

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

Par : **ASSAYAG Ilan**

Titre du mémoire : Le risque de longévité dans le cadre d'un viager immobilier

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

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

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

*Entreprise : **Generali Vie***

Nom : **Mme. FOUGEA Elvire**

Signature : 

Directeur de mémoire en entreprise :

*Membres présents du jury de la
filière*

Nom : **Mme. QUINOL Sandra**

Signature : 

Invité :

Nom :

Signature :

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

Signature du responsable entreprise



Signature du candidat



Résumé

Le viager immobilier est une transaction où l'acheteur d'un bien promet en contrepartie le versement d'une somme d'argent. Une rente viagère est communément associée à ce versement. La vente de viager expose l'acheteur du bien à des risques dont le risque de longévité.

En effet, les rentes viagères sont encadrées par un cadre légal et l'utilisation d'une table de mortalité visant à prédire l'espérance de vie du rentier fait partie des obligations associées. Le risque de longévité intervient dès lors où la durée de vie du rentier dépasse l'espérance de vie estimée à partir de cette table de mortalité.

L'un des objectifs de ce mémoire consiste à créer une table de mortalité prospective pour les individus de plus de 60 ans réalisée à partir des données du portefeuille de l'assureur. Après estimation des taux bruts de mortalité, plusieurs méthodes ont été appliquées afin de modéliser ces taux de décès sous plusieurs points de vue. Des méthodes de positionnement à partir d'une table de référence, un modèle de projection Cairns-Blake-Dowd et une régression logistique basée sur plusieurs des variables du portefeuille ont été appliqués. Des méthodes de fermeture de table ont également été appliquées sur les grands âges.

Une comparaison entre ces modèles a permis de faire une sélection des meilleurs modèles. L'effet de l'utilisation de l'une de ces tables à travers une rente viagère classique a été étudié et comparé aux résultats obtenus par la table réglementaire.

Mots-clefs : viager, mortalité, risque de longévité, senior, table prospective, Kaplan-Meier, Hoem, Cairns-Blake-Dowd, positionnement, fermeture de table, régression logistique.

Abstract

A “viager” is a transaction where the buyer of an asset promises to pay a sum of money in return. A life annuity is commonly associated with this payment. The sale of a “viager” exposes the purchaser of the property to risks including longevity risk.

Indeed, the life annuities are framed by a legal framework and the use of a mortality table aiming at predicting the life expectancy of the annuitant is part of the associated obligations. The longevity risk arises when the life expectancy of the annuitant exceeds the life expectancy estimated from the mortality table.

One of the objectives of this thesis is to create a prospective mortality table for individuals over 60 years of age based on data from the insurer's portfolio. After estimating the crude mortality rates, several methods were applied to model these death rates from several perspectives. Positioning methods based on a reference table, a Cairns-Blake-Dowd projection model and a logistic regression based on several of the portfolio variables were applied. Table closure methods were also applied on the older ages.

A comparison between these models allowed a selection of the best models. The effect of using one of these tables through a conventional life annuity was studied and compared to the results obtained by the regulatory table.

Keywords: life annuity, mortality, senior, longevity risk, prospective table, Kaplan-Meier, Hoem, Cairns-Blake-Dowd, positioning, table closure, logistic regression.

Note de synthèse

Le présent document s'articule autour du risque de longévité dans le cadre d'une vente de viager. L'association du viager immobilier à la rente viagère expose l'acheteur du bien à des risques. La rente viagère est associée légalement à une estimation de l'espérance de vie du rentier. Dans un contexte où la durée de vie du rentier excède l'espérance de vie estimée, l'acheteur paiera la rente plus longtemps que prévu et sera donc perdant : c'est le risque de longévité.

Contexte actuel et motivation

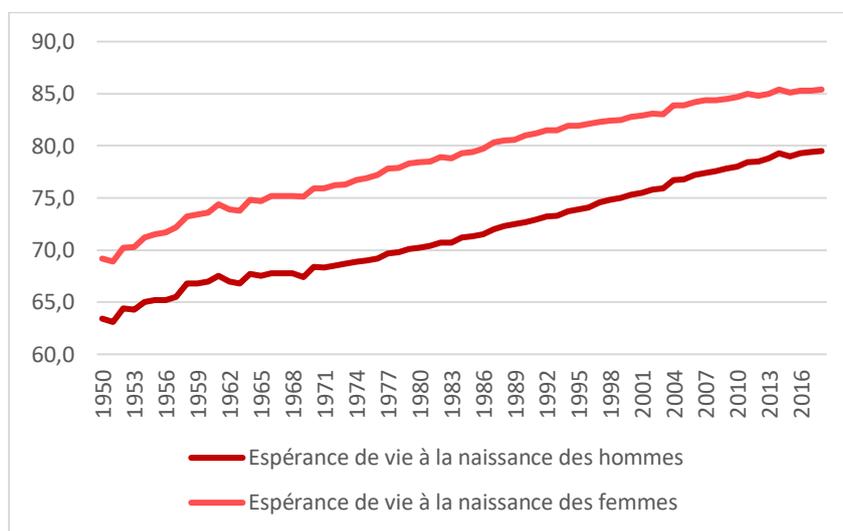


Figure 0-1 : Evolution des espérances de vie à la naissance en France depuis 1950 à partir des tables de l'INSEE

L'allongement de la durée de vie humaine est un phénomène qui ne peut laisser indifférent. Il accentue, d'années en années, le risque de longévité porté par les assureurs pour qui estimer l'espérance de vie de façon précise se complexifie. Cet allongement est visible sur le graphique ci-dessus représentant l'évolution de l'espérance de vie à la naissance par sexe en France, de 1950 à 2016. Cette évolution semble croître assez linéairement et il est difficile de prédire si cette tendance se maintiendra par la suite.

A l'heure actuelle, la mise en place de rente viagère est réglementée par un cadre légal. Plusieurs facteurs ne peuvent être laissés à la bienveillance de l'assureur. L'estimation de l'espérance de vie est réglementée par le biais de l'utilisation de tables de mortalité précises. Le jeu de tables de mortalité générationnelles de 2005 (TGF05 et TGH05) est le jeu de tables en vigueur pour les calculs de rentes viagères. Néanmoins, la réglementation laisse une certaine marge de manœuvre à l'assureur en autorisant l'utilisation de table d'expérience à condition qu'une certification ait été préalablement réalisée. L'utilisation d'une table d'expérience a pour objectif d'augmenter la qualité d'estimation de l'espérance de vie et donc de diminuer le risque de longévité ; le tout en se basant sur le portefeuille que l'assureur a à sa disposition.

C'est dans ce contexte qu'a été trouvée la motivation de construire une table de mortalité prospective d'expérience, axée sur les seniors, à partir du portefeuille épargne de Generali vie.

Modélisation des taux de mortalité

La base de données utilisée représente le portefeuille épargne Generali vie, sur une tranche d'observations de 10 ans et orientée sur les plus de 60 ans uniquement. Cette base est séparée en deux par la variable sexe afin de dissocier la modélisation pour chacun des sexes.

Le calcul des expositions au risque et des taux bruts de mortalité se basent sur une décomposition temporelle par période ; se distinguant de la décomposition par cohorte (Planchet et Guilbert, 2017).

Deux estimateurs des taux bruts de mortalité sont appliqués au portefeuille : l'estimateur d'Hoem et l'estimateur de Kaplan-Meier.

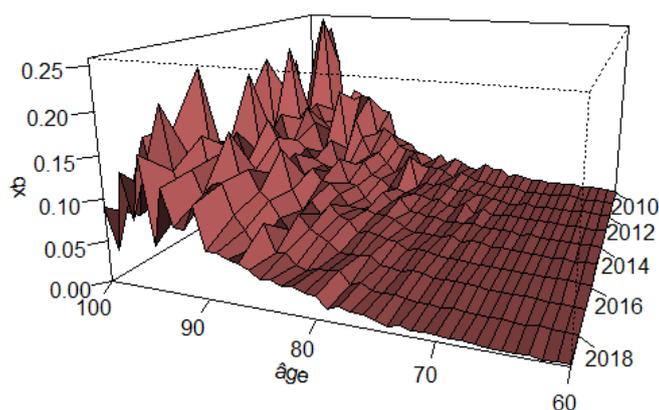


Figure 0-2 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les hommes

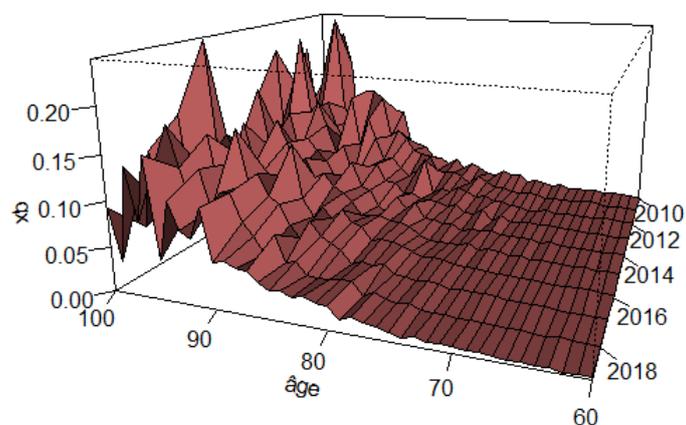


Figure 0-3 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les hommes

Bien que les deux estimateurs fournissent des résultats assez proches dans l'ensemble, la comparaison par la variance de ces deux estimateurs favorise l'estimateur d'Hoem. C'est cet estimateur qui est utilisé dans la suite de l'étude.

Dans un premier temps, c'est un modèle stochastique de projection de la mortalité qui a été étudié. Le modèle retenu s'inscrit dans la continuité du modèle de Lee-Carter (1992) puisque Cairns, Blake et Dowd reprennent, 14 ans plus tard, plusieurs de ces principes pour construire un modèle d'estimation des *logits* des taux de mortalité à partir de deux paramètres. Ce modèle se base sur les taux de mortalités bruts pour le lissage des premiers taux.

Quatre modèles suivent Cairns-Blake-Dowd, ces quatre modèles se basent sur les mêmes principes : les informations du portefeuille sont liées d'une manière ou d'une autre à une table de référence, en l'occurrence la TGH05 pour les hommes et TGF05 pour les femmes. Ces méthodes sont appelées des méthodes de positionnement.

La première méthode consiste à appliquer un coefficient d'ajustement à la table de référence ; ce coefficient se base sur des données du portefeuille. C'est le positionnement par *Standardized Mortality Ratio* (SMR).

En second lieu, la méthode appliquée est un modèle relationnel à deux paramètres. Ce modèle, variant du modèle régressif de Brass (1971), est également un modèle régressif où le *logit* des taux de mortalité est basé sur deux paramètres. L'un de ces paramètres se base sur les taux de mortalité bruts (calculés par l'estimateur d'Hoem par période) et le second est appliqué à la table de référence.

La troisième méthode de positionnement se base sur l'idée que les décès du portefeuille suivent une loi de Poisson avec un paramètre associé aux taux de risque instantanés. Les taux de risques instantanés sont ensuite exprimés en fonction de plusieurs paramètres dépendants de l'âge, de l'année et des taux de référence.

La méthode clôturant cette partie positionnement est un modèle non-paramétrique. Ce modèle applique aux taux bruts une fonction de lissage. Cette fonction sera ensuite soumise à une dérive de mortalité liée aux taux issus de la table de référence. Cette méthode est à prendre avec des pincettes puisque c'est une méthode assez complexe, qui laisse une grande place à l'interprétation lors de son application. L'estimation des paramètres de la fonction de lissage est notamment une tâche laborieuse.

L'ensemble de ces quatre méthodes sont soumises à des tests de validation axés sur deux critères : la proximité aux données brutes et la régularité de l'ajustement.

Les modèles obtenus par méthode de positionnement, mais aussi par modélisation stochastique (Cairns-Blake-Dowd), se voient appliquer une fermeture de table. Extrapoler les taux de décès aux grands âges par l'une de ces méthodes permet de réduire les biais d'ajustement liés à la volatilité des données aux âges les plus élevés. Une analyse des résultats obtenus sur ces modèles permet de retenir pour chacun d'entre eux l'âge à partir duquel sera appliquée l'extrapolation.

Quatre méthodes de fermeture de table sont introduites, appliquées et comparées pour chacun des modèles. Une comparaison par proximité, à partir de quantités (X^2 , déviance...) aux données brutes permet d'obtenir l'extrapolation la plus adaptée à chaque modèle. Un autre critère qui a influé sur le choix est le niveau de surestimation de la mortalité ; compte tenu de l'allongement de la longévité au fil des années, il convient que la mortalité soit au minimum surestimée par la méthode utilisée. Les fermetures de tables abordées sont : une extrapolation exponentielle, la méthode de Denuit & Goderniaux, la fermeture de Coale-Kisker et Kannistö.

Dans l'ensemble, l'extrapolation exponentielle fournit de bons résultats et la méthode de Denuit & Goderniaux est souvent associée aux méthodes de positionnement appliquées à la population féminine. Néanmoins, toutes les fermetures ont été utilisées au moins une fois.

L'un des souhaits de ce mémoire est également de pouvoir mettre à contribution différentes variables qui influent d'une manière ou d'une autre sur la mortalité dans un modèle calculant des taux de mortalité. Le portefeuille permet d'utiliser plusieurs variables qui peuvent être explicatives des taux de décès ; c'est pour cette raison qu'une régression logistique est réalisée. Cette régression se base sur une multitude de variables internes et externes au portefeuille et permet d'obtenir des taux de décès dépendants de ces variables. La censure a été intégrée dans ce modèle par le biais de la prise en compte de l'exposition au risque en tant qu'*offset* sur cette régression. Une sélection des variables les plus utiles au modèle a été faite à partir de méthodes de sélection *Backward*, *Forward*, *Stepwise*. Les résultats des prédictions de ce modèle sont bons ; mais à prendre avec attention. En effet, deux classes sont comparées : vie et décès ; mais l'une d'entre elles est fortement disproportionnée et c'est pour cette raison que certains estimateurs de qualité donnent des résultats anormalement élevés.

L'utilisation de ce modèle semble par ailleurs difficile. La quantité d'informations demandée pour chacune des estimations n'est pas applicable. De plus, le cadre légal interdit de proposer un tarif différent par rapport à certaines variables ; ce qui accroît la complexité d'utilisation.

Comparaison des modèles et applications

Dès lors que l'ensemble des modèles sont finalisés, une comparaison est nécessaire afin de se rendre compte des niveaux de fiabilité de chacun.

Dans le but de comparer ces modèles par critère de proximité aux données brutes, des quantités sont calculées pour l'ensemble des modèles obtenus avec fermeture de table. Ces comparaisons visent à jauger à quel point les taux lissés et extrapolés sont éloignés de la réalité du portefeuille. La comparaison se fait donc sur les 100 premiers âges et sur les 10 premières années d'observations ; les projections au-delà de ces âges et années ne peuvent être comparées à la réalité. La proximité entre le nombre de décès estimé et le nombre de décès brut est également étudiée. Ce paramètre permet de prendre en compte l'application des taux à l'exposition au risque et donc à comprendre les caractéristiques de chaque modèle.

Un autre critère de comparaison est le niveau de risque obtenu pour chacun des modèles. Comme pour la comparaison des fermetures de tables, il est important que la mortalité ne soit pas abondamment surestimée. L'allongement de la durée de vie humaine peut possiblement perdurer dans le temps et il est donc important d'obtenir une table qui fait prendre peu de risque à l'assureur.

Dans l'ensemble, les modèles issus de la population féminine du portefeuille offrent de meilleurs résultats. Certaines de ces tables sont ensuite appliquées au calcul d'une rente viagère classique. En considérant deux individus : l'Individu 1 est un assuré né en 1970 qui voit sa rente débiter à 60 ans ; l'Individu 2 est un assuré né en 1970 qui voit sa rente débiter à 90 ans. Le calcul des rentes pour un capital constitutif de 100 000€ et pour 4 différentes tables de mortalité donne les résultats suivants :

Méthode	Positionnement : relationnel F	Positionnement : lissage non- paramétrique F	Cairns-Blake- Dowd F	TGF-05
Individu 1	2 971,33 €	2 839,30 €	2 884,50 €	2 861,73 €
Individu 2	12 267,36 €	9 627,01 €	9 259,76 €	10 752,43 €

Tableau 1 : Comparaison des montants de rente par individu et modèle (F=femme)

L'application de ces tables « d'expériences » offrent des résultats différents mais pas complètement éloignés de la TGF05.

C'est de cette façon qu'a été pensée la construction d'une table de mortalité prospective « d'expérience ». D'autres façons de limiter ce risque de longévité existent. Ces points sont également abordés dans le cadre de cette étude.

Executive summary

The present document is articulated around the risk of longevity within the framework of a sale of a “viager”. The combination of the “viager” with a life annuity exposes the buyer of the property to risk. The life annuity is legally associated with an estimate of the life expectancy of the annuitant. In a context where the life expectancy of the annuitant exceeds the estimated life expectancy, the buyer will pay the annuity longer than expected and will therefore lose out: this is the longevity risk.

Current context and motivation

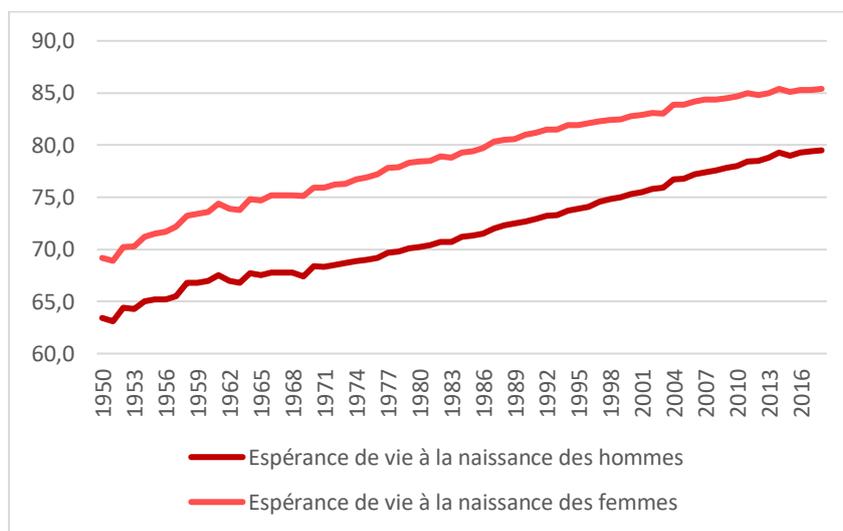


Figure 0-4 : Evolution of life expectancy at birth in France since 1950, based on INSEE tables

The increase in human life expectancy is a phenomenon that cannot leave anyone indifferent. Year after year, it accentuates the longevity risk borne by insurers, for whom estimating life expectancy accurately is becoming increasingly complex. This increase can be seen in the graph above representing the evolution of life expectancy at birth by sex in France from 1950 to 2016. This evolution seems to grow quite linearly, and it is difficult to predict whether this trend will continue thereafter.

At present, the establishment of life annuities is regulated by a legal framework. Several factors cannot be left to the benevolence of the insurer. The estimation of life expectancy is regulated with precise mortality tables. The 2005 set of generational mortality tables (TGF05 and TGH05) is the current set of tables for life annuity calculations. Nevertheless, the regulations leave some leeway to the insurer by allowing the use of experience tables provided that a certification has been carried out beforehand. The purpose of using an experience table is to increase the quality of the life expectancy estimate and therefore to reduce the longevity risk; all this is based on the portfolio that the insurer has at his disposal.

It is in this context that the motivation was found to build a prospective mortality table of experience, focused on seniors, from the savings portfolio of Generali vie.

Mortality rates modelling

The database used represents the Generali vie savings portfolio, over a 10-year observation period and focused on the over-60s only. This database is separated in two by the gender variable in order to dissociate the modelling for each gender.

The calculation of risk exposures and crude mortality rates is based on a temporal decomposition by period; this differs from the cohort decomposition (Planchet and Guilbert, 2017).

Two estimators of the crude death rates are applied to the portfolio: the Hoem estimator and the Kaplan-Meier estimator.

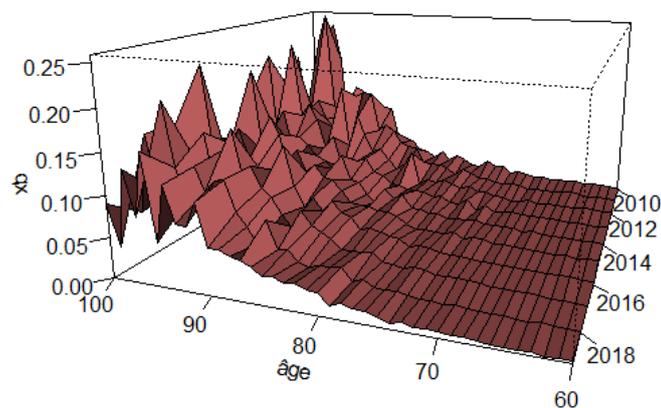


Figure 0-5 : Death rates of the Hoem estimator by age and year for men

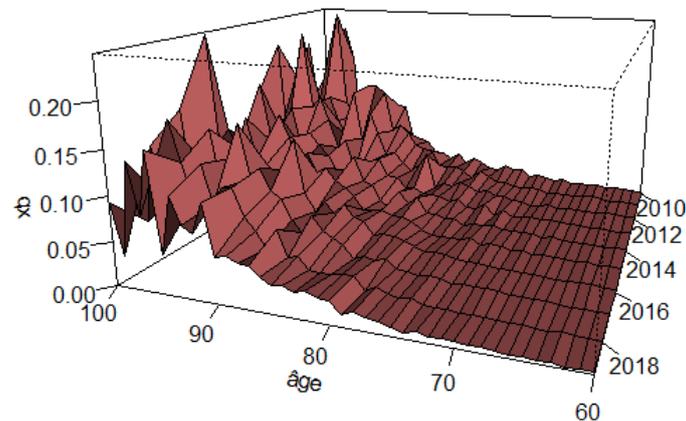


Figure 0-6 : Death rates of the Kaplan-Meier estimator by age and year for men

Although the two estimators provide close results overall, the comparison by variance of these two estimators favors the Hoem estimator. It is this estimator that is used in the rest of the study.

Initially, a stochastic model of mortality projection was studied. The model chosen is in line with the Lee-Carter model (1992) since Cairns, Blake and Dowd, 14 years later, took up several of its

principles to construct a model for estimating logit mortality rates from two parameters. This model is based on crude mortality rates for the smoothing of the first rates.

