nouvelles années d’expérience.

Conclusions

Dans un premier temps, l'objectif de notre étude était de projeter selon une approche fiable et rigoureuse la mortalité d'un portefeuille de rentes. Nous avons vu que pour l'assureur, le risque d'un tel portefeuille est difficile à appréhender principalement à cause de l'allongement de l'espérance de vie. Il devient alors nécessaire de déterminer la mortalité spécifique du portefeuille. L'assureur pourra se servir par la suite de cette table de mortalité pour la projection de ses prestations futures et la tarification des rentes, sous certaines conditions.

Nous avons mis en compétition deux modèles d'estimation du taux brut afin de choisir celui qui correspond le mieux à nos données. Le modèle traditionnel de Lee Carter a permis une première approche prospective. Cependant, les résultats obtenus étant peu satisfaisants, parce que nous n'avons pas suffisamment de données, nous avons donc appliqué les modèles de positionnement via le package ELT à savoir la méthode 1 (Approche à un paramètre avec le SMR), la méthode 2 (Approche à deux paramètres avec un modèle relationnel) et la méthode 3 (Modèle linéaire généralisé de Poisson) développés dans le cadre d'un groupe de travail de l'IA sur la mortalité.

Dans notre travail, les critères de validation que nous avons retenus étaient principalement basés d'une part sur le SMR et la qualité du lissage réalisé, et d'autre part sur la cohérence des projections obtenues par le modèle implémenté. Cette cohérence est appréhendée au travers de l'espérance de vie résiduelle. Les taux de mortalité prédits par le modèle retenu, serviront pour la projection des rentes à un horizon donné. La table de mortalité obtenue sera donc le point d'entrée dans la construction d'un outil de pilotage des rentes viagères.

Une analyse a priori nous aurait laissé penser que la méthode 3 par le package ELT, qui est un modèle linéaire généralisé de Poisson incluant des interactions avec l'âge et l'année calendaire, conduirait à de meilleurs résultats. L'étude confirme bien cet a priori. Toutefois, la méthode 2 présente des résultats similaires. Ceci peut s'expliquer par le fait que cette dernière présente de meilleurs résultats concernant l'évolution du SMR en plus nous avons éliminé la composante qui permet l'interaction entre l'âge et l'année calendaire dans le modèle du fait du manque d'effectif. Toutefois, le positionnement par rapport à une table externe est une approche qui permet de pallier à la limite des données. Lorsque les données le permettront, il serait alors intéressant d'utiliser un modèle endogène de type Bongaarts ou encore des modèles de positionnement permettant de prendre en compte l'interaction entre les âges et les périodes. De plus, nous avons vu que selon les âges étudiés, les tables de mortalité d'expérience se complètent. Il serait intéressant alors de construire une table

d'expérience comme combinaison des différentes tables obtenues par tranche d'âge. C'est notamment ce que réalise Bennet & al (2017) dans le cadre de la construction d'une table prospective à partir de 21 modèles de projection où chacun des modèles contribue aux projections finales.

Bibliographies

Olivier CLEMENT. Elaboration d'une table d'expérience : comparaison de méthodes de lissage analytique et d'ajustement statistique application à la population d'un régime de retraite marocain. Master's thesis, EURIA, 2003.

PETAUTON P.[2004]. Théorie et pratique de l'assurance vie. DUNOD.

DELWARDE.A,DENUIT.M[2006] Construction de tables de mortalité périodiques et prospectives.ECONOMICA.

PLANCHET F., THEROND.P[2006] Modèle de durée. ECONOMICA.

Cours Modèle de durée M. Olivier LOPEZ.

PLANCHET F. [2006] Tables de mortalité d'expérience pour des portefeuilles de rentes note de présentation, Document de travail, Institut des Actuaire.

PLANCHET F., THÉRON P.E. [2007], pilotage technique d'un régime de rentes viagères, Paris : Economica. [11] PLANCHET F. [2009] Risque de longévité et détermination du besoin en capital, Document de travail ISFA, laboratoire SAF.

B. GOMPERTZ. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London, 115 :513 – 585, 1825.

W.M. MAKEHAM. on the law of mortality and the construction of annuity tables. Journal of the Institute of Actuaries, 6 :301 – 310, 1860.

KLEIN J.P. , MOESCHBERGER M.L. [1997] Survival Analysis : Techniques for censored and truncated data

KONTIS V., BENNET J.E., MATHERS C. D., LI G., FOREMAN K., EZZATI M. [2017] Future life expectancy in 35 industrialised countries projections with a Bayesian model ensemble Lancet, vol. 389 no. 10076 p1323-1335

LANGMEIER R. [2000] Étude de différentes méthodes d'ajustement de tables de mortalité : application aux données d'une compagnie d'assurance.

LEE R. [2000] The Lee-Carter method for forecasting mortality with various extensions and applications North American Actuarial Journal vol. 4 no.1

LELIEUR V. [2005] Utilisation des méthodes de Lee Carter et Log Poisson pour l'ajustement de tables de mortalité dans le cas de petits échantillons

LELIEUR V., PLANCHET F. [2007] Utilisation des méthodes Lee Carter et log Poisson pour l'ajustement de tables de mortalité dans le cas de petits échantillons

LIDDELL F.D.F [1984] Simple exact analysis of the standardised mortality ratio Journal of Epidemiology and Community Health, 1984, 38, 85-88

MUSSET A.S [2016] Risque de longévité : la référence à a tendance de la population nationale est-elle justifiée ?

PLANCHET F. [2006] Tables de mortalité d'expérience pour les portefeuilles de rentiers (Tables TGH 05 et TGF 05) Notice de présentation

PLANCHET F. [2017a] Méthodes de lissage et d'ajustement Support de cours disponible sur le site : ressources-actuarielles

RENSHAW A., HABERMAN S. [2005] Mortality reduction factors incorporating cohort effects Actuarial research paper (case business school)

Sites

Statistiques OCDE : <https://data.oecd.org/fr>

Statistiques INSEE : <https://www.insee.fr/fr/statistiques>

Cours modèles de durée : <https://www.ressources-actuarielles>

Tracer des courbes (histogramme, density...) : https://stt4230.rbind.io/tutoriels_

etudiants/hiver_2015/histogramme_ggplot2/ freakonometrics : <https://freakonometrics>.

[hypotheses.org/tag/lee-carter](https://freakonometrics)