Four models follow Cairns-Blake-Dowd, these four models are based on the same principles: the portfolio information is linked in one way or another to a reference table, in this case the TGH05 for men and TGF05 for women. These methods are called positioning methods.

The first method consists of applying an adjustment coefficient to the reference table; this coefficient is based on portfolio data. This is the Standardized Mortality Ratio (SMR) positioning.

Secondly, the method applied is a relational model with two parameters. This model, which varies from the regressive model of Brass (1971), is also a regressive model where the logit of mortality rates is based on two parameters. One of these parameters is based on the crude mortality rates (calculated by Hoem's period estimator) and the second is applied to the reference table.

The third positioning method is based on the idea that the portfolio deaths follow a Poisson distribution with a parameter associated with the instantaneous hazard rates. The instantaneous hazard rates are then expressed as a function of several parameters depending on the age, the year and the reference rates.

The method closing this positioning section is a non-parametric model. This model applies a smoothing function to the raw rates. This function will then be subjected to a drift of mortality related to the rates resulting from the reference table. This method is to be taken with precautions since it is a rather complex method, which leaves a great place to the interpretation during its application. The estimation of the parameters of the smoothing function is a laborious task.

All these four methods are subjected to validation tests based on two criteria: proximity to the raw data and regularity of fit.

The models obtained by the positioning method, but also by stochastic modeling (Cairns-Blake-Dowd), are subjected to a table closure. Extrapolating death rates to older ages by one of these methods reduces the bias in the fit due to the volatility of the data at older ages. An analysis of the results obtained with these models makes it possible to select the age at which the extrapolation will be applied.

Four table closure methods are introduced, applied, and compared for each model. A comparison by proximity, based on quantities (X^2 , deviance...) to the raw data allows to obtain the most adapted extrapolation for each model. Another criterion which influenced the choice is the level of overestimation of mortality; considering the increase in longevity over the years, it is advisable that mortality is at least overestimated by the method used. The table closures discussed are an exponential extrapolation, the Denuit & Goderniaux method, the Coale-Kisker and Kannistö closure.

Overall, exponential extrapolation provides good results and the Denuit & Goderniaux method is often associated with positioning methods applied to the female population. Nevertheless, all closures have been used at least once.

One of the wishes of this dissertation is also to be able to bring different variables that affect mortality in some way into a model calculating mortality rates. The portfolio allows for the use of

several variables that may be explanatory of death rates; for this reason, a logistic regression is performed. This regression is based on a multitude of variables internal and external to the portfolio and allows to obtain death rates depending on these variables. Censoring has been incorporated into this model through the inclusion of risk exposure as an offset on this regression. A selection of the most useful variables for the model was made using Backward, Forward, Stepwise selection methods. The results of the predictions of this model are good but should be taken with care. Indeed, two classes are compared: life and death; but one of them is strongly disproportionate and therefore some quality estimators give abnormally high results.

The use of this model also seems difficult. The amount of information required for each estimate is not applicable. In addition, the legal framework prohibits proposing a different rate for certain variables, which increases the complexity of use.

Comparison of models and applications

Once all the models are finalized, a comparison is necessary in order to realize the levels of reliability of each one.

To compare these models by proximity criteria to the raw data, quantities are calculated for all the models obtained with table closure. These comparisons aim to gauge how far the smoothed and extrapolated rates are from the reality of the portfolio. The comparison is therefore made over the first 100 ages and the first 10 years of observations; projections beyond these ages and years cannot be compared with reality. The proximity between the estimated number of deaths and the gross number of deaths is also studied. This parameter makes it possible to take into account the application of rates to risk exposure and thus to understand the characteristics of each model.

Another comparison criterion is the level of risk obtained for each of the models. As with the comparison of table closures, it is important that mortality is not overestimated. The increase in human lifespan may possibly persist over time, so it is important to obtain a table that takes little risk for the insurer.

Overall, models derived from the female population in the portfolio provide better results. Some of these tables are then applied to the calculation of a conventional life annuity. Considering two individuals: Individual 1 is an insured born in 1970 whose annuity starts at age 60; Individual 2 is an insured born in 1970 whose annuity starts at age 90. The calculation of the annuities for a constitutive capital of 100 000€ and for 4 different mortality tables gives the following results:

Method	Positioning : relational F	Positioning : non-parametric smoothing F	Cairns-Blake-Dowd F	TGF-05
Person 1	2 971,33 €	2 839,30 €	2 884,50 €	2 861,73 €
Person 2	12 267,36 €	9 627,01 €	9 259,76 €	10 752,43 €

Table 2 : Comparison of pension amounts per individual and model (F=woman)

The application of these "experience" tables offer different results but not completely different from the TGF05.

It is in this way that the construction of a prospective mortality table "of experience" was thought. Other ways to limit this longevity risk exist. These are also discussed in this study.

Remerciements

Je tiens tout d'abord à remercier Madame Elvire FOUGEA et l'ensemble de l'équipe Epargne de la Technique Assurance Generali VIE qu'elle dirige de m'avoir accueilli en tant qu'alternant.

Plus particulièrement, je tiens à remercier Madame QUINOL Sandra, responsable d'études actuarielles, qui a également été ma tutrice au cours de cette alternance. Sandra m'a apporté son expertise et sa pédagogie, me permettant d'assimiler rapidement les enjeux de ce mémoire et d'avoir un suivi régulier. Elle a également rendu ce mémoire possible en l'orientant vers une problématique d'intérêt.

Au sein même de cette équipe, Madame MOUQUET Laura, Madame Géraldine PARAT et Monsieur SOULAS Thomas ont notamment été d'un grand soutien de par leur disponibilité et leur maîtrise des différents sujets d'études. Au même titre, je remercie Monsieur DU GRAND PLACITRE Damien et Madame FIANT Emma avec qui les différentes réflexions concernant la modélisation ont été d'une grande aide.

Je tiens également à remercier ma tutrice académique Madame DION-BLANC Charlotte pour l'ensemble des conseils techniques permettant de réaliser ce mémoire. L'ensemble de l'équipe pédagogique de l'Institut de Statistiques de l'Université de Paris m'a apporté les connaissances de bases nécessaires à l'élaboration de ce mémoire.

Sans oublier ma famille et mes proches que je remercie pour leur soutien indispensable. Madame BELLAICHE Jessica a notamment été d'un grand soutien au quotidien.

Introduction

La vente de viager est un mécanisme qui permet à un propriétaire de se servir de son bien pour en tirer une somme d'argent. Plusieurs types de viagers existent mais dans le cas général une rente est associée à cette vente, en faveur du vendeur. Pour les personnes âgées, le viager peut donc être vu comme une bonne alternative aux retraites complémentaires.

Associer le viager à une rente viagère constitue un risque pour l'assureur. En effet, prédire l'âge de décès est d'une importante complexité aux vues de plusieurs critères. Les actuaires calculent l'espérance de vie du rentier à partir de tables de mortalité réglementaires. Cette espérance rentre en considération dans la tarification ou dans le calcul de rente. Si un assuré vient à dépasser l'espérance de vie qui lui a été prédit, le risque est que l'assureur doive payer la rente plus longtemps que ce qui a été prévu au départ. L'assureur sera donc en perte du fait d'une tarification biaisée par un mauvais calcul d'espérance de vie. C'est le risque de longévité qui est porté par l'assureur.

Dans cet environnement de risque pour l'assureur, de quelle manière est-il possible de limiter le risque de longévité ?

Le but de cette étude sera donc de construire une table de mortalité prospective en respectant deux grands critères : fiabilité et prudence. Cette table de mortalité portera sur les individus de plus de 60 ans.

Ce mémoire s'articulera donc en 4 grands axes. Le premier concernera en partie le mécanisme du viager et tout ce qui l'entoure. Les aspects réglementaires et fiscaux y seront notamment abordés. Tandis que la seconde partie sera orientée vers le risque de longévité, introduisant la théorie actuarielle des modèles de durées et le risque de longévité associé à la population française.

Dans le troisième grand axe, il conviendra de parcourir plusieurs méthodes de modélisations possibles permettant de mettre en place une table de mortalité prospective. Ces modèles se baseront sur le portefeuille client de la section épargne de Generali vie. Les données à disposition y seront présentes et elles permettront de mettre en place différents types de modèles. Ces modèles seront utilisés dans la dernière partie puisque celle-ci sera comparative des différentes méthodes. Les tables y seront également appliquées à la tarification afin de se rendre compte de l'effet

En parallèle de ces différents axes, un nouveau type de vente de viager sera présenté. Ce nouveau mécanisme permet à l'assureur de limiter le risque de longévité tout en restant intéressant pour le propriétaire. Une étude approfondie de ce viager sera faite au cours de l'étude.

Table des matières

Introduction.....	0
I. Le viager immobilier.....	1
A. Présentation et fonctionnement.....	1
B. Risque et cadre légal	2
C. Un nouveau mécanisme pour le viager ?	3
II. Notions de base de modèle de durée	6
A. Notations actuarielles et notion de censure.....	6
B. Les méthodes de fermeture de table	8
1. L'extrapolation exponentielle	8
2. Denuit & Goderniaux	9
3. Coale-Kisker.....	9
4. Kannistö	10
C. Longévité et population française.....	10
D. Risque de longévité en assurance et tables d'expérience.....	14
III. Données utilisées	16
A. Présentation de la base	16
B. Prétraitement.....	17
C. Les données manquantes	18
D. Valeurs aberrantes et traitement variable par variable	19
E. Statistiques descriptives	20
IV. Construction d'une table de mortalité prospective d'expérience	29
A. Validité par critère de Cochran	30
B. Exposition au risque et calcul des taux bruts de mortalité.....	31
1. L'estimateur binomial.....	33
2. L'estimateur d'Hoem du moment	33
3. L'estimateur de Kaplan-Meier.....	34

4.	Comparaison des méthodes	36
C.	Modèles stochastiques de projection de la mortalité.....	40
1.	Lee-Carter	40
2.	Cairns-Blake-Dowd	40
D.	Méthodes de positionnement à partir d'une table existante	56
1.	Positionnement par <i>Standardized Mortality Ratio</i> (SMR)	56
2.	Modèle relationnel à deux paramètres	66
3.	Modèle linéaire généralisé de Poisson	68
4.	Lissage non-paramétrique et dérive de mortalité.....	71
E.	Estimation par régression logistique.....	75
1.	Théorie.....	75
2.	Prérequis.....	76
3.	Application	77
V.	Comparaisons et application à la tarification	80
A.	Adéquation aux données.....	81
B.	Niveaux de risques	82
C.	Application aux calculs de rentes et autres comparaisons.....	83
	Conclusion	86
	Bibliographie	88
	Annexes.....	89

I. Le viager immobilier

A. Présentation et fonctionnement

Le principe du viager s'apparente à une vente de bien immobilier classique. Le propriétaire d'un bien immobilier met en vente ce bien, on l'appellera crédirentier. L'acheteur, ou débirentier, s'engage à verser en échange une somme d'argent au crédirentier. Ce versement peut apparaître sous plusieurs formes.

- Il peut contenir un « bouquet », c'est-à-dire un versement d'argent comptant au moment de la signature. Le bouquet est souvent la partie qui suscite le plus d'intérêts au crédirentier. La part du bouquet ne dépasse que rarement 50% de la valeur totale du bien.
- Le débirentier peut verser périodiquement une somme d'argent au propriétaire d'origine. Cette rente est le plus souvent de nature viagère, elle est versée jusqu'au décès du crédirentier. Il y a tout de même certaines exceptions dans lesquelles la transaction ne fait pas appel à une rente.

Un autre élément peut prendre une place importante lors de la mise en place de ce type de contrat : l'occupation éventuelle du bien immobilier. Il est donc important de bien faire la distinction entre un viager libre et occupé. Dans le premier cas, l'acquéreur récupère immédiatement le bien tandis que dans le second, le vendeur conserve son droit d'usage et d'habitation pour une durée déterminée ou non.

La variété des éléments qui composent cette transaction offre aux deux parties une multitude de choix lors de la mise en place du viager. Le vendeur du bien peut notamment adapter le paramétrage de la transaction pour satisfaire au mieux ses besoins. Parallèlement, l'acquéreur peut fixer ses exigences afin de jauger au mieux les risques encourus.

C'est à travers ces engagements réciproques entre crédirentiers et débirentiers que s'inscrit un cadre réglementaire. Le viager immobilier entre dans le cadre des contrats aléatoires d'après l'article 1968-1976 du code civil concernant les conditions requises pour la validité du contrat de rente viagère. Cette notion d'aléa est donc un élément important sur ce type de contrats. Malgré la multitude d'aménagements possibles lors de la mise en place d'une vente en viager, le montant total de la transaction doit demeurer inconnu dans le but de conserver le caractère aléatoire du contrat. C'est la seule « contrainte » d'aménagement observable.

B. Risque et cadre légal

Le viager offre des avantages au vendeur du bien. En effet, la fiscalité relative à cette transaction est avantageuse. L'imposition du bouquet viager pour le vendeur n'est pas soumise à la taxation des plus-values lorsqu'il est occupé au titre de résidence principale. A cela vient s'ajouter un abattement fiscal progressif en fonction de l'âge du vendeur :

Age	Abattement octroyé
50 ans et moins	30%
50 à 59 ans	50%
60 à 69 ans	60%
69 ans et plus	70%

Tableau 3 : Barème des abattements octroyés aux rentes issues de vente de viager

De plus, cette rente n'est pas considérée comme un loyer. Cet aspect intervient comme un avantage pour le crédientier qui n'est pas imposable à l'impôt foncier sur cette rente. C'est l'acquéreur qui a à sa charge la taxe foncière puisque l'occupation ne jouit que d'un droit d'usage et d'habitation.

Le prix de la transaction n'est quant à lui pas inscrit dans cadre légal très précis. Aucune loi ne permet de déterminer précisément le prix du bien au préalable. Plusieurs paramètres entrent toutefois en ligne de compte. A l'instar d'une transaction classique, le prix du bien ne peut pas se faire uniquement à partir de l'offre et de la demande. Le caractère aléatoire du contrat de viager fait intervenir des paramètres liés au bien en lui-même mais aussi à des paramètres propres à l'acheteur et au vendeur.

Dans un premier temps est fixée une valeur de base du bien, assimilable à la prime unique. Cette valeur peut être liée à plusieurs critères notamment le nombre de mètres carrés mais aussi le nombre de contrats de ce type élaborés chaque année. Ce montant est ensuite ajusté avec des caractéristiques propres au vendeur ; le sexe et l'âge notamment. Ces données sont d'autant plus importantes puisque la transaction fait intervenir une rente. Elles permettent donc d'estimer l'espérance de vie du vendeur et d'ajuster inversement proportionnellement le montant de la rente. Plus le vendeur vit longtemps, plus l'acheteur soutiendra la charge de la rente longtemps.

L'espérance de vie est donc un paramètre important de la transaction et influe sur la rentabilité du bien pour l'acheteur. Le calcul d'espérance de vie fait donc intervenir un élément essentiel : les tables de mortalité. Les tables de mortalité sont utilisées par les assureurs et viagéristes mais ces derniers mettent en avant la difficulté à estimer de façon précise l'espérance de vie. Les espérances de vie sur les âges avancés viennent à augmenter d'années en années. Ce facteur influe sur la justesse des tables de mortalité. Le risque de longévité est un phénomène qui touche indéniablement les

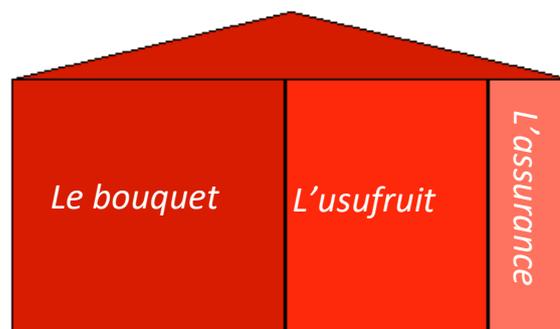
ventes immobilières de viager à cause de cette dépendance à des tables de mortalité qui sont parfois obsolètes.

De plus, le cadre légal mis en place pour ce type de transaction ne permet pas à l'acheteur en viager d'imposer de visite médicale ou d'avoir recours à des informations à des facteurs influant sur l'espérance de vie (comme la catégorie socio-professionnelle notamment). Cette disposition ne fait qu'accroître la dépendance aux tables de mortalité et est un véritable problème pour l'acheteur.

C. Un nouveau mécanisme pour le viager ?

Cette partie a pour objectif de présenter un nouveau mécanisme de vente de viager qui permettrait de limiter le risque porté par l'assureur.

Le viager immobilier présente des risques et inconvénients incontestables autant côté acheteur que vendeur. C'est dans le but de limiter et de faire disparaître certains de ces inconvénients qu'a été pensé un nouveau modèle de viager. Dans ce nouveau type de viager, la valeur du bien est décomposé en trois parties :



- Un **bouquet**.

C'est la part en cash résultant de la transaction, cette valeur correspond à la plus grande part de la valeur totale de la transaction.

- Un **usufruit temporaire** suivant un nombre d'années fixé.

Le nombre d'années est fixé en fonction de l'espérance de vie du crédientier. Cet usufruit permet au crédientier de conserver son droit de logement et de l'exploiter à son gré pendant le temps défini au préalable. L'usufruit possède une valeur décroissante au fil du temps qui devient nulle à la fin de la dernière année fixée. Cette part reste dans le patrimoine du vendeur et est transmissible et cessible aux ayants droits. En cas de décès prématuré du crédientier, la part restante de l'usufruit est versée comme complément de prix aux bénéficiaires.

- Une **assurance** associée à une rente viagère au moment de la transaction.

Au moment de la transaction, un contrat d'assurance est souscrit par le crédirentier. Il y voit son encours se valoriser comme pour un contrat d'épargne classique. L'encours de ce contrat est destiné à devenir une rente viagère. Cette rente viagère est différée et se déclenche dans deux situations :

- Il y a déclenchement des prestations de la rente si le crédirentier décide de quitter le bien immobilier. Le contrat sera liquidé et il y récupérera immédiatement le bénéfice lié à cette rente viagère.
- Le second processus de déclenchement est de nature temporel. La transaction étant liée à un contrat d'assurance ; celui-ci dure jusqu'aux 95 ans du crédirentier. Si le vendeur est toujours en vie à l'issue du contrat, les prestations commenceront à être émises. Dans ce cas précis, et puisque le crédirentier conserve son libre arbitre quant à l'occupation du domicile, la rente viagère différée s'assimilera à un loyer s'il devient acquéreur (qu'il occupe le bien) ou à un revenu complémentaire s'il devient assuré et ne souhaite pas rester sur place.

La rente est trimestrielle à terme échu et son montant est déterminé à partir du taux technique en vigueur au moment de la liquidation et de la table de mortalité en vigueur à ce moment (TGF05 actuellement). Si une seule tête est considérée, la prime P du contrat est unique et calculée à partir du montant de la rente R avec la formule suivante :

$$P = \frac{\left(R \times {}_nE_x \times \left(\frac{N_{x+n+1} + \frac{m-1}{2}}{D_{x+n}} \right) \right)}{1-\alpha}$$

Avec : x , l'âge de l'assuré à la date de souscription du contrat
 n , la durée du différé
 α , le taux de chargement sur prime
 m , la périodicité de la rente
 ${}_nE_x$, N_x et D_x seront abordés dans la section II.A

Ce mécanisme n'est visiblement pas assimilable à une vente de viager classique. Son fonctionnement atypique et novateur relève d'avantages incontestables.

- Tout d'abord côté vendeur :

Comme pour un viager classique, cette transaction permet au crédirentier de disposer immédiatement de liquidités afin d'améliorer son train de vie, d'anticiper les besoins futurs liés à l'âge et de transmettre à ses ayants droits, à ses proches.

Le vendeur du bien possède alors une certaine liberté, le fait que l'usage du logement soit garanti à vie permet de pouvoir prolonger l'occupation du bien et être seul décisionnaire à ce niveau. Autre aspect important, la flexibilité de l'usufruit permettant au senior de conserver une part supplémentaire de la valeur totale dans son patrimoine jusqu'à épuisement. Cette part est également transmissible, il n'y a donc pas de lèse du legs.

De plus, ce mécanisme se différencie d'une vente classique puisque c'est un pari sur la vie de l'assuré et non sur sa mort. C'est cet aspect qui permet que la transaction soit sereine et équilibrée.

- L'investisseur tire également quelques avantages à cette situation :

L'investissement en nue-propiété permet de se confronter à moins de risques, moins de coûts et moins de frottements fiscaux côté débirentier. Les risques liés à une vente de viager classique ayant déjà été abordés, il est clair que proposer ce type de transaction est un véritable pari pour le débirentier.

La valorisation progressive liée à la reconstitution de la pleine propriété est également un aspect intéressant pour un Assureur bien que le niveau de rentabilité soit connu et stable quel que soit l'échéance. L'Assureur est en effet garanti de percevoir des revenus dès la fin de l'usufruit temporaire.

Outre le fait que plusieurs aléas passent à la trappe (aléa locatif, aléa sur la durée de vie), et que la fiscalité soit plus favorable, le réel changement apporté par ce mécanisme est éthique. La vente de viager est éthiquement un problème du point de vue assureur. Le fait de « parier sur la mort » du propriétaire est souvent mal perçu et contribue certainement au fait que la vente de viager soit peu pratiquée en France. Ce mécanisme, qui n'est pas un pari sur la mort, apporte donc une réponse à une problématique sociétale.

II. Notions de base de modèle de durée

A. Notations actuarielles et notion de censure

T est la variable aléatoire continue qui modélise la durée de survie. Elle peut être positive ou nulle. Soit S , fonction de survie, définie comme la probabilité qu'un individu survit jusqu'à un instant t positif. S s'exprime comme : $S(t) = P(T \geq t)$.

La fonction de répartition de T se déduit facilement de la fonction de survie :

$$F(t) = 1 - S(t) = P(T \leq t)$$

F s'interprète comme la probabilité de mourir avant l'instant t . S a également comme point en commun avec F le fait qu'il définit la loi de distribution. Par les propriétés de la fonction de répartition, nous retrouvons facilement la densité de probabilité de T :

$$f(t) = \lim_{h \rightarrow 0} \frac{P(t < X < t + h)}{h} = F'(t)$$

En se basant sur les résultats précédents, une relation entre f et S peut également être relevée :

$$f(t) = -S'(t)$$

Dans le cadre de ce mémoire, nous étudions la durée de survie d'individus à différents âges et sur différentes générations (matérialisées par des années). Plaçons-nous donc dans le cas discret où x représente l'âge d'un individu. Soit L_x le nombre d'individus encore vivant à l'âge x . En se plaçant sur l'année t , $L_x(t)$ est le nombre d'individus de x années révolues durant l'année t . En général, l'observation se base sur un échantillon initial de $l_0 = 100\,000$ personnes qui décroît au fil des âges suite aux décès.

Parallèlement, le nombre de décès à l'âge x , D_x correspond au nombre de décès à l'âge x . Pour l'année t , le nombre de décès à l'âge x est noté $D_x(t)$. La réalisation de cette variable s'exprime également comme $d_x(t) = l_x(t) - l_{x+1}(t)$; la différence entre le nombre de survivants à l'âge x et à l'âge $x + 1$.

Soit ${}_t p_x$, taux de survie ou la probabilité de survivre un temps supplémentaire τ sachant que l'individu a déjà survécu jusqu'à x . Ce taux de survie peut être calculé à partir d'une des formules suivantes :

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

La variable sur lequel se porte l'intérêt de notre étude est le taux de mortalité ${}_t q_x$. Cette valeur se retrouve à partir du taux de survie ou avec la formule suivante :

$${}_t q_x = 1 - {}_t p_x = P(T < x + \tau | T \geq x)$$

Ces deux taux ont un lien direct avec la fonction de survie : $q_x = 1 - \frac{S(x+1)}{S(x)}$.

L'étude se fait sur un temps discret, âge par âge, les taux de mortalités et de survie s'écrivent alors : ${}_1q_x = q_x$ et ${}_1p_x = p_x$. Le taux de mortalité à l'âge x pour l'année t s'écrit alors : $q_x(t)$.

Le taux instantané de mortalité μ_x est la probabilité de décéder entre x et $x + \Delta x$; soit un laps de temps assez court. Il sera considéré ici comme un taux annualisé s'exprimant de la manière suivante :

$$\mu_x = -\frac{1}{l_x} \frac{dl_x}{dx}$$

En raisonnant en années, la relation suivante existe entre taux instantané de mortalité et taux de décès.

$$\mu_x(t) = -\ln(1 - q_x(t))$$

Dans le calcul des rentes, les instruments qui rentrent en jeu sont les commutations. Les commutations se basent sur la table de mortalité classique avec des l_x pour chaque âge et année ; et les quantités suivantes sont utilisées pour simplifier les calculs :

Soit i un taux d'intérêt technique fixé et w l'âge limité de la table :

$$D_x = l_x(1 + i)^{-x}$$

$$N_x = \sum_{k=0}^{w-x} D_{x+k}$$

Un paramètre important à prendre en compte lors de ce type d'étude est la complexité de recueillir des données abondamment étalés dans le temps. Sur ce type d'études, la fenêtre d'observation est placée entre deux dates $t_{début}$ et t_{fin} . Cet intervalle, qui n'est que très rarement l'intervalle total d'observation du portefeuille, permet d'avoir une information sur un individu à l'unique condition qu'il soit observable à une date t située entre $t_{début}$ et t_{fin} . Si cet individu a une durée d'observations qui ne correspond pas à la durée totale d'observations sur laquelle l'étude se place ; ces données peuvent être considérées comme incomplètes. Deux phénomènes rendent ces observations incomplètes : la censure et la troncature (Huber-Carol, 1994).

Censure signifie que l'observation est incomplète à cause d'une cause aléatoire. Il existe plusieurs types de censures divisées en deux axes : la censure à gauche et à droite. En se basant sur une variable d'intérêt X ; cette variable est censurée par une variable de censure C si C est observé au lieu de X . Si $X > C$, il y a censure à droite et dans le cas inverse, on parle de censure à gauche.

La durée est censurée à droite si l'individu n'a pas subi l'évènement à sa dernière observation. La censure à droite est souvent présente dans les études de longévité en assurance vie. Lorsqu'un individu quitte le portefeuille assureur avant de décéder (par terme de contrat, fin d'observation ou rachat du contrat) ; son information est censurée puisque la date du décès n'a pas pu être observée.

La troncature est un phénomène différent de la censure pas incompatible. Il y a troncature gauche (resp. droite) lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un

seuil $c > 0$ (resp. supérieure à un seuil $C > 0$). Contrairement à la censure, la troncature est liée à une perte totale d'information sur les observations en dehors de la plage. La censure possède une information qui n'est pas complète.

$q_x(t)$ est la variable d'intérêt lorsque l'étude porte sur la durée de vie humaine. C'est ce taux qui sera estimé dans un premier temps à partir des données brutes. $\hat{q}_x(t)$ sera la notation du taux de mortalité brut. Une fois les taux bruts obtenus, ils seront lissés à partir de différentes méthodes de modélisation ; les taux de mortalité estimés seront alors notés $\tilde{q}_x(t)$. Il faudra donc estimer ces valeurs en prenant bien en considération les phénomènes de censure-troncature dans la suite de l'étude.

B. Les méthodes de fermeture de table

L'estimation de la mortalité avec un modèle de mortalité n'est souvent possible que jusqu'à un certain âge, notamment car aux grands âges il y a peu de données. C'est pour cette raison que plusieurs méthodologies de « fermeture de table » peuvent être appliquées. Ces méthodes s'appliquent à partir d'un certain âge x^F , choisi arbitrairement.

Même si des auteurs préconisent d'utiliser une méthode à partir d'un certain âge, les extrapolations restent le plus souvent modulables. x^F peut notamment être choisi à partir d'une analyse des résultats obtenus.

Ces méthodes d'extrapolation permettent donc de calculer des taux de décès à partir de x^F et jusqu'à un âge final qui peut aller au-delà des observations. De manière générale, les fermetures de tables se basent sur les résultats obtenus pour les âges antérieurs à x^F .

1. L'extrapolation exponentielle

L'extrapolation exponentielle se base sur deux paramètres α et β . Cette méthode est simple à appliquer car α est le seul paramètre du modèle qui prend en compte les résultats précédents. Si x^F est l'âge à partir duquel la fermeture de table débute et x^Z l'âge où se termine la fermeture, les paramètres se calculent de la façon suivante :

$$\alpha = \frac{-\ln(q_{x^F-1})}{x^Z - (x^F - 1)} \quad \text{et} \quad \beta = \exp(-\alpha x^Z)$$

Ces paramètres sont appliqués aux taux de décès à partir de x^F de la façon suivante :

$$q_x = \beta * \exp(\alpha x)$$

Cette méthode qui est une méthode généralement peu utilisée offre de bons résultats dans la suite de l'étude. Elle permet de prendre en compte, pour chaque taux de décès, l'âge et les taux de décès obtenus avant la fermeture. De plus, le calcul permet d'obtenir un taux de décès à 100% pour l'âge final $q_{x^Z} = 1$.

2. Denuit & Goderniaux

La méthode de Denuit & Goderniaux (2005) consiste quant à elle à ajuster par moindres carrés le modèle log-quadratique en se basant sur des observations relatives aux âges les plus élevés.

$$\ln(q_x) = a + bx + cx^2 + \varepsilon_x$$

avec ε_x iid suivant une loi normale centrée de variance σ^2 .

Soit q_x' , dérivée première de q_x (fonction de l'âge x), deux contraintes relatives à l'âge x^Z apparaissent :

$$q_{x^Z} = 1 \text{ et } q_{x^Z}' = 0$$

En prenant l'âge final $x^Z = 130$ ans, ces contraintes donnent le résultat suivant :

$$a + bx + cx^2 = c(130^2 - 260x + x^2)$$

De plus, ces contraintes ont pour incidence une allure de courbe concave aux grands âges mais aussi une tangente horizontale au point x^Z . L'allure concave existe en particulier par le biais de la contrainte sur dérivée, empêchant également q_x de décroître aux âges avancés.

3. Coale-Kisker

La méthode de Coale & Kisker (1990) propose de recalculer les taux de mortalité bruts à partir de 65 ans minimum et de les extrapoler aux grands âges via la formule de taux instantanés suivante :

$$\mu_x = \mu_{65} * \exp(k_x * (x - 65))$$

k_x est le taux de croissance de la mortalité à l'âge x .

Coale-Kisker est donc utilisable à partir de 65 ans. Il est tout de même possible de choisir un âge à partir duquel x^F débute. Pour extrapoler cette formule, il faut supposer que le taux de croissance de la mortalité suit la fonction suivante :

$$k_x = k_{x^F} + s(x - x^F) \text{ pour } x \geq x^F$$

Le coefficient s s'obtenant quant à lui à partir de la condition $q_{x^Z} (= q_{130}) = 1$:

$$\mu_{x^Z} = \mu_{x^F-1} \exp\left(\sum_{x=x^F}^{x^Z} k_x\right) = q_{x^F-1} \exp\left(\sum_{x=x^F}^{x^Z} k_{x^F} + s(x - x^F)\right)$$

$$s = - \frac{\ln\left(\frac{\mu_{x^F-1}}{\mu_{x^Z}}\right) + (x^Z - x^F + 1)k_{x^F}}{\sum_{x=x^F}^{x^Z} (x - x^F)}$$

Les taux de mortalité se calculent donc à partir de ces paramètres et par la suite suivante :

$$\mu_x = \mu_{x-1} * \exp(k_{x^F} + s(x - x^F))$$

Pour tout $x \in [x^F, x^Z]$.

4. Kannistö

Le modèle de Kannistö (Thatcher et al., 1998) calcule les taux de mortalité extrapolés à partir de deux coefficients a et b et de la formule suivante :

$$\mu_x = \frac{a * \exp(bx)}{1 + a * (\exp(bx) - 1)}$$

Ce modèle est tiré du modèle logistique classique. C'est en quelque sorte une simplification de ce dernier puisqu'il ne retient que deux paramètres contre 4 pour le modèle logistique classique. Les paramètres sont estimés par la méthode des moindres carrés pondérés et un algorithme de Newton-Raphson est utilisé pour la résolution du système. Le choix des paramètres initiaux peut s'avérer délicat.

Ces quatre méthodes font partis de la grande famille des méthodes de fermeture de table. Elles ont été introduites dans cette partie car ce sont celles qui seront utilisées dans la modélisation.

C. Longévité et population française

Plusieurs indicateurs permettent d'analyser la survie et donc d'étudier la mortalité. Parmi eux, l'espérance de vie (à la naissance ou à 60 ans), les taux de mortalité, l'âge modal des décès mais aussi d'autres indicateurs liés aux supercentenaires ou la mortalité infantile.

L'espérance de vie est le temps moyen pour un organisme de rester en vie sachant son âge actuel. L'espérance de vie humaine dépend de nombreux facteurs : le sexe, l'année de naissance, le pays de résidence, l'environnement, le travail, le mode de vie... L'espérance de vie peut être catégorisée sous deux types :

- L'espérance de vie par période qui est l'espérance de vie pour une année donnée sans amélioration future de la mortalité.
- L'espérance de vie par cohorte qui concerne une cohorte donnée, avec les améliorations futures de la mortalité.

L'espérance de vie qui est généralement étudiée par les instituts de statistiques est l'espérance de vie à la naissance pour une période donnée.

En France, l'espérance de vie à la naissance et à 60 ans fait partie des meilleurs au monde. Comme l'atteste le tableau ci-dessous récapitulatif des espérances de vie à la naissance par pays et sexe en 2015. Ce tableau est basé sur des données de l'OMS.

Pays	Rang pour les deux sexes	Rang pour les hommes	Espérance de vie hommes	Rang pour les femmes	Espérance de vie femmes
 Japon	1	1	82,4	1	86,8
 Suisse	2	6	80,3	6	85,3
 Singapour	3	10	80	2	86,1
...
 Suède	9	4	80,7	12	84
 France	9	16	79,4	5	85,4
...
 Sierra Leone	183	183	49,3	183	50,8

Tableau 4 : Espérance de vie à la naissance par pays et sexe avec les données de l'OMS de 2015

De manière générale, la France se situe bien en termes d'espérance de vie en affichant notamment la cinquième meilleure pour une femme à la naissance. En observant de plus près le classement général, il est clair que l'espérance de vie est dépendante du niveau de richesse du pays. Les pays les plus pauvres se trouvent en queue de classement.

L'autre distinction qui se fait assez clairement est la différence d'espérance de vie entre les hommes et les femmes. Cette différence montre que le sexe est un facteur important dans le calcul d'espérance de vie et plus généralement dans les analyses de survie. La différence entre les hommes et les femmes est d'autant plus significative du fait qu'aucun pays ne montre une espérance de vie à la naissance plus haute pour les hommes que pour les femmes. L'écart minimal s'élève à 0,1 année pour la population malienne ; au contraire, l'écart maximal s'élève à 11,6 ans en Russie sur les données de l'OMS de 2015.

D'un pays à l'autre, les causes principales de décès sont totalement différentes. En effet, les causes diffèrent d'autant plus entre deux pays d'économies différentes. Les maladies les plus présentes dans un pays à faible revenus ne sont pas les mêmes que celles d'un pays à haut revenus.

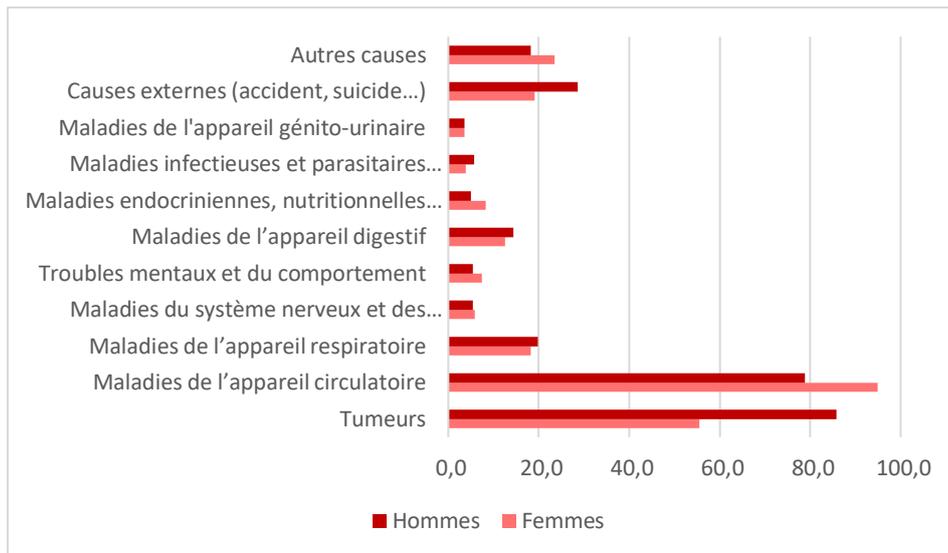


Figure II-1 : Causes de décès recensées en 2015 en France par sexe en milliers d'après les données de l'INSEE

En France, les causes de décès semblent être réparties globalement de la même façon pour les hommes et les femmes. Les causes présentes en grande quantité sont les tumeurs et les maladies de l'appareil circulatoire (cœur, vaisseaux sanguins...).

Afin de se rendre compte de ces résultats par âge, la courbe ci-dessous a été construite à partir de la table T68 – Table de mortalité des années 2013 – 2015 avec des données provisoires arrêtées à fin mars 2017 disponibles sur le site de l'INSEE. Cette courbe relate les taux de mortalités par âge basée sur une population de 10 000 hommes.

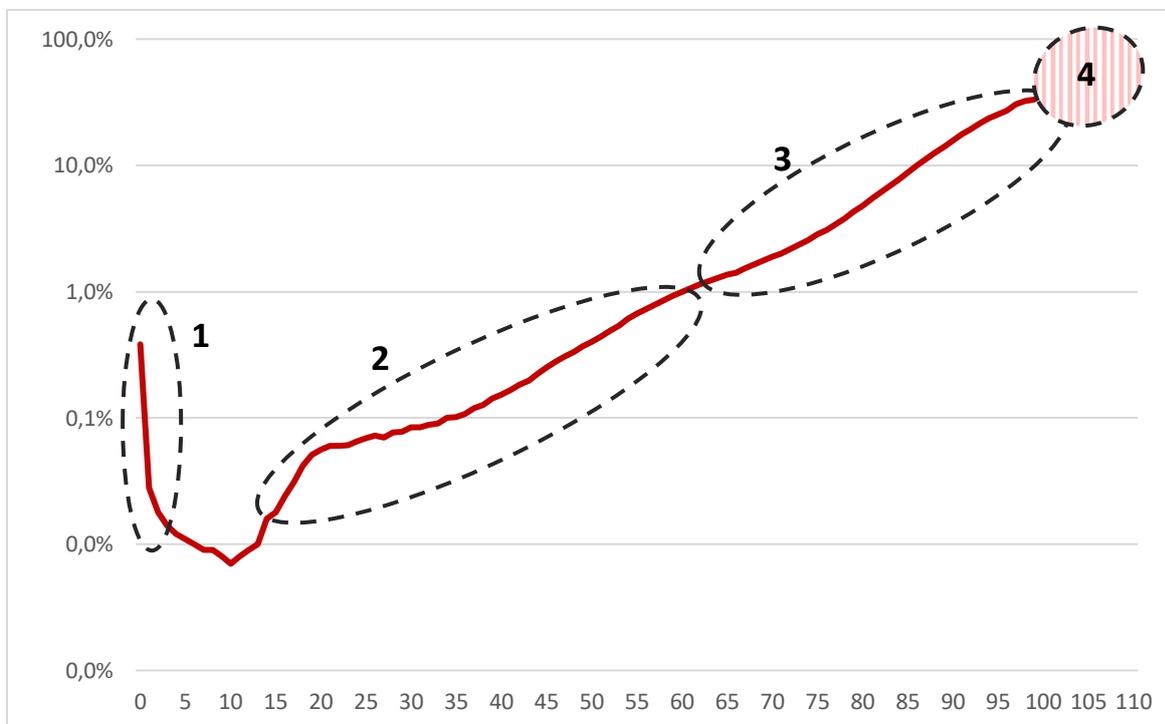


Figure II-2 : Taux de mortalité des hommes pour les années 2013 à 2015 basée sur la T68

Les taux ci-dessus ont été tracés à une échelle logarithmique afin de mieux se rendre compte des phénomènes à chaque âge. En observant de plus près, 4 phénomènes se distinguent sur cette courbe :

1. Le premier phénomène relaté concerne la mortalité des moins d'un an. Ce phénomène s'appelle la mortalité infantile. C'est un phénomène qui touche beaucoup moins les pays développés que les pays en développement. La mortalité infantile s'explique par les conditions sanitaires à l'accouchement mais aussi par le manque de soins aux nouveaux nés.

La mortalité infantile en France est décroissante au fil des années grâce aux progrès liés aux soins des nouveaux nés ; comme l'atteste la courbe ci-dessous retraçant les taux de mortalité infantile français d'après les tables de l'INSEE.

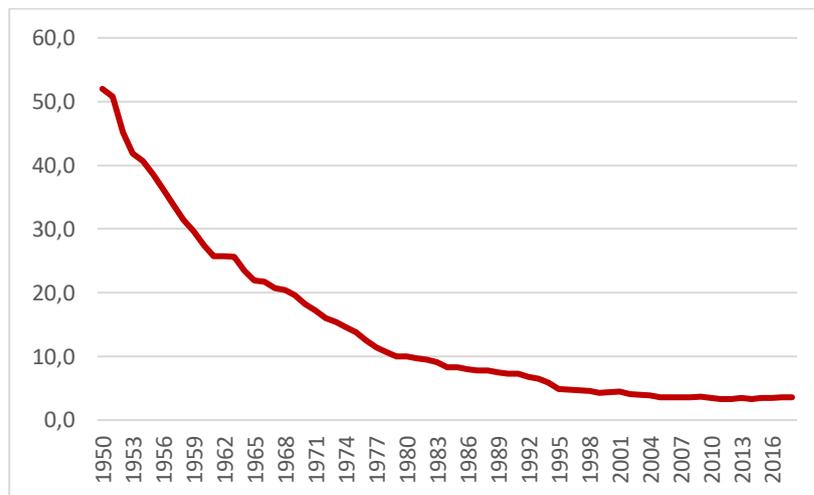


Figure II-3 : Taux de mortalité infantile en France d'après les données de l'INSEE

Comme pour la plupart des pays développés, la mortalité infantile en France ne dépasse pas les 1%. La mortalité infantile comprend la mortalité néonatale qui concerne uniquement la mortalité des nouveaux nés de moins de 28 jours.

2. Le second phénomène est appelé la mortalité « accidentelle ». Ce phénomène relatif aux âges allant de 15 à 60 ans est appelé de cette façon pour l'aspect inhabituel et anormal de la mortalité à ces âges. C'est une tranche d'âge pour laquelle les causes externes (accidents, suicides...) sont assez fréquentes.
3. Entre 60 et 100 ans, c'est la mortalité des seniors qui est abordée. Ce sont les âges sur lesquels la mortalité est la plus fréquente. L'âge modal des décès est notamment, pour la plupart des pays développés, inclus dans cette tranche. L'âge modal est le pic maximal du nombre de décès ; c'est l'âge auquel le plus de personnes décèdent.
4. Le dernier phénomène qui concerne les plus de 100 ans est la mortalité des plus âgés. Elle comporte notamment la mortalité des supercentenaires. Les supercentenaires sont les personnes vivant jusqu'à 110 ans ou plus. Ces personnes sont en grande majorité des femmes et restent assez rares.

Ce sont ces deux derniers points qui seront abordés dans le cadre de cette étude.

Néanmoins, les progrès de la médecine éliminent ou diminuent petit à petit certaines maladies qui s'imposaient comme cause de décès principales. A cela s'ajoutent les améliorations de la qualité de vie et la stabilité géopolitique. Au fil des années, il y a donc dû avoir une certaine évolution des causes de décès et donc une tendance positive des espérances de vie.

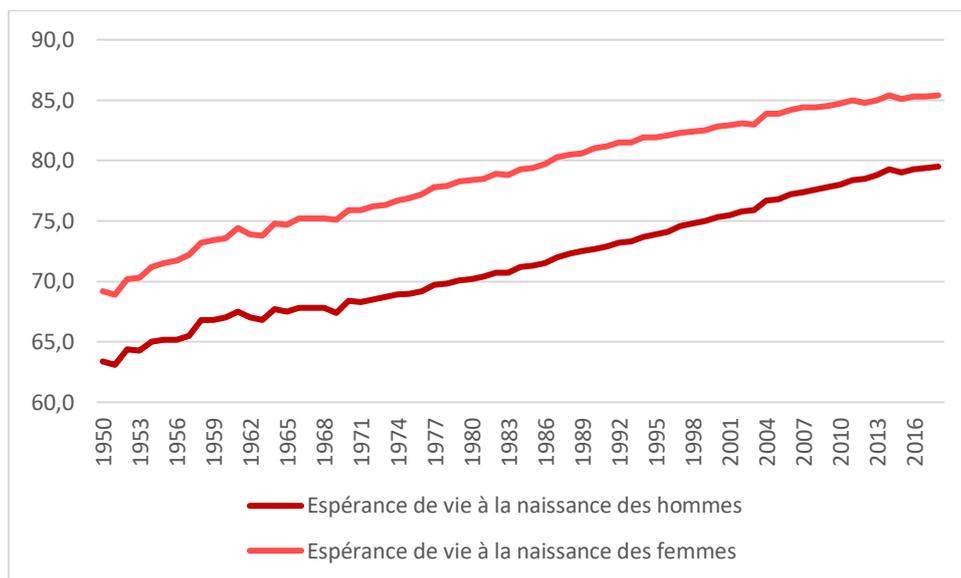


Figure II-4 : Evolution des espérances de vie à la naissance depuis 1950 à partir des tables de l'INSEE

Ces courbes mettent en avant pour chaque sexe, l'espérance de vie croît au fil des années. La tendance (globalement) croissante de l'espérance de vie peut être liée aux améliorations faites chaque année, autant d'un point de vue qualité de vie que d'un point de vue santé. Le décalage entre l'espérance de vie des femmes et celui des hommes semble rester plus ou moins le même au fil des années (entre 5 et 6 ans) et ne semble pas se résorber.

Les améliorations d'espérance de vie touchent les personnes de plus de 65 ans. La part des personnes de plus de 65 ans est en augmentation au fil des années en France, s'élevant à 30% en 2014. Cette augmentation, au même titre que l'évolution de l'espérance de vie, devrait perdurer dans les années à venir.

La France s'affiche donc comme l'un des tous meilleurs pays en termes d'espérance de vie. Plusieurs facteurs influent sur ce niveau d'espérance de vie comme la richesse, le sexe ou les habitudes. Le temps fait pour le moment évoluer positivement l'espérance de vie et la longévité de la population. L'accroissement de longévité passé et à venir peut donc constituer un risque pour les assureurs pour qui la durée de vie de l'assuré est un pari.

D. Risque de longévité en assurance et tables d'expérience

Le risque de longévité s'inscrit dans la hiérarchie des risques de la « Formule standard » du calcul de SCR de l'assureur proposé par la directive Solvabilité II. Le SCR, ou *Solvency Capital Requirement*, représente le capital cible nécessaire pour absorber le choc provoqué par un risque

majeur ; et la directive en vigueur propose une formule qui repose sur un éventail de risques dont le risque de longévité.

Le risque de longévité est le risque de perte, ou de changement défavorable de la valeur des engagements d'assurance, résultant de fluctuations affectant le niveau, l'évolution tendancielle ou la volatilité des taux de mortalité, lorsqu'une baisse de ces taux entraîne une augmentation de la valeur des engagements d'assurance.

La tarification et le provisionnement des engagements d'assurance vie sont encadrés par la législation de manière assez stricte. Des contraintes existent pour les taux techniques utilisés mais aussi pour les tables de mortalité. L'article A335-1 du Code des Assurances (respectivement A931-10-10 du Code de la Sécurité Sociale et A212-10 du Code de la Mutualité) fixe le cadre légal. Ce cadre laisse la liberté à l'assureur d'utiliser une table d'expérience dans le but d'accroître la précision des taux de mortalité utilisés en se basant sur les données de son portefeuille ; et donc de limiter le risque de longévité. Une table d'expérience peut être construite à partir de diverses méthodes. La table doit ensuite être certifiée par les règles de l'Institut des actuaires pour pouvoir être utilisée. La certification permet de mesurer le risque du portefeuille de manière réaliste et raisonnablement prudente (Planchet, 2007).

Pour la mise en place d'une rente viagère, les assureurs doivent obligatoirement utiliser :

- le jeu de table TGH05-TGF05 (table par génération homme/femme version 2005). Ce jeu de table établi par l'INSEE sont des tables prospectives par génération. Ces tables comportent toutes les générations de 1900 à 2005 ; pour des âges allant de 0 à 120 ans.
Si l'assureur choisi ce jeu de table, il a la possibilité de choisir une table unique pour les deux sexes. Dans la pratique, c'est la TGF05 qui est choisie pour les deux sexes car c'est celle qui surestime le moins la mortalité parmi les deux.
- une table d'expérience certifiée par un actuaire indépendant agréé à cet effet.

L'assureur est indéniablement exposé à un risque de longévité qui risque de croître au vu de la tendance que semble prendre la longévité. Une fois les bases actuarielles posées, les méthodes d'extrapolation introduites et les tables d'expériences définies ; il convient maintenant de construire une table de mortalité prospective d'expérience à partir des données du portefeuille Generali vie.

III. Données utilisées

Un contrat d'assurance vie engage l'assureur à verser à l'assuré en contrepartie de la perception de primes, une rente ou un capital à des bénéficiaires. Plusieurs types de contrats existent : contrat en cas de vie, en cas de décès, ou contrat vie et décès. Durant la durée de vie du contrat, l'assuré place son encours sur des supports de différents types et accumule des intérêts. C'est sur base de données composée de ce type de contrats qu'a été effectuée cette étude.

A. Présentation de la base

L'étude est basée sur le portefeuille épargne des CGI et grands comptes de Generali VIE. Les contrats d'épargne Generali sont commercialisés à partir de différents canaux (agents/courtiers, CGPI, internet, banques, réseau salariés). Le portefeuille retenu porte uniquement sur les contrats issus de CGPI, d'internet ou de banques. Ces différents canaux sont eux même divisés en segments : chaque contrat est donc issu d'une segmentation liée à l'un de ces 3 canaux.

La base de données finale est le résultat de la juxtaposition de 2 bases :

- Une première composée de l'ensemble des contrats de Generali associés à l'Assuré, ses caractéristiques (encours, informations personnelles etc.). Les éventuels rachats, termes ou décès sont également précisés dans l'extraction. Cette base est la plus lourde des deux par la quantité d'informations qu'elle comporte.
- La seconde base concerne le portefeuille des rentiers de Generali. Tous les individus dont le contrat n'est plus en cours mais qui perçoivent tout de même une rente sont présents dans la base.

Puisque l'intérêt de notre étude concerne plus l'assuré que le contrat en lui-même, le numéro d'assuré a été choisi comme variable principale. C'est un code propre à chaque personne physique qui souscrit un contrat auprès de Generali. Le numéro d'assuré permet de se détacher du ou des contrats de l'assuré mais aussi de pouvoir faire le lien entre les deux bases. C'est donc à partir de cette variable commune que la jointure a été faite.

Plusieurs variables composent cette base :

- L'identifiant de l'Assuré ;
- Le sexe de l'Assuré ;
- La date de naissance de l'Assuré ;
- La date de décès de l'Assuré ;
- Le numéro de contrat ;
- La situation familiale de l'Assuré ;
- La situation actuelle du contrat, décrite par un libellé ;
- La segmentation par laquelle est issu le contrat ;

- La date d'effet ainsi que la date de terme du contrat. Marquant respectivement le début et la fin du contrat sous réserve de sinistre ou de rachat ;
- Le mode de gestion du contrat ;
- La catégorie socio-professionnelle de l'Assuré ;
- L'indicateur de garantie plancher ;
- L'indicateur de nantissement ;
- La région de résidence de l'Assuré (étranger si ce n'est pas sur le territoire français) ;
- La somme des primes versées sur le contrat ;
- L'encours total à la dernière observation (date de sortie) ;
- L'éventuelle date de rachat total, évènement clôturant le contrat.

L'assuré peut être titulaire de plusieurs contrats au sein même du portefeuille. Pour pallier ce problème, un seul contrat a été conservé par numéro d'assuré ; le contrat en cours avec la plus haute ancienneté est retenue. Si tous les contrats ont été rachetés, c'est le plus ancien qui est retenu. En parallèle, la somme des encours est calculée et affectée à l'observation retenue.

B. Prétraitement

La plage d'observations retenue dans le cadre de cette étude est d'une durée de 10 ans, se situant donc entre $t_0=01/01/2010$ et $t_1=31/12/2019$. Dans le but de se concentrer sur les seniors, seuls les individus ayant dépassés la soixantaine ont été conservés ; cet âge est déterminé de façon arbitraire.

Les observations pour lesquelles une de ces propositions est vraie sont donc retirées de la base :

- $t_{effet} > t_1$: les contrats qui ont été souscrits après 2019,
- $\min(t_{terme}, t_{décès}) < t_0$: les contrats qui ont vu leur terme arriver avant 2010 et ceux qui ont eu un sinistre avant 2010,
- $age_{t_1} < 60$: les assurés n'ayant pas atteint les 60 ans fin 2019.

L'une des particularités de cette base est que les Assurés voyant leur contrat arriver à terme peuvent décider de récupérer leur encours sous forme de rente. Ils deviennent donc rentiers et restent dans le portefeuille. C'est une particularité qui a une influence positive sur l'étude car elle permet d'avoir une durée supplémentaire d'observation. Il est donc important de préciser que la date de terme devient donc la date de terme de la rente s'il y a transformation en rente :

$$t_{terme} = \begin{cases} t_{terme} (rente) & \text{si l'Assuré devient rentier} \\ t_{terme} (contrat) & \text{sinon} \end{cases}$$

Puisque certaines rentes sont de natures viagères et n'ont donc pas de date de terme, le 31/12/2999 leur est affectée comme date de terme.

Un décès peut intervenir à la suite du terme du contrat de base, au cours de la rente. Le libellé de situation de contrat a été traité pour retenir 3 états différents : en cours, regroupant les *contrats et rentes en cours* ; *racheté* ou *décès au cours de la vie du contrat ou de la rente*.

L'ensemble des variables de dates ont été mises sous le même format « AAAA/MM/JJ » afin de simplifier les calculs. Des variables date d'entrée et date de sortie ont été créées de la façon suivante :

- $t_{sortie} = \min(t_{terme}, t_{rachat\ total}, t_{d\acute{e}c\grave{e}s}, t_1)$
- $t_{entr\acute{e}e} = \max(t_{effet}, t_{60}, t_0)$

Avec t_{60} , le 60^{ème} anniversaire de l'assuré.

Dans la base brute extraite, il se peut que quelques observations aient donné lieu à des résultats aberrants ou même manquants à cause d'erreurs d'extraction ou de saisie.

C. Les données manquantes

Lors de la saisie des données ou de l'extraction, certaines valeurs ont été oubliées dans la base.

Dans cette étude, certaines variables sont indispensables et il n'est pas possible de conserver ces données manquantes ou même de les remplacer. C'est notamment le cas des variables qui permettent de différencier les observations : *Code contrat*, *Identifiant Assuré*. C'est également le cas des variables de dates : *effet*, *terme*, *naissance*. La totalité des contrats possédant des données manquantes au niveau de ces variables ont été supprimés par souci de prudence.

D'autres variables possèdent des valeurs manquantes. C'est le cas du sexe de l'assuré. Le sexe est une variable très influente dans le calcul d'espérance de vie. Il a donc fallu donner de l'importance au bon traitement de cette variable. Bien heureusement, la quantité de valeurs manquantes reste négligeable comparée au nombre d'observations : 3,50 % données de sexe manquaient à notre base. Nous avons donc décidé d'y affecter une méthode d'imputation par moyenne. Sur les données non-manquantes, 45,47 % étaient des femmes et 54,53 % des hommes. L'hypothèse prise est que ces données manquantes ne devraient pas grandement modifier la répartition par sexe. Cette répartition a donc été appliquée aux données manquantes par une implémentation par pourcentage.

1,23 % contrats ne fournissaient pas d'information sur la segmentation dont ils sont issus. Du fait de la faible quantité de valeurs manquantes, elles ont été implémentées avec la même méthode utilisée pour la variable sexe.

L'imputation par moyenne a également été utilisée pour la variable donnant une indication sur le lieu de résidence de l'assuré. Cette variable est à ce stade définie par des numéros de départements ou 00 si l'individu réside à l'étranger.

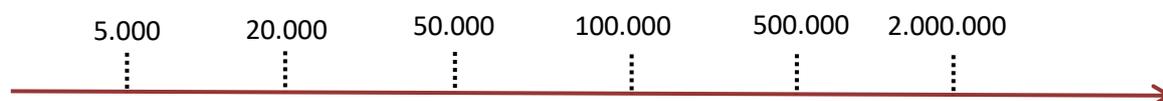
La catégorie socio-professionnelle (CSP) et la situation familiale ont été retirées de l'étude en raison d'un trop grand nombre de valeurs manquantes. Avec respectivement 95,40 % et 89,12 % de données manquantes, ces variables ont été jugées comme non significatives et une méthode d'imputation biaiserait les résultats.

D. Valeurs aberrantes et traitement variable par variable

Certaines dates possèdent des anomalies. La base a donc été restreinte aux observations pour lesquelles la *Date d'effet*, *de décès*, *de rachat* et la *Date de naissance* ne dépassent pas 2021 et ne précèdent pas 1890. Les Dates d'effet aberrantes (inférieures à 1930) ont également été retirées. La *Date de terme*, *de rachat ou de décès* doit toujours arriver après la *Date d'effet*. Il se peut que la *Date de terme* soit fixée en 2999, ceci n'est pas vu comme une anomalie. La prudence concernant les dates est très importante pour cette étude et c'est pour cette raison que chaque anomalie est retirée.

La variable primes, qui sera considérée comme un indicateur de richesse, est une variable continue. Des valeurs y sont parfois négatives, ce ne sont pas des anomalies mais des annulations de versement. Pour chaque contrat, la somme des primes versées a été réalisée. Il est important de bien conserver les valeurs négatives pour ne pas compter les versements annulés.

Par soucis de simplification, les primes ont été segmentées en 6 classes :



L'autre indicateur de richesse est la provision mathématique. La provisions mathématique (PM) juste avant t_{sortie} a été retenues. Cela permet, pour les assurés décédés, d'avoir la PM du contrat au moment du décès. Les seules valeurs aberrantes liées à cette variable sont quelques PM négatives (moins de 1,00% de la totalité). Ces PM sont transformées en valeur manquantes et ne seront pas utilisées dans la suite de l'étude.

Les numéros de départements des assurés est une variable discrète. L'ensemble des numéros de départements sujets à des DOM-TOM ont été transformés en 97. Les départements varient finalement de 0 à 98. 0, 96 et 98 n'existent pas en tant que département. Après comparaison avec une variable indicateur non-résident, ces observations ont été regroupées sous la valeur 0 qui sera considérée comme le numéro de département des individus étrangers.

E. Statistiques descriptives

Le portefeuille se compose après modifications de **274 734** observations. Puisque la base est composée dans son intégralité de personnes physiques, ces 274 734 observations sont également des individus.

Parmi ces observations, 19 378 décès ont été recensés sur cette période et tranche d'âge. 1 890 contrats ont été rachetés dans leur intégralité, 21 374 contrats qui sont arrivés à terme pendant la période. Il est important de distinguer ces trois variables pour bien prendre en compte les phénomènes de censures.

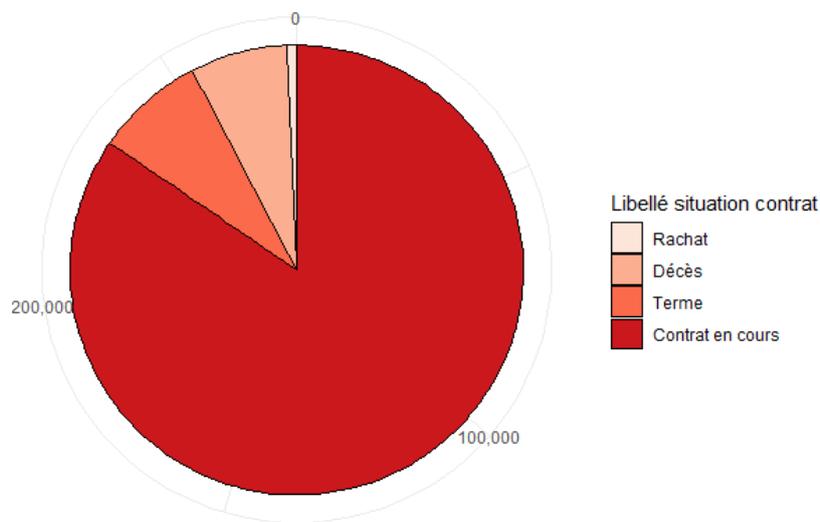


Figure III-1 : Répartition des libellés de situation de contrats

La plupart des contrats sont encore en cours, ce qui est naturel pour un portefeuille d'épargne.

Les assurés du portefeuille proviennent d'un peu partout en France mais aussi de pays étrangers, frontaliers ou non. Dans la base, la différence n'est pas faite entre les différents pays étrangers mais on a observé la répartition sur l'ensemble du territoire français. Il y a une forte majorité d'assurés vivant en Ile-de-France (presque 71 000 soit $\frac{1}{4}$ du portefeuille) mais on voit également trois autres régions qui sont assez présentes : l'Occitanie, l'Auvergne-Rhône-Alpes et la Nouvelle Aquitaine. Il y a un peu moins de 4 000 assurés étrangers.

A l'image du portefeuille intégral de Generali, la segmentation principale est Internet. Comme on peut le constater dans le tableau ci-dessous, presque la moitié de notre base de données est composée de contrats souscrits sur internet.

Banques Privées	Banques Régionales	Inspection	Internet	Plateforme
34 467	25 685	55 645	127 727	31 210
12,55%	9,35%	20,25%	46,49%	11,36%

Tableau 5 : Répartition des contrats par segmentations

La somme des primes émises sur le contrat est un indicateur de la richesse de l'assuré. Voici comment sont réparties les sommes de primes dans ce portefeuille :

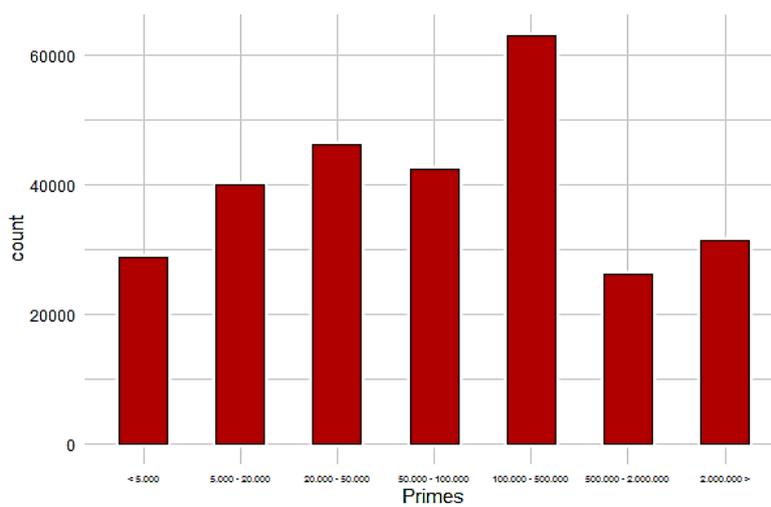


Figure III-2 : Répartition des différents échelons de primes

Les primes sont assez bien réparties entre ces 6 tranches avec une légère domination des contrats ayant entre 100 et 500 mille euros de primes encaissés.

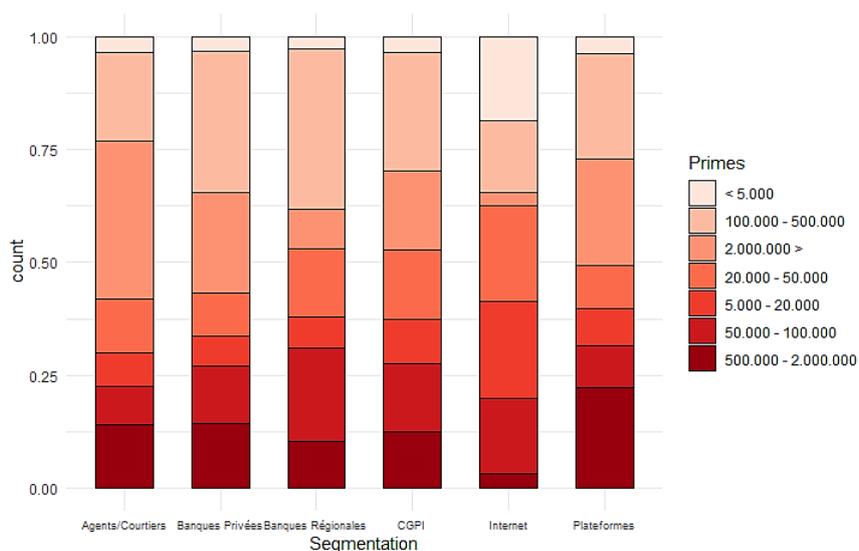


Figure III-3 : Proportion des primes versées par segmentation

La segmentation Internet est le plus souvent associée aux contrats ayant le moins de primes émises. Cela permet de constater que le portefeuille est composé en majorité de contrats internet mais que ce n'est pas forcément sur ce segment qu'il y a le plus d'encours.

Afin d'avoir une idée de comment est réparti l'âge des assurés sur ce portefeuille, voici un petit zoom sur la répartition de l'âge des assurés (vivants) de plus de 60 ans au 31/12/2019, date marquant la fin de notre étude.

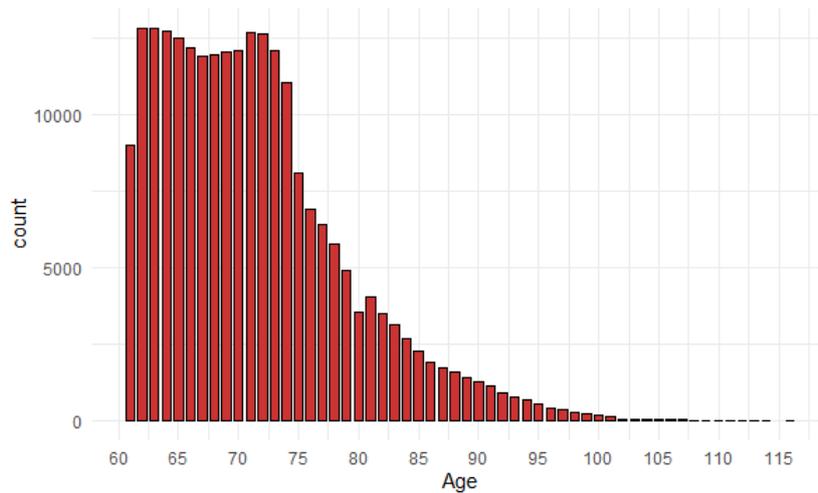


Figure III-4 : Répartition des âges d'assurés du portefeuille au 31/12/2019

Outre le fait que la répartition est logiquement décroissante, une forte affluence d'assurés âgés de moins de 75 ans est observable. C'est d'ailleurs à 76 ans que la tendance décroissante débute. Avec un âge moyen de 71 ans un âge médian à 70 ans et un âge modal à 61 ans ; l'hypothèse de forte concentration d'individus de moins de 75 ans semble réelle.

Chacun des segments a des caractéristiques totalement différents :

Segment	Banques Privées	Banques Régionales	Inspection	Internet	Plateforme
Age moyen	74,77	71,76	73,53	71,11	73,05
PM moyenne *	196 206 €	159 670 €	98 550 €	49 189 €	73 427 €

Tableau 6 : Caractéristiques propres à chaque segment (âge moyen, produit principalement vendu, PM moyenne)

* La PM retenue pour chaque observation est celle à sa date de sortie du portefeuille.

L'âge moyen diffère entre chaque segment, de manière assez distincte. Internet est le segment qui a le portefeuille le plus jeune mais aussi avec la PM moyenne la moins élevée. Les banques privées, qui ont l'âge moyen le plus élevé, sont également associées aux PM les plus élevées. Il est possible qu'une corrélation existe entre la PM et l'âge.

Pour éclairer ce phénomène, voici un graphique de la PM moyenne par âge :

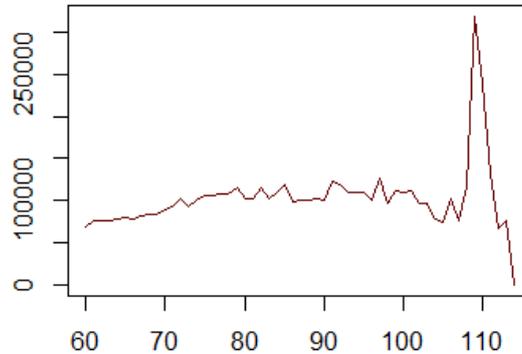


Figure III-5 : Courbe retraçant les PM moyennes par âge à la sortie

Il n'y a pas de réelle tendance croissante, il est donc difficile de dire que ces deux variables soient directement corrélées. Le pic à 109 ans met en avant la volatilité des données aux grands âges. Cette observation est certainement basée sur peu d'observations.

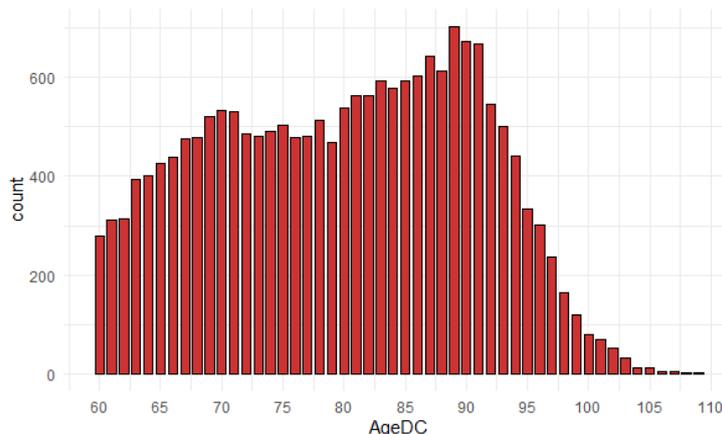


Figure III-6 : Histogramme des âges de décès des individus du portefeuille

Cet histogramme retraçant le nombre de décès à chaque âge sur la dizaine d'années observée montre plusieurs tendances. Une première augmentation de 60 à 70 ans est suivie par une période stable sur les 10 années qui suivent. Le nombre de décès augmente ensuite jusqu'à 90 ans pour décroître indéfiniment. Cette forte chute est due au fait que de moins en moins d'individus dépassent les années à partir de cet âge.

En observant de plus près, 2 pics peuvent être mis en avant : un léger aux alentours de 70 ans puis un pic assez remarquable aux alentours de 90 ans. L'âge modal de décès est de 89 ans. L'âge médian et l'âge moyen de décès du portefeuille sont respectivement 81 et 80 ans.

Comme l'a déjà été souligné précédemment, le sexe est un facteur très influant de l'espérance de vie. Le portefeuille est composé d'une majorité d'hommes : 54,58% d'hommes contre 45,42% de femmes. Parmi eux, 10 239 hommes et 8 874 femmes sont décédés. Les décès du portefeuille sont

donc constitués à 46,55% de femmes et à 53,45% d’hommes. Le pourcentage de femmes dans le portefeuille est plus faible que le pourcentage de femmes décédées, malgré un résultat assez proche. Le fait que les pourcentages soient assez proches laisse émettre comme hypothèse que les hommes décèdent en moyenne plus tôt que les femmes.

Ce résultat est vérifiable en comparant les répartitions par âges et par sexe des individus encore en vie dans notre portefeuille :

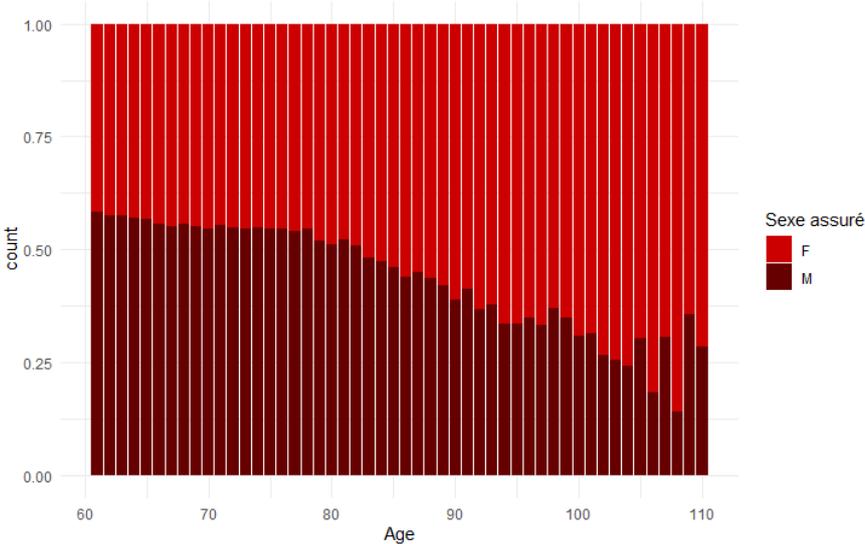


Figure III-7 : Proportion d’hommes et de femmes vivants par âge

La part de femmes dans le portefeuille est très clairement croissante au fil des âges. La part de femme est dominante dans le portefeuille à partir de 83 ans. Cette affirmation doit avoir un lien avec la répartition de l’âge de décès par sexe.

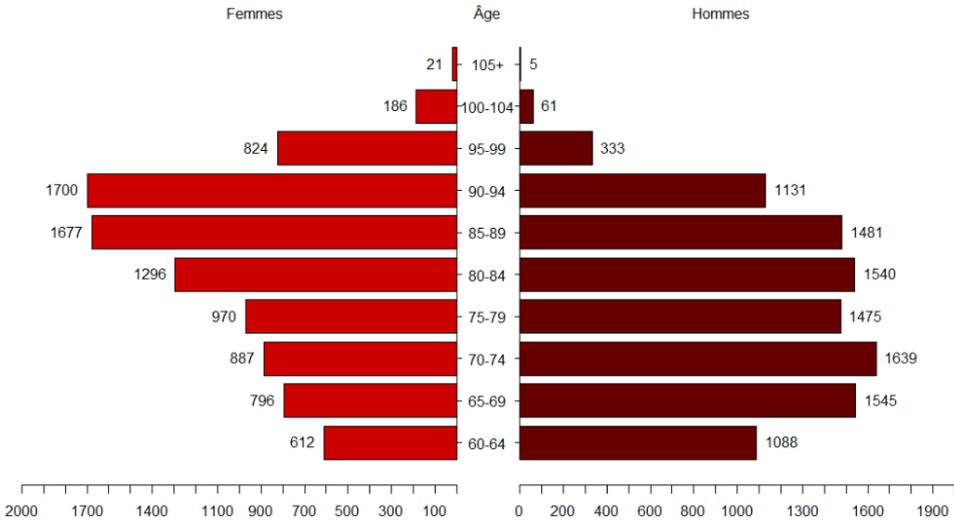


Figure III-8 : Histogramme des répartitions de décès par âge et sexe

Les hommes dominent les âges de mortalités les plus précoces. Par la suite, ce sont les femmes qui meurent plus puisqu’il reste davantage de femmes que d’hommes à ces âges.

En retraçant l'ensemble des résultats mis en avant concernant les âges et les répartitions par sexe, on obtient le tableau suivant :

Age	Moyen	Médian	Modal
Portefeuille au 31/12/2019	71	70	61
Femmes du portefeuille	72	71	73
Hommes du portefeuille	70	70	61
Décès sur la période	80	81	89
Femmes décédées	82	84	89
Hommes décédés	77	77	70

Tableau 7 : Récapitulatif d'âge moyen médian et modal par sexe, pour les décès ou au 31/12/2019

Les femmes décèdent en général à des âges plus tardifs que les hommes et c'est en corrélation avec la répartition des âges sur le portefeuille du 31/12/2019. Cette observation est à l'image de celles faites à l'échelle de la population française.

En observant maintenant les décès à l'échelle des années, les résultats sont tout autres. Il convient dans un premier temps d'observer l'exposition au risque par sexe. L'exposition au risque des graphiques ci-dessous sont calculés avec la définition retenue dans la partie suivante (c.f. IV, B).

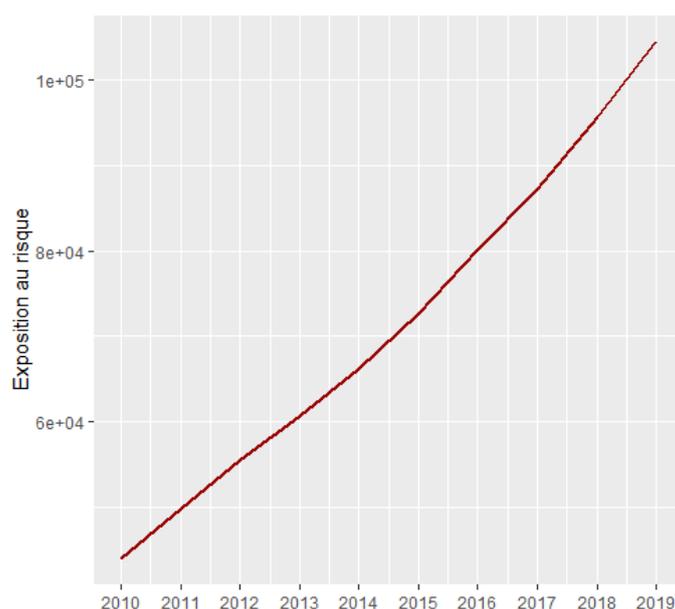


Figure III-9 : Exposition au risque de la population masculine par année

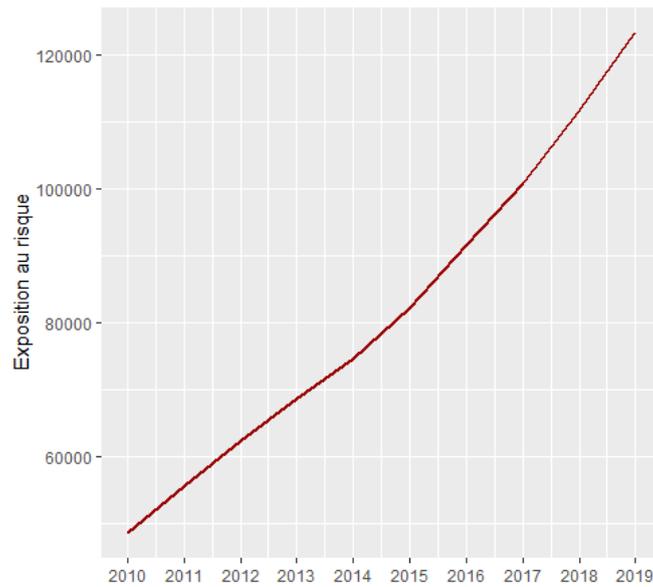


Figure III-10 : Exposition au risque par année sur la population féminine

L'exposition au risque évolue de façon assez linéaire au fil des années, quel que soit le sexe. Cette tendance linéairement croissante est liée à l'accumulation de nouveaux souscripteurs au sein de notre base mais aussi à l'augmentation temporelle du nombre d'individus dépassant les 60 ans.

Il serait intéressant de voir si le nombre de sinistres évolue de la même façon dans notre portefeuille.

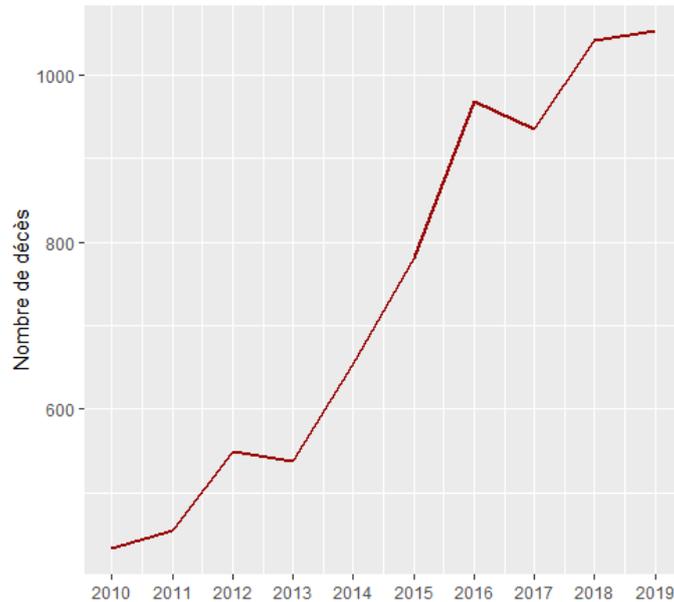


Figure III-11 : Nombre de décès par année pour la population masculine

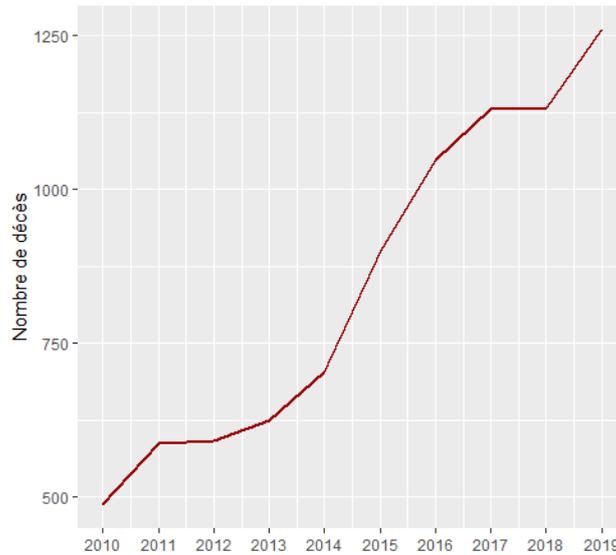


Figure III-12 : Nombre de décès par an pour la population féminine

Le nombre de décès n'évolue pas proportionnellement à l'exposition au risque. Malgré une tendance croissante assez claire, on observe sur certaines années de légers ralentissements pour chacun des sexes. Ces propos peuvent être illustrés par les chutes de 2013 et 2017 pour les femmes.

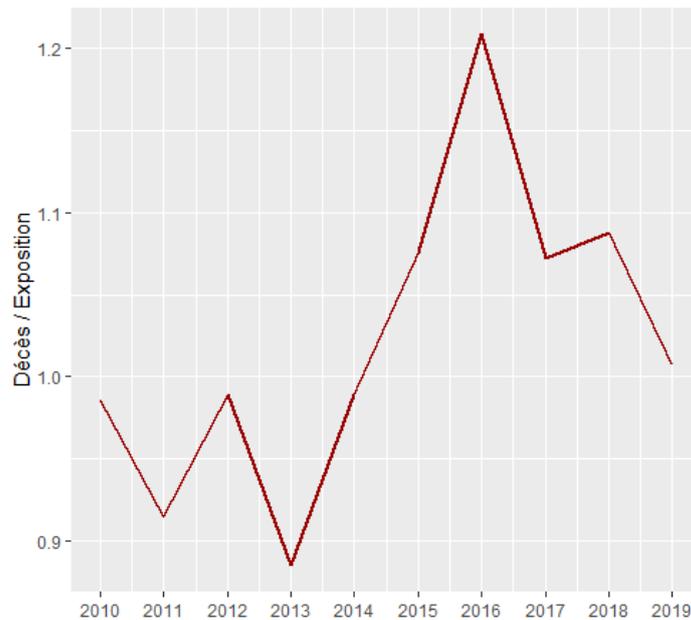


Figure III-13 : Rapport entre décès et exposition par année chez les hommes

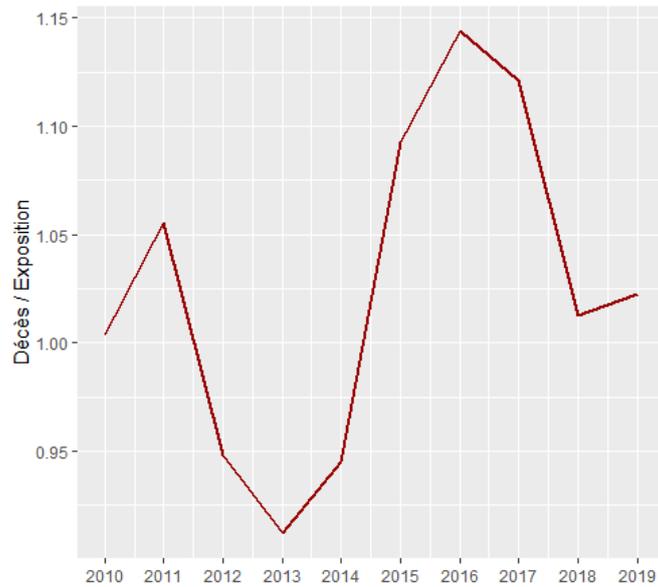


Figure III-14 : Rapport entre nombre de décès et exposition par année pour les femmes

En observant le rapport entre le nombre de décès et l'exposition au risque par année, aucune tendance explicite ressort. Cet aspect montre la non-proportionnalité du nombre de décès par rapport à l'exposition au risque sur notre portefeuille. Plusieurs raisons peuvent permettre d'interpréter ce résultat : la rencontre entre l'amélioration de la survie au fil des années et l'augmentation du nombre d'individus exposé à un décès ; mais aussi la faible quantité d'observations de notre portefeuille. Ce résultat pourrait se répercuter sur les estimations réalisées dans la suite de notre étude.

IV. Construction d'une table de mortalité prospective d'expérience

Le but de cette étude est de construire une table de mortalité prospective pour les individus de plus de 60 ans en s'appuyant sur le portefeuille assuré. La table de mortalité prospective devra aller au-delà des années d'observations et devra être pensée pour une tranche d'âge précise. Les observations vont de 2010 à 2019 pour des individus de plus de 60 ans. La tranche d'âge retenue pour cette estimation varie de 60 à 130 ans.

Une multitude de méthodes sont envisageables mais 3 grands axes ont été retenus :

- Une première vision s'appuyant sur les modèles de durées et les méthodes stochastiques. Parmi ces modèles, celui de Lee-Carter reste le plus connu et est considéré comme précurseur des modèles stochastiques de construction de table. Le modèle qui sera appliqué dans le cadre de cette partie est celui de Cairns-Blake-Dowd, souvent considéré, d'après les différentes lectures, comme un modèle qui fournit de bons résultats aux grands âges.
- Une seconde partie s'appuyant sur des tables de références (en l'occurrence les TGF05 et TGH05) sera étudiée. Cela permettra de partir sur une base assez solide en la présence de tables de références et de pouvoir appliquer des méthodes de positionnement pour ajuster les tables de références aux données.
- Le dernier axe est associé au souhait de construire un modèle prenant en considération l'ensemble des ressources mises à disposition par notre base de données. Les variables propres à l'assuré pourront donc être utilisées pour pouvoir prédire des probabilités de décès plus propres à chaque individu. Dans le cadre de ce dernier axe, c'est une régression logistique binomiale qui a été faite.

Des méthodes de fermeture de table seront également appliquées sur les modèles issus des deux premiers axes afin de pouvoir capter le modèle le plus fiable mais aussi le plus révélateur du portefeuille d'observations.

A. Validité par critère de Cochran

Avant de construire une table d'expérience, il faut vérifier que la base est assez riche. Le critère de Cochran émet l'hypothèse qu'au minimum 80% des classes d'âges doivent posséder au minimum cinq sinistres et que toutes les classes d'âges doivent en posséder au moins 1.

Soit $\hat{q}_x = D_x/N_x$ l'estimateur empirique du taux de mortalité à l'âge x . Il est sans biais, convergent et asymptotiquement normal. Cochran veut que pour chaque année d'observation :

$$\begin{cases} N_x * \hat{q}_x \geq 5 \\ N_x * (1 - \hat{q}_x) \geq 5 \end{cases} \text{ pour 80\% des } x \quad (1)$$

$$N_x * \hat{q}_x \geq 1 \text{ pour tout } x \quad (2)$$

En commençant par la tranche d'âge allant de 60 à 110 ans, les résultats obtenus sont les suivants :

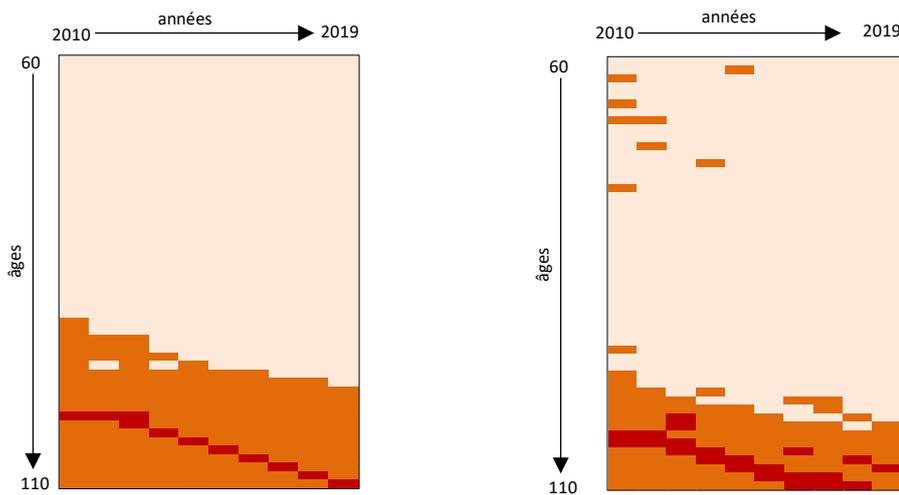


Figure IV-1 : Zonier de validation des critères de Cochran pour les hommes (à gauche) et les femmes

Sur ce zonier, les âges et les années qui ne respectent pas le premier critère de Cochran (1) sont visibles en orange. Le rouge met avant le non-respect du second critère (2). Au fil des âges, le portefeuille compte de moins en moins d'observations ; rendant la validité de la base de données plus compliqué aux grands âges. Il est donc impossible, d'après le critère de Cochran, de construire une table prospective avec cette base en allant jusqu'à un âge d'observation maximum de 110 ans.

En réajustant notre portefeuille avec un maximum s'élevant à 100 ans, les critères de validité de Cochran sont respectés. C'est cette nouvelle tranche d'âge qui servira pour la suite de l'étude. Le nombre de décès restant s'élève à 19 115.

B. Exposition au risque et calcul des taux bruts de mortalité

D'un point de vue pratique, pour évaluer $q_x(t)$ le taux de mortalité à l'âge x sur l'année t ; il est indispensable de bien distinguer la décomposition temporelle réalisée (Planchet et Guilbert, 2017) :

- Il est possible de se placer dans le cas d'une décomposition par cohorte, également appelée par génération. Cela revient à regrouper l'ensemble des individus nés à l'an $t - x$. En prenant σ comme date de naissance d'un individu on aura la formule générale :

$$q_x(t) = P(T < x + 1 | T \geq x, \sigma \in]t - x, t - x + 1])$$

- La seconde option est le cadre d'une décomposition par période où sont regroupés uniquement les individus atteignant l'âge x sur l'année t :

$$q_x(t) = P(T < x + 1, T + \sigma < t + 1 | T \geq x, T + \sigma \geq t)$$

Soient $L_x(t)$, $D_x(t)$ respectivement le nombre de personnes et le nombre de décès à l'âge x sur l'année calendaire t . Dans un cadre sans censure, D_x se note aussi : $D_x(t) = L_x(t) - L_{x+1}(t)$.

C'est grâce à cette base qu'est définie l'exposition au risque. L'exposition au risque est le temps passé en vie des personnes ayant un âge donné entre ces deux valeurs sur la période d'observation. $E_x(t)$ est donc l'exposition au risque sur un an pour l'âge x et l'année t , elle se calcule avec la formule suivante : $E_x(t) = \int_0^1 L_{x+\varepsilon}(t + \varepsilon) d\varepsilon$.

Dans l'approche par période, sur une même année, un individu aura une exposition au risque sur deux âges différents séparés par sa date d'anniversaire. $E_x(t)$ s'écrit comme la somme entre :

- $E_x^U(t)$, le temps passé par les individus nés en $t - x - 1$ avant leur anniversaire de l'année t
- Et $E_x^D(t)$, le temps passé par les individus nés en $t - x$ après leur anniversaire t .

$$E_x(t) = E_x^U(t) + E_x^D(t)$$

De façon analogue, le nombre de décès se calcule par la somme suivante : $D_x(t) = D_x^U(t) + D_x^D(t)$.

Soient :

- $(e_x(t))_i$: date d'entrée à l'âge x sur l'année t :

C'est le maximum entre $01/01/t$, le premier jour de l'année t ; le x -ième anniversaire de l'assuré et date d'effet du contrat de l'assuré i , s'il est souscrit dans l'année t

- $(s_x(t))_i$: date de sortie à l'âge x sur l'année t :

C'est le minimum entre $31/12/t$, le dernier jour de l'année t ; le $x+1$ -ième anniversaire et la date de sortie du portefeuille si elle est dans l'année t .

L'exposition au risque a donc été calculée comme la différence, en nombre de jours, entre ces deux dates :

$$(E_x(t))_i = (s_x(t))_i - (e_x(t))_i$$

Sur une même année, une observation peut donc être accordée à deux âges différents. En se plaçant dans le cas où l'information de l'observation j sur l'année t n'est pas censurée ; cette même observation d'âge x sur en début d'année participera à $E_x(t)$ avant son j -ième anniversaire puis il participera à $E_{x+1}(t)$. Afin de se rendre compte de l'évolution de cette variable dans notre portefeuille, nous avons tracé en 3 dimensions les expositions au risque par âge, sexe et année :

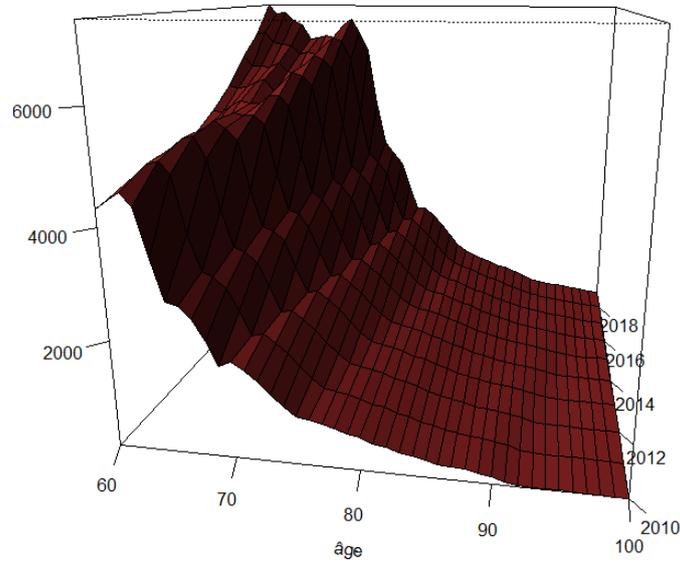


Figure IV-2 : Exposition au risque en jours par âge et année pour les hommes du portefeuille

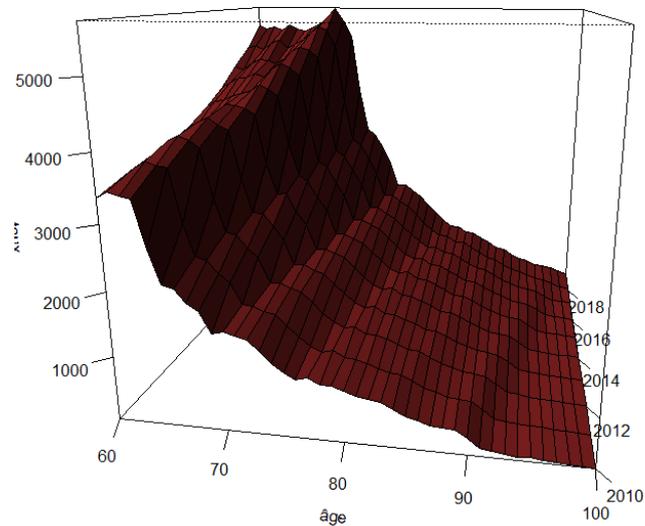


Figure IV-3 : Exposition en jours par âge et année pour les femmes du portefeuille

La tendance est clairement décroissante au fil des âges ; quel que soit le sexe. Une légère croissance de l'exposition au risque au fil des années est également observable. Cette croissance peut être due à une baisse de mortalité en fonction de l'année observée et à une augmentation du nombre de contrats dans la base. Les hommes sont plus exposés que les femmes mais cette affirmation est corrélée avec le fait que les hommes sont majoritaires dans le portefeuille.

Les taux bruts de mortalité, q_x , se calculent à partir des informations du portefeuille à disposition uniquement. Elles permettent de fournir une première indication sur la mortalité. Cet indicateur est totalement dépendant de nos données et ne permet pas d'évaluer la survie avec un modèle précis ; les q_x sont plus considérés comme point de départ à la modélisation.

1. L'estimateur binomial

L'estimateur binomial est le plus naturel qu'il existe. Deux hypothèses sont importantes pour l'appliquer. D'une part, l'indépendance des décès ; chaque décès est indépendant des autres. Et d'autre part, le nombre de décès à l'âge x doit suivre une loi binomiale $B(n_x, q_x)$ avec le paramètre n_x étant le nombre d'individus en vie de l'âge x à l'âge $x + 1$. C'est en utilisant cette base que sont déterminés les taux bruts estimés \hat{q}_x ; par maximum de vraisemblance notamment.

L'estimateur binomial se formule donc de la façon suivante :

$$\hat{q}_x = \frac{d_x}{n_x}$$

d_x s'inscrit comme une réalisation de D_x .

Cet estimateur ne prend visiblement pas en considération la censure. La formule utilisée sous-entend que la seule cause de sortie possible est le décès. Il est donc difficile d'utiliser cet estimateur aux observations du portefeuille qui multiplient les sorties pour rachat notamment.

2. L'estimateur d'Hoem du moment

C'est dans la continuité de l'estimateur binomial que s'inscrit l'estimateur d'Hoem. Cette nouvelle méthode permet de prendre en considération les phénomènes de censures et troncatures.

L'estimateur d'Hoem se base sur le fait que chaque observation i est définie comme un individu peut avoir 2 états X_i : vivant ou décédé. X_i s'inscrit donc comme une variable de Bernoulli prenant pour valeur 1 si l'individu décède et 0 sinon. C'est dans ce cadre que le nombre de décès D_x devient la somme des X_i . Il suit donc une loi de type binomiale avec pour paramètres n , le nombre d'individus observés sur la période (les n observations pour lesquelles $(E_x)_i > 0$) et q_x le taux de mortalité à l'âge x sur l'année.

Une autre hypothèse à prendre en compte est celle qu'un individu x vivant en x décède dans l'intervalle $[(e_x(t))_i, (s_x(t))_i]$ avec une probabilité $((s_x(t))_i - (e_x(t))_i) * q_x$.

La conséquence de cette relation est la mise en avant d'un estimateur de q_x appelé Estimateur d'Hoem :

$$\hat{q}_x = \frac{D_x}{\sum_i (E_x)_i}$$

L'estimateur du moment d'Hoem appliqué à nos données donne les résultats suivants :

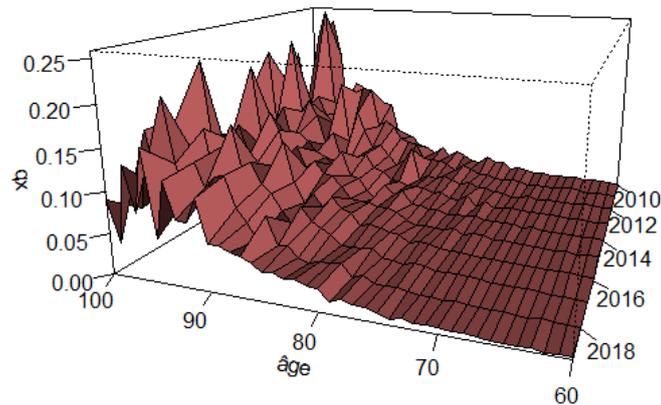


Figure IV-4 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les hommes

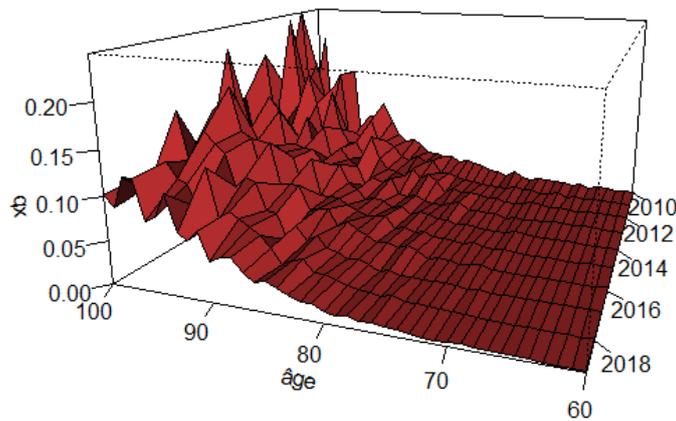


Figure IV-5 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les femmes

La tendance est la même pour les hommes et pour les femmes, les taux augmentent remarquablement à partir de 80 ans pour conserver une croissance allant jusqu'aux âges les plus élevés. En étant précautionneux sur les échelles utilisées, il est possible d'observer que les taux mettent plus de temps à atteindre leur tendance maximale chez les femmes. Outre cette différence, des taux « anormaux » sont visibles aux grands âges, quel que soit le sexe. Cette volatilité des résultats aux grands âges est la conséquence de la chute du nombre d'observations à disposition.

3. L'estimateur de Kaplan-Meier

« Être vivant à l'instant t c'est d'avoir survécu jusqu'à l'instant t et ne pas décéder à l'instant t ». C'est sur ce principe assez trivial que repose le modèle de Paul MEIER et Edward L.KAPLAN (1983), Kaplan-Meier. L'estimateur de Kaplan-Meier est un estimateur non paramétrique de la fonction de

survie qui permet de prendre en compte les phénomènes de censure. Il est estimé à partir de la définition de la fonction de survie et de quelques principes de conditionnement. Soit $t' > t$:

$$S(t') = P(T \geq t') = P(T > t', T > t) = P(T > t' | T > t)P(T > t) = P(T > t' | T > t)S(t)$$

Il y a donc une relation entre la fonction de survie à t' et à t : $S(t') = P(T > t' | T > t)S(t)$

Or la probabilité $P(T > t' | T > t)$ est la probabilité de survivre après t' sachant que l'on a survécu au-delà de t . Si l'on prend $t = t_{i-1}$ et $t' = t_i$ l'instant suivant, il y aura :

$$P(T > t_i | T > t_{i-1}) = 1 - P(T \leq t_i | T > t_{i-1})$$

Et donc $1 -$ la probabilité de décéder en t_i sachant que l'individu a vécu jusqu'à t_{i-1} . Ce qui équivaut à prendre $1 -$ le rapport entre le nombre de décès et le nombre de personnes en t_i .

$$P(T > t_i | T > t_{i-1}) = 1 - \frac{D_i}{N_i}$$

Avec N_i le nombre d'individus observés sur la période i et n_i sa réalisation.

On retrouve finalement :

$$\hat{S}(t) = S(t-1) \left(1 - \frac{d_t}{n_t}\right) = \prod_{i=1}^t \left(1 - \frac{d_i}{n_i}\right)$$

Cet estimateur empirique de la fonction de survie permet de retrouver facilement une estimation des taux bruts de mortalité :

$$\hat{q}_x = 1 - P(T < x + 1 | T > x) = 1 - \prod_{i=x}^{x+1} \left(1 - \frac{d_i}{n_i}\right)$$

En appliquant cet estimateur à notre portefeuille, les graphiques suivants sont le résultat des taux bruts de Kaplan-Meier :

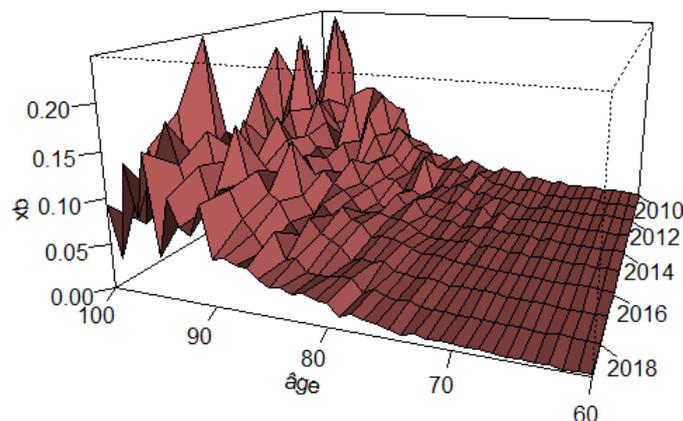


Figure IV-6 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les hommes

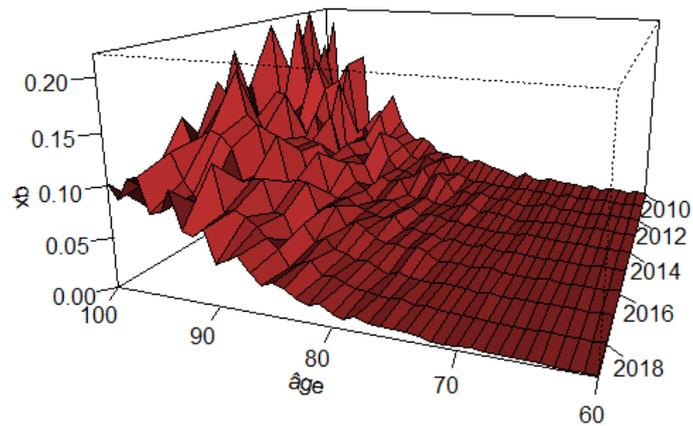


Figure IV-7 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les femmes

Les observations qui découlent de ces taux restent les mêmes que celles de l'estimateur du moment. Une forme « en escalier » se dégage à partir de 80 ans. Il conviendra maintenant de comparer les résultats obtenus par les deux méthodes

4. Comparaison des méthodes

Ces deux estimations par période de taux bruts de mortalité s'inscrivent sur une tendance assez similaire. Ces estimateurs restent tout de même assez différents dans la conception et il conviendra de n'en garder qu'un seul pour la suite de l'étude.

Le fait que Kaplan-Meier soit un estimateur non paramétrique permet de ne pas avoir à adopter une quelconque spécification de loi tandis que Hoem permet d'avoir des résultats satisfaisants à condition que les hypothèses paramétriques adoptées soient proches de la loi de mortalité.

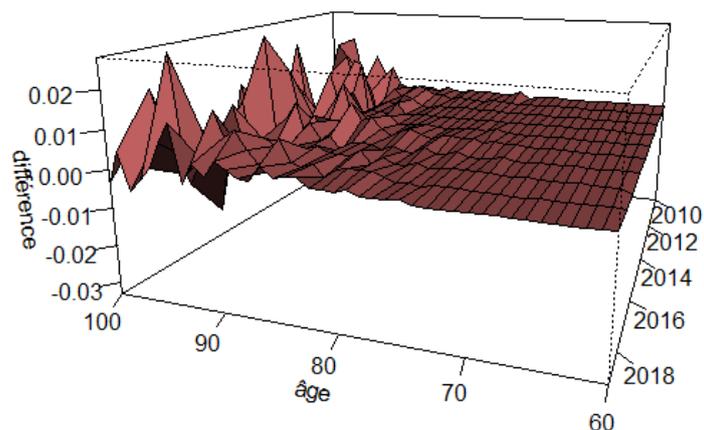


Figure IV-8 : Différence des taux de décès entre Hoem et Kaplan-Meier pour les hommes

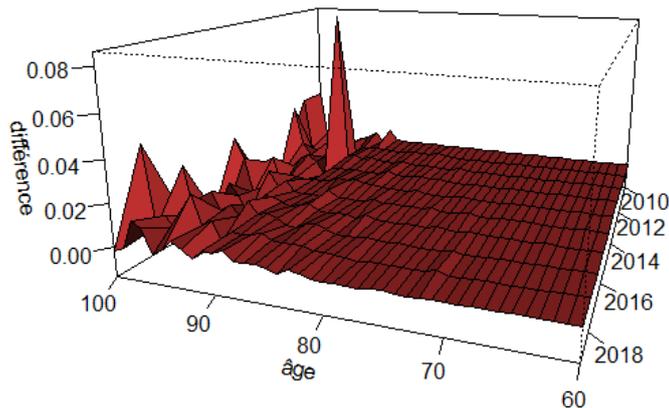


Figure IV-9 : Différence des taux de décès entre Hoem et Kaplan-Meier pour les femmes

En représentant graphiquement les différences des \hat{q}_x d'Hoem et de Kaplan-Meier, une tendance plus ou moins égale sur les premiers âges d'observations est visible. C'est à partir d'environ 90 ans que la différence est flagrante entre les deux méthodes. Ces différences restent minimales puisqu'elles ne dépassent que très rarement les 0,04.

Il est difficile d'interpréter les différences entre ces deux estimateurs et donc de les départager. C'est pour cette raison que la variance intervient :

- Variance des \hat{q}_x calculés avec l'estimateur d'Hoem du moment :

L'estimateur d'Hoem se sert de l'hypothèse forte dans laquelle la somme des décès dans l'année suit une loi binomiale de paramètres n_x et q_x . En utilisant l'approximation normale de cette loi binomiale, la variance de q_x se déduit facilement avec la formule suivante :

$$V(\hat{q}_x) = \sqrt{\frac{\hat{q}_x(1 - \hat{q}_x)}{E_x}}$$

- Variance des \hat{q}_x calculés avec l'estimateur de Kaplan-Meier :

Il existe plusieurs façons d'obtenir la variance de l'estimateur de Kaplan-Meier ; parmi elles, la méthode la plus utilisée reste par l'estimateur de Greenwood.

Pour rappel :

$$\hat{S}(t) = \prod_{i=1}^t \left(1 - \frac{d_i}{n_i}\right)$$

$$\ln(\hat{S}(t)) = \sum_{i=1}^t \ln\left(1 - \frac{d_i}{n_i}\right) = \sum_{i=1}^t \ln(1 - \hat{q}_i)$$

Or $\ln(1 - \hat{q}_i)$, sont des variables indépendantes. La méthode delta peut donc s'appliquer pour estimer la variance.

Méthode delta : $V(f(X)) \approx \left(\frac{\delta f}{\delta x} E(X)\right)^2 V(X)$

$$V(\ln(1 - \hat{q}_i)) \approx \left(\frac{\delta}{\delta x} \ln(1 - \hat{q}_i)\right)^2 V(1 - \hat{q}_i) = \frac{\hat{q}_i}{n_i(1 - \hat{q}_i)}$$

La variance est déduite du logarithme de la fonction de survie :

$$V(\ln(\hat{S}(t))) = \sum_{i=1}^t \frac{\hat{q}_i}{n_i(1 - \hat{q}_i)} = \sum_{i=1}^t \frac{d_i}{n_i(n_i - d_i)}$$

Puis à nouveau par la méthode delta :

$$V(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i=1}^t \frac{d_i}{n_i(n_i - d_i)}$$

C'est $\gamma(t) = \sqrt{\sum_{i=1}^t \frac{D_i}{n_i(n_i - D_i)}}$ appelé estimateur de Greenwood. Il est utile à la construction d'intervalles de confiance asymptotique pour l'estimateur de Kaplan-Meier ; notamment en appuyant la normalité asymptotique de cette variance.

En s'appuyant sur les formules de \hat{q}_x suivantes :

$$\hat{q}_x = 1 - \prod_{i=x}^{x+1} \left(1 - \frac{d_i}{n_i}\right) \quad \text{et} \quad \hat{q}_x = 1 - \frac{\hat{S}(t+1)}{\hat{S}(t)}$$

La variance des taux bruts de mortalité se calcule avec la formule suivante :

$$V(\hat{q}_x) = (1 - \hat{q}_x)^2 \sum_{i=x}^{x+1} \frac{d_i}{n_i(n_i - d_i)}$$

En appliquant ces formules aux résultats obtenus, des variances par âge et année sont obtenues. En ne s'intéressant qu'aux variances par année, les graphiques suivants sont tracés :

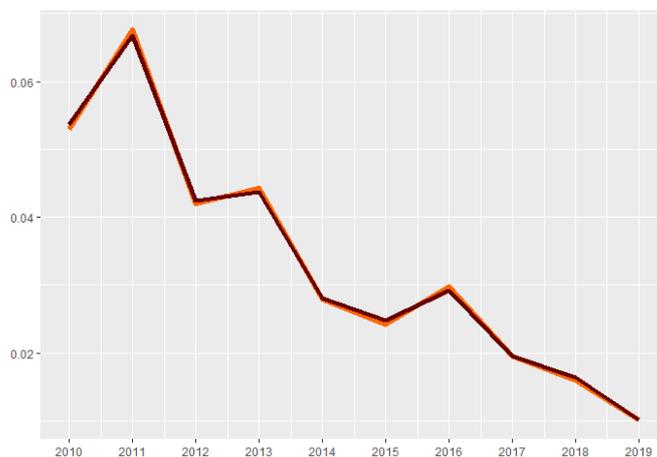


Figure IV-10 : Variance des taux de décès d'Hoem et de Kaplan-Meier par année pour les hommes

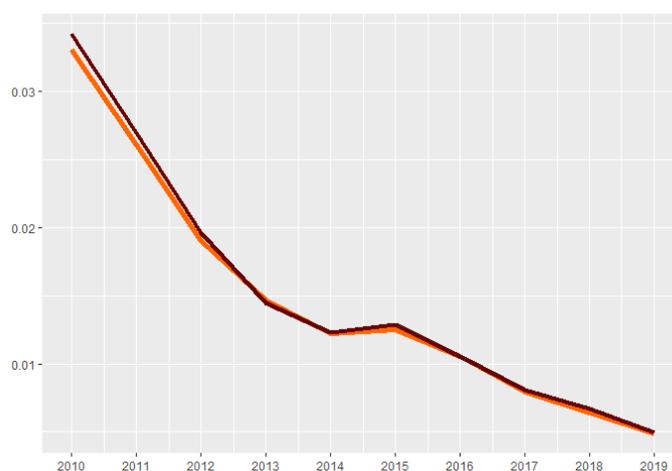


Figure IV-11 : Variance des taux de décès d'Hoem et de Kaplan-Meier par année pour les femmes

En bordeaux, la variance par année de l'estimateur de Kaplan-Meier et en orange celles des taux bruts d'Hoem du moment. Une première observation porte sur la tendance décroissante de ces courbes. Cette chute est en corrélation avec l'augmentation du nombre d'individus par année et donc avec le nombre d'observations à disposition. De plus, la différence entre la variance de l'estimateur Kaplan-Meier et d'Hoem n'est pas flagrante. Les deux courbes sont à chaque fois très proches. Les variances sont proches en raison de leur forte sensibilité au nombre d'observations.

La courbe orange reste tout de même globalement au-dessous de la rouge. C'est assez visible pour la population féminine et c'est également le cas pour les hommes (malgré des dépassements à certains pics). Ce résultat laisse penser que prendre l'estimateur des moments de Hoem est moins risqué que prendre celui de Kaplan-Meier.

C. Modèles stochastiques de projection de la mortalité

1. Lee-Carter

Le modèle de LEE Ronald et CARTER Lawrence de 1992 (Lee-Carter) est une méthode d'extrapolation des tendances passées datant de 1992 qui était à l'origine utilisée sur les données américaines. Lee & Carter proposent de décomposer la force de mortalité à une composante propre à l'âge puis d'extrapoler la composante relative au temps. Le modèle est donc dépendant de l'âge et du temps. La formule des taux instantanés de mortalité est la suivante :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$$

avec $\varepsilon_{x,t}$ des variables aléatoires i.i.d selon une loi normale centrée et de variance σ^2 .

Il convient alors d'ajuster une structure paramétrique et un phénomène aléatoire à cette série des logarithmes de taux de décès instantanés.

Trois paramètres se distinguent de cette formule :

- α_x , ce paramètre s'interprète comme le comportement moyen des taux instantanés au cours du temps ;
- k_t est l'évolution générale de la mortalité au cours du temps ;
- β_x est un coefficient qui traduit la sensibilité instantanée à l'âge x par rapport à l'évolution générale explicitée ci-dessus. Plus ce coefficient sera haut, plus k_t aura un impact sur cet âge.

2. Cairns-Blake-Dowd

a) Théorie

C'est dans la continuité du modèle de Lee-Carter que s'inscrit le modèle de CAIRNS Andrew J. G., BLAKE David et DOWD Kevin de 2006 (CBD). Souvent considéré dans la littérature comme un modèle qui est performant sur les grands âges, le modèle CBD considère les probabilités de décès selon la structure suivante :

$$\text{logit}(\tilde{q}_x(t)) = \log\left(\frac{\tilde{q}_x(t)}{1 - \tilde{q}_x(t)}\right) = K_t^{(1)} + K_t^{(2)}(x - \bar{x}) + \varepsilon_{x,t}$$

Ce modèle se décompose en plusieurs facteurs :

- Le premier capture la tendance générale dans l'évolution (du *logit*) des probabilités de décès. $K_t^{(1)}$ est un processus stochastique qui représente l'*intercept* du modèle. Il affecte tous les âges de la même façon et donne le niveau de mortalité au temps t .

Si $K_t^{(1)}$ a une tendance décroissante en fonction du temps, cela se répercutera sur les taux de décès qui suivront la même tendance.

- Le second facteur permet de laisser évoluer les probabilités de décès à travers le temps en fonction de variables spécifiques de chaque tranche d'âge. Comme $K_t^{(1)}, K_t^{(2)}$ est un processus stochastique qui représente la « pente » du modèle. Chaque âge y est affecté différemment. Si dans notre portefeuille, les améliorations de mortalité sont plus importantes aux âges inférieurs qu'aux âges supérieurs ; $K_t^{(2)}$ augmente avec le temps. On applique ce paramètre sur $x - \bar{x}$, l'âge centré (\bar{x} étant l'âge moyen).
- Comme dans le modèle de Lee-Carter, les $\varepsilon_{x,t}$ sont également i.i.d et suivent une $\mathcal{N}(0, \sigma^2)$.

En se plaçant sur des âges allant de 60 à 100 ans, de 2010 à 2019, le modèle de CBD peut être visualisé de la façon suivante :

$$\begin{pmatrix} \text{logit } \tilde{q}_{60}(2010) & \text{logit } \tilde{q}_{60}(2011) & \dots & \text{logit } \tilde{q}_{60}(2019) \\ \text{logit } \tilde{q}_{61}(2010) & \text{logit } \tilde{q}_{61}(2011) & \dots & \text{logit } \tilde{q}_{61}(2019) \\ \vdots & \vdots & \ddots & \vdots \\ \text{logit } \tilde{q}_{100}(2010) & \text{logit } \tilde{q}_{100}(2011) & \dots & \text{logit } \tilde{q}_{110}(2019) \end{pmatrix} = \begin{pmatrix} K_{2010}^{(1)} & K_{2011}^{(1)} & \dots & K_{2020}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ K_{2010}^{(1)} & K_{2011}^{(1)} & \dots & K_{2020}^{(1)} \end{pmatrix} + \begin{pmatrix} -20 \\ -19 \\ \vdots \\ 20 \end{pmatrix} (K_{2010}^{(1)} \dots K_{2020}^{(1)})$$

En se concentrant sur un âge et une année en particulier ($x = 95, t = 2015$ par exemple), les paramètres estimés sont ceux d'une tendance linéaire sur l'âge :

$$\text{logit } \tilde{q}_{95}(2015) = K_{2015}^{(1)} + 15 \times K_{2015}^{(2)}$$

L'expression (du *logit*) du taux de décès permet de bien distinguer ce modèle à celui de Lee-Carter pour lequel aucune valeur à β_x doit être estimée et n'est pas imposée. Ici, β_x est assimilé à $x - \bar{x}$.

Il existe plusieurs méthodes pour estimer les paramètres du modèle ; c'est à partir des \hat{q}_x , taux bruts estimés avec l'estimateur d'Hoem, que les estimations sont faites. Le maximum de vraisemblance ou la méthode des moindres carrés ordinaires peuvent notamment être utilisés pour cette estimation. La méthode des moindres carrés ordinaires sera retenue pour l'estimation des paramètres par soucis de simplicité.

Voici l'application théorique de la méthode :

Soient A le nombre d'âges présents dans l'échantillon et B le nombre d'années.

$$J = \sum_{x=x_1}^{x_A} \sum_{t=t_1}^{t_B} \left(\log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) - K_t^{(1)} - K_t^{(2)}(x - \bar{x}) \right)^2$$

En dérivant J par nos deux paramètres, on obtient :

$$\frac{\partial J}{\partial K_t^{(1)}} = \sum_{x=x_1}^{x_A} 2 \left(\log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) - K_t^{(1)} - K_t^{(2)}(x - \bar{x}) \right)$$

$$\frac{\partial J}{\partial K_t^{(2)}} = \sum_{x=x_1}^{x_A} 2(x - \bar{x}) \left(\log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) - K_t^{(1)} - K_t^{(2)}(x - \bar{x}) \right)$$

Le but étant de minimiser la fonction J , voici les formules permettant d'estimer pour chaque année les deux paramètres :

$$\frac{\partial J}{\partial K_t^{(1)}} = 0 \Leftrightarrow \widehat{K_t^{(1)}} = \frac{\sum_{x=x_1}^{x_A} \log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) - \widehat{K_t^{(2)}} \sum_{x=x_1}^{x_A} (x - \bar{x})}{A}$$

$$\frac{\partial J}{\partial K_t^{(2)}} = 0 \Leftrightarrow \widehat{K_t^{(2)}} = \frac{A \sum_{x=x_1}^{x_A} (x - \bar{x}) \log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) - \sum_{x=x_1}^{x_A} \log \left(\frac{\hat{q}_x(t)}{1 - \hat{q}_x(t)} \right) \sum_{x=x_1}^{x_A} (x - \bar{x})}{A \sum_{x=x_1}^{x_A} (x - \bar{x})^2 - \left(\sum_{x=x_1}^{x_A} (x - \bar{x}) \right)^2}$$

b) Application

En appliquant ces formules sur la population masculine du portefeuille, les paramètres estimés sont les suivants :



Figure IV-12 : $\widehat{K_t^{(1)}}$ estimé sur les 10 premières années pour la population masculine

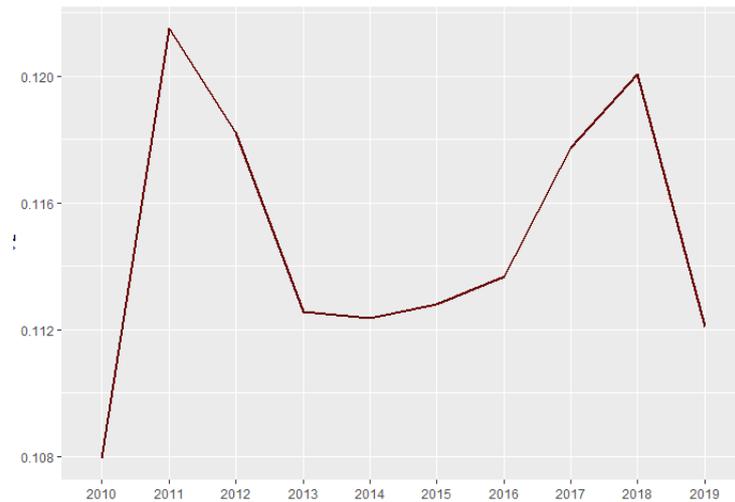


Figure IV-13 : $\widehat{K}_t^{(2)}$ estimé sur les 10 premières années pour la population masculine

Le premier graphique vise à capter l'évolution du premier paramètre $\widehat{K}_t^{(1)}$ estimé par la méthode des moindres carrés. Le résultat attendu sur ce graphique est une courbe décroissante en fonction du temps. Ce paramètre, qui reflète la tendance de la courbe de mortalité au fil du temps, doit avoir une tendance décroissante du fait de la réduction de la mortalité au fil des années. Nous n'obtenons pas le résultat souhaité.

Deux grands axes peuvent expliquer ce phénomène : la faible quantité d'observations à disposition dans notre portefeuille mais aussi un intervalle de temps trop modeste pour observer une réelle tendance. Il est important de rappeler que ce premier paramètre ne prend pas en compte les âges. Le niveau reflété par ce paramètre est à extrapoler sur tous les âges observés.

L'estimation du second paramètre $\widehat{K}_t^{(2)}$ est visible sur le second graphique. Le niveau de cette estimation donne une indication sur les âges les plus affectés par une éventuelle amélioration de la mortalité. On s'attend à une tendance croissante au fil des années ; indiquant une plus grande influence sur les âges les plus faibles. La courbe obtenue sur nos 10 années d'observations ne nous permet pas d'affirmer une tendance particulière.

Compte tenu du fait que la population masculine n'offre pas les tendances attendues sur les paramètres estimés ; et que les hommes sont majoritaires dans le portefeuille client, les paramètres estimés pour la population féminine ne devraient pas non plus afficher la tendance attendue.

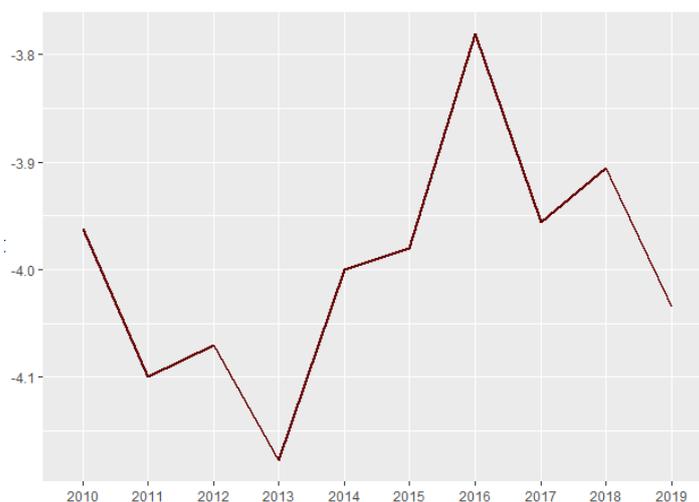


Figure IV-14 : $\widehat{K}_t^{(1)}$ estimé sur les 10 premières années pour la population féminine

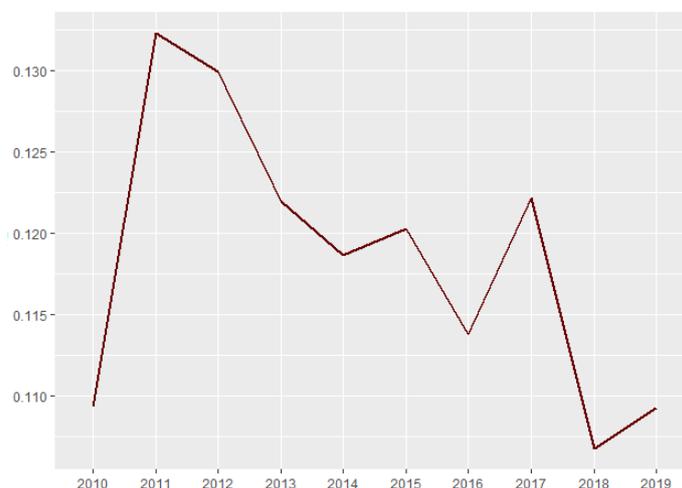


Figure IV-15 : $\widehat{K}_t^{(2)}$ estimé sur les 10 premières années pour la population féminine

Comme pour les hommes, les estimations de paramètres ne permettent pas d'avoir une tendance particulière. Le $\widehat{K}_t^{(2)}$ obtenu à partir de la population féminine va même à l'encontre de nos hypothèses en affichant une tendance décroissante à partir de 2011. Ce résultat voudrait dire que la réduction de la mortalité est davantage observable sur les âges les plus élevés.

Ces observations sont à prendre avec des pincettes. En effet, l'échelle sur laquelle les paramètres ont été projetés reste assez faible, avec des différences de \widehat{K} n'excédant que rarement les 0,02 d'une année à l'autre. Nous pouvons même dire que globalement, l'ensemble des estimations obtenues restent globalement constantes sur les années d'observation.

La suite du processus vise à projeter ces paramètres sur les années estimées, nous pourrions alors nous rendre compte si une tendance particulière s'y dégage. Avant cela, il est important de s'attarder sur les résidus du modèle. Les résidus sont supposés i.i.d et normaux.

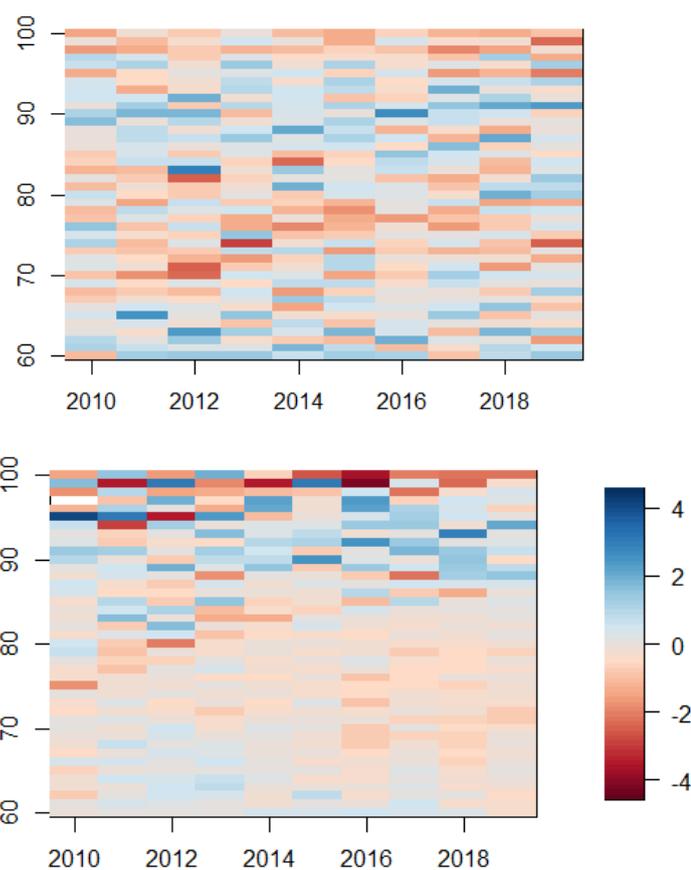


Figure IV-16 : Cartographie des résidus du modèle estimé pour les hommes et les femmes (respectivement)

Ces graphiques cartographient les résidus des deux modèles (homme puis femme). Cela permet d’avoir une première impression de la qualité de notre modèle. Plus les résidus sont répartis aléatoirement, plus l’hypothèse qu’ils sont normaux et i.i.d est respectée et plus le modèle est fiable.

La répartition est bonne pour la population masculine où les résidus semblent répartis aléatoirement au fil des âges et des années. Concernant la population féminine, une légère concentration semble exister sur les premiers âges (60 à 70 ans). Cette concentration est difficilement explicable ; elle est certainement liée à la faible quantité de données à disposition, influant sur la fiabilité du modèle.

En intégrant les paramètres estimés à la formule principale du modèle, les taux de mortalité modélisés pour chaque âge et année observée sont obtenus :

$$\text{logit}(\tilde{q}_x(t)) = \widehat{K}_t^{(1)} + \widehat{K}_t^{(2)}(x - \bar{x})$$

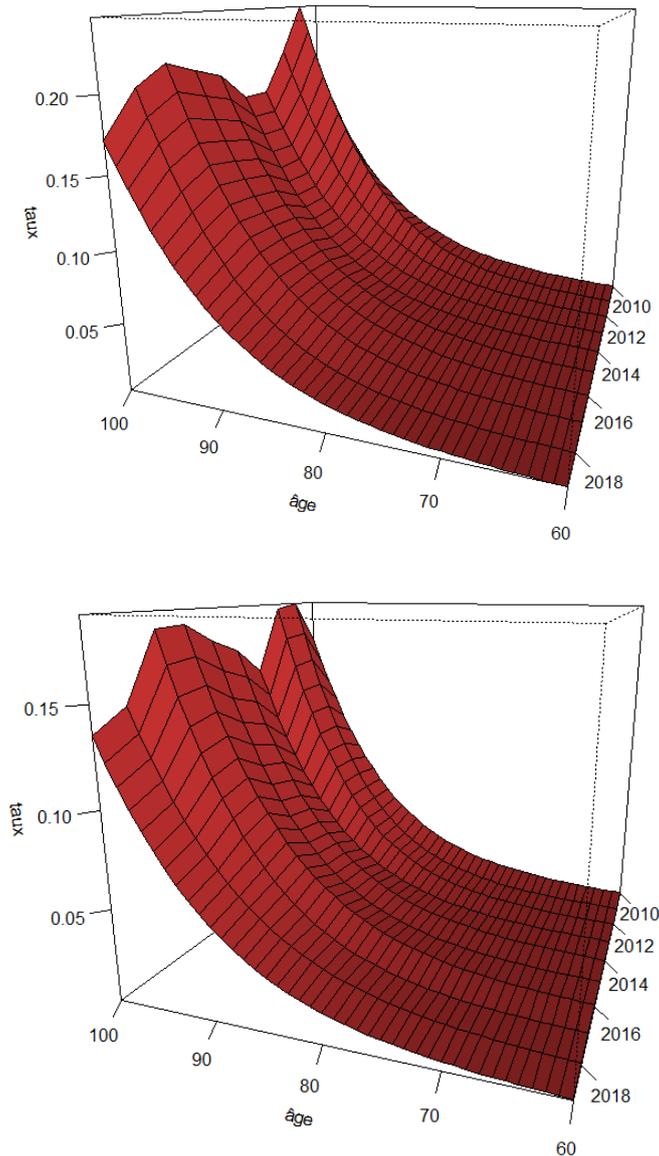


Figure IV-17 : $\tilde{q}_x(t)$ estimés pour la population masculine et féminine (respectivement)

Pour la population masculine, comme féminine, les probabilités de décès prennent la forme de « vagues » à certains moments. Un pic en 2011 est notamment visible.

Ces estimations en vague, ainsi que les paramètres estimés du modèle, se voient biaisées en partie par l'observation faite sur les statistiques descriptives, concernant l'évolution non-proportionnelle de la mortalité et de l'exposition au risque sur le portefeuille.

Nous souhaitons modéliser des taux de mortalité pour les 70 années suivantes. Il faut donc pouvoir prédire nos paramètres $K_t^{(1)}$, $K_t^{(2)}$ sur les années qui suivent l'observation.

Le modèle prédictif suggéré par Cairns, Blake & Dowd est une marche aléatoire à dérive bidimensionnelle qui s'articule de la façon suivante :

$$K_{t+1} = \mu + K_t + Cu(t + 1)$$

$$K_{t+1} = \begin{pmatrix} K_{t+1}^{(1)} \\ K_{t+1}^{(2)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} K_t^{(1)} \\ K_t^{(2)} \end{pmatrix} + \begin{pmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{pmatrix} \begin{pmatrix} u_{1,t+1} \\ u_{2,t+1} \end{pmatrix}$$

μ est un vecteur de dérive calculé comme la moyenne arithmétique de la série de paramètres estimés. C est une matrice qui découle directement de la matrice variance-covariance V de $K_t^{(1)}, K_t^{(2)}$. On veut $V = CC'$ avec comme contrainte C étant une matrice triangulaire supérieure. Cette matrice est aussi appelée la matrice de Choleski de V . $u(t + 1)$ est un vecteur aléatoire normal à deux dimensions. Nous suivons la pratique standard et estimons les systèmes par maximum de vraisemblance.

En se plaçant entre 2020 et 2089 et à partir des observations obtenues sur les 10 années passées, nous avons obtenu pour les hommes l'évolution des K_t suivante :

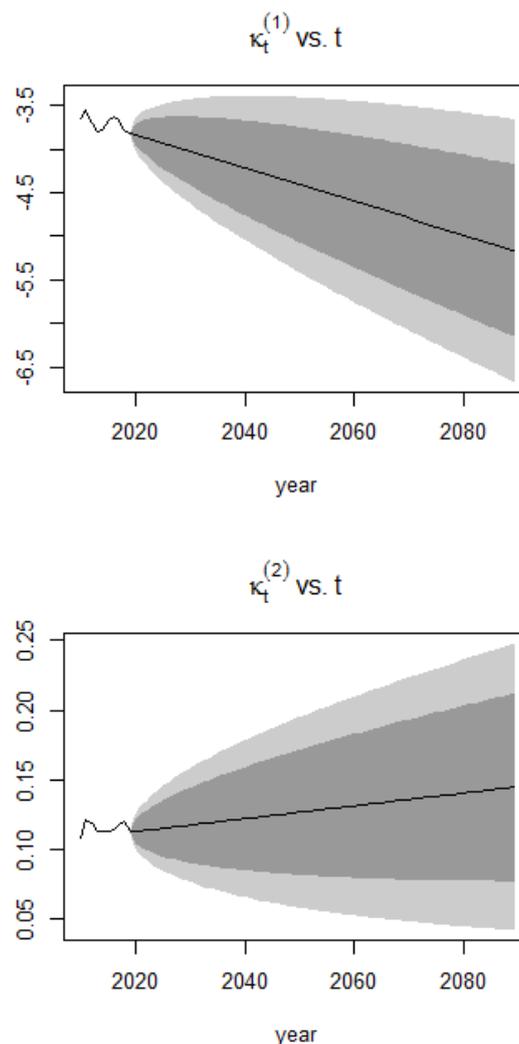


Figure IV-18 : Estimation prédictive des K_t du modèle sur la population masculine

L'estimation prospective des paramètres est encourageante pour la population masculine ; les tendances obtenues sont celles souhaitées avec notamment un premier paramètre décroissant du temps et un second qui semble prendre une direction croissante. En gris, les intervalles de confiance sont représentés avec 95% en clair et 80% en foncé. Les intervalles de confiance montrent que la marge d'erreur reste néanmoins assez large.

En générant des simulations aléatoires de K_t , par le biais de la fonction *simulate* du package *StMoMo*, les projections pour la population masculine donnent le résultat suivant :

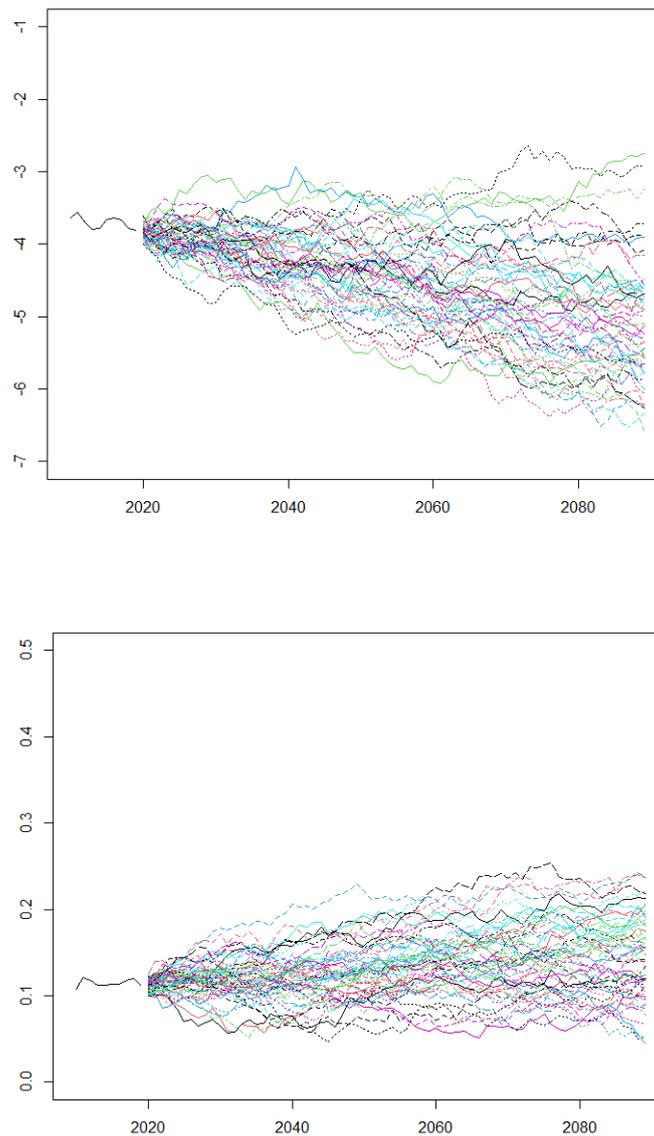


Figure IV-19 : Projections de scénarios aléatoires sur les paramètres K_t prospectifs pour les hommes

Le mouvement décroissant du premier paramètre semble majoritaire au vu des 100 projections effectuées. Concernant le second paramètre, il est plus difficile de capter une tendance croissante.

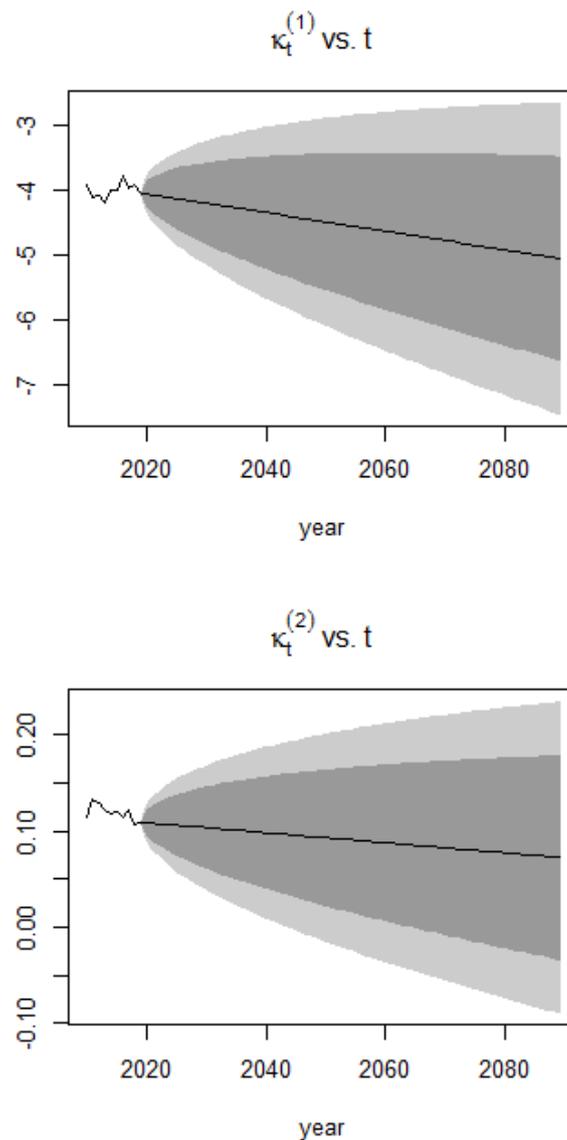


Figure IV-20 : Estimation prédictive des K_t du modèle sur la population féminine

Les projections pour les individus de sexe féminin sont un peu moins bonnes. Malgré qu'une légère tendance décroissante semble se dessiner pour le premier paramètre, nous n'avons pas la hausse attendue pour le second. De plus, les intervalles de confiance sont assez larges et ne permettent pas de s'appuyer sur des résultats de grande précision.

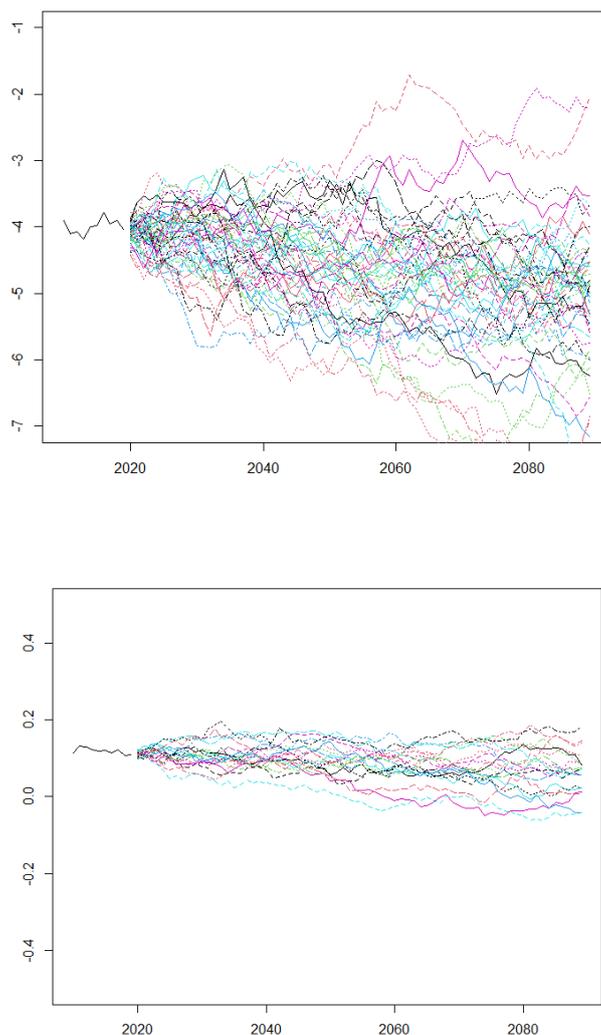


Figure IV-21 : Projections de scénarios aléatoires sur les paramètres K_t prospectifs pour les femmes

Il est plus difficile de capter sur ce premier graphique la baisse du premier paramètre. Par ailleurs, cette projection permet d'affirmer une certaine constance de $K_t^{(2)}$ sur la population féminine. Cette constance s'interprète en soulignant que l'âge n'a pas un lien direct avec le niveau de la réduction de la mortalité chez les femmes.

Ce modèle s'inscrit bien dans la continuité de celui de Lee-Carter. L'inconvénient de ce modèle, comparé à un Lee-Carter, vient de la valeur de $\beta_x (= x - \bar{x})$. Imposer une croissance linéaire en fonction de l'âge au *logit* ne reflète pas la réalité pour certains âges ; en particulier les plus jeunes. Ce modèle devient adéquat à partir d'un certain âge, qui tourne autour de 50-60 ans. Cette problématique ne devrait pas interférer dans notre modélisation.

Au contraire, CBD est avantageux sur certains points et permet de le distinguer des autres modèles du même type. Un avantage est notamment que les facteurs ne varient pas après l'ajout d'observations. Par exemple, sur un Lee-Carter, l'ajout d'une nouvelle année à t impacte les paramètres estimés sur des années antérieures.

Le biais pris sur les grands âges, notamment lié à la faible quantité d'informations, nous incite à utiliser un modèle de fermeture de table pour clore la modélisation.

c) Méthodes de fermeture de table et comparaisons

Afin de pouvoir extrapoler les résultats obtenus aux grands âges, et en particulier jusqu'à 130 ans, des méthodes de fermeture de table sont utilisées. Dans un premier temps, il convient de repérer à partir de quel âge l'extrapolation sera utilisée. Même si certaines méthodes sont par habitude utilisées à un certain âge (Coale-Kisker à 85 ans), il faut tout de même que l'âge retenu pour le début de l'extrapolation soit cohérent avec les résultats obtenus.

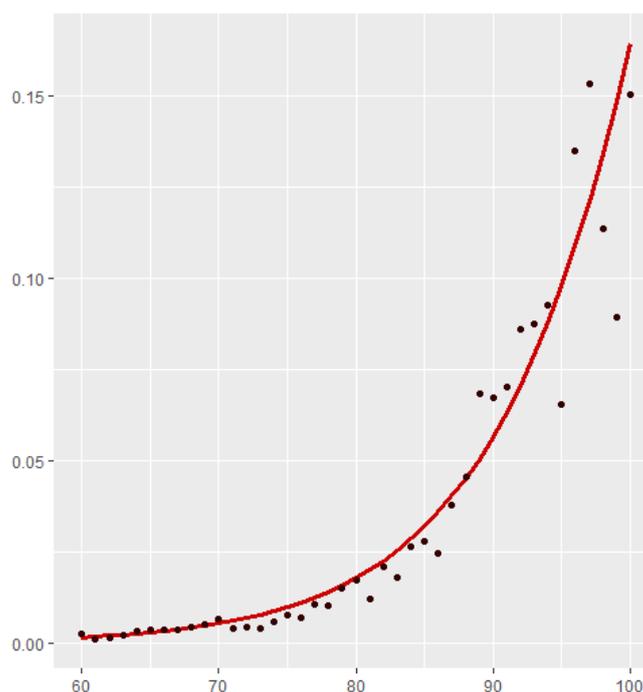


Figure IV-22 : Taux bruts estimés et observés sur la population féminine de 2014

Le graphique ci-dessus met en avant les \tilde{q}_x estimés à l'aide du modèle CBD en rouge et les q_x bruts estimés au préalable à partir de la méthode des moments d'Hoem, en noir ; pour la population féminine de 2014. Sur les premiers âges étudiés, la courbe ajustée semble proche des taux observés. Plus les âges avancent et moins la courbe s'ajuste aux observations ; en s'accompagnant de résidus de plus en plus grands. L'âge retenu comme âge où l'ajustement semble dériver est de 85 ans. C'est à partir de cet âge que seront appliquées nos méthodes de fermeture sur la population féminine.

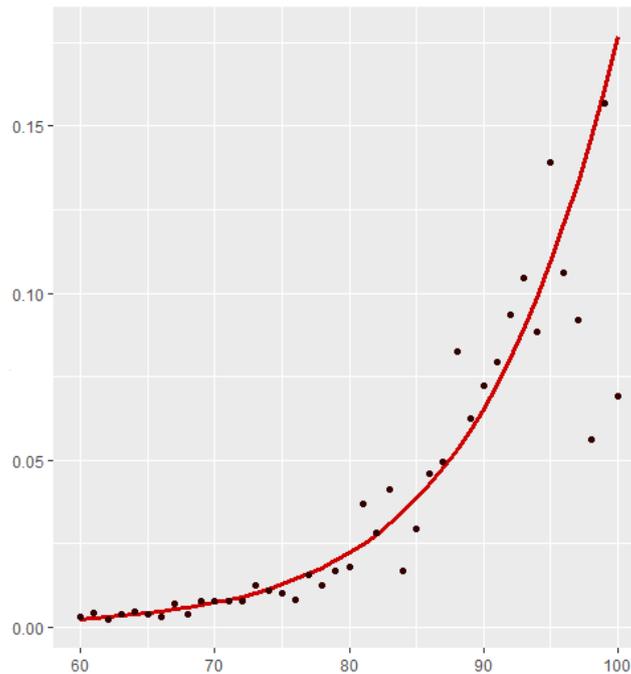


Figure IV-23 : Taux bruts estimés et observés sur la population masculine de 2014

En traçant ce même graphique, à la même année, pour la population masculine ; les observations sont les mêmes. Par soucis de fiabilité mais aussi de simplifications, c'est 85 ans qui sera l'âge retenu pour la population masculine également.

Quatre méthodes de fermeture de table ont été appliquées à notre modèle : Exponentielle, Denuit et Goderniaux, Coale-Kisker et Kannisto. Une fois les fermetures appliquées, des taux bruts après fermeture, \tilde{q}_x^F , sont obtenus pour la tranche d'âge [85,130] ans.

Afin de comparer ces méthodes de fermetures, plusieurs quantités ont été utilisées :

La déviance est une quantité mesurant la qualité de l'ajustement. Si les décès sont supposés suivre une loi de poisson : $D_x(t) \sim P(E_x(t)q_x(t))$; la déviance s'exprime alors de la façon suivante :

$$\left\{ \begin{array}{l} Dev_x(t) = 2 \left(D_x(t) \ln \left(\frac{q_x(t)}{\tilde{q}_x(t)} \right) - (D_x(t) - E_x(t)\tilde{q}_x(t)) \right) \text{ si } D_x(t) > 0 \\ Dev_x(t) = 2(E_x(t)\tilde{q}_x(t)) \text{ sinon} \end{array} \right.$$

Ainsi, la déviance totale du modèle est la somme des déviances à chaque âge et année observée. Une déviance faible indique que le modèle est bon.

Le X^2 est une distance qui se calcule en faisant pour chaque âge et année le ratio entre la différence taux lissés, taux bruts au carré et les taux bruts ; le tout appliqué à l'exposition au risque. Cette distance donne une indication sur le bon ajustement des taux lissés aux taux bruts. Le X^2 est souhaité être le plus bas possible.

$$X^2 = \sum_{x,t} E_x(t) \frac{(\tilde{q}_x(t) - q_x(t))^2}{q_x(t)}$$

Le *MAPE*, ou *Mean Absolute Percentage Error*, est un pourcentage d'erreur permettant de mesurer l'exactitude de l'ajustement par rapport aux données. Le *MAPE* est une distance se basant sur les mêmes principes que le χ^2 . Voici la formule utilisée :

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

n correspond au nombre de taux à lisser.

Les tableaux ci-dessous peuvent servir d'outil de comparaison entre ces méthodes, en utilisant notamment la distance du χ^2 entre les taux obtenus et les observations, la déviance et le MAPE. D'autres méthodes de comparaisons ont été faites concernant le nombre de décès sujet à chaque méthode. Puisque le nombre d'observations devient trop faible à partir de 101 ans (c.f. critère de validité de Cochran), les comparaisons se font entre 85 et 100 ans.

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	212,76	193,47	148,54	149,42
MAPE	32,88	25,99	25,24	26,83
Déviance	2731	1837	2020	1895

Tableau 8 : Comparaison des différentes fermetures de table pour un CBD sur la population masculine

Les quantités de comparaisons semblent afficher de meilleurs résultats pour la méthode de Coale-Kisker. Cette comparaison donne donc comme résultat que cette méthode est la moins risquée des quatre.

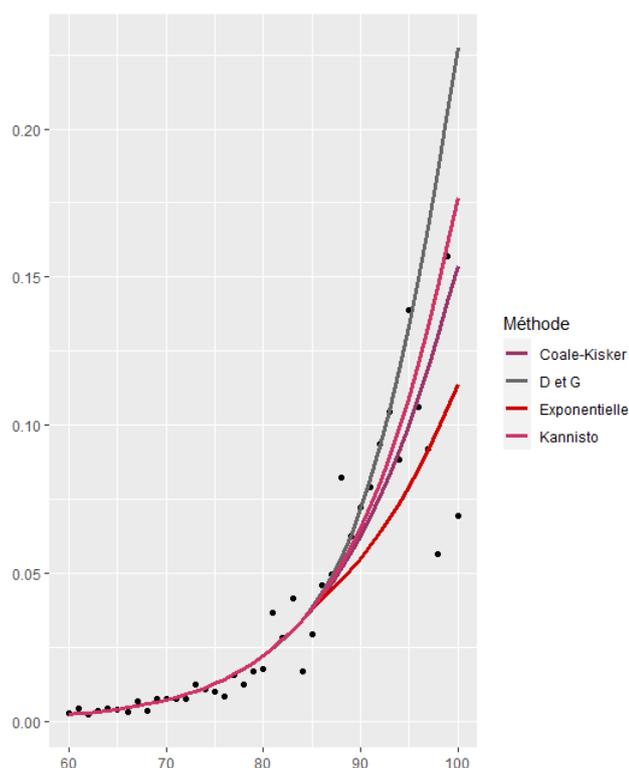


Figure IV-24 : Taux bruts et estimés par différentes fermetures sur les hommes en 2014

L'analyse graphique permet de constater qu'à partir de 90 ans, les taux de décès semblent être surestimés par les méthodes Denuit & Goderniaux et légèrement par Coale-Kisker. Le plus prudent, au vu du risque de survie, serait de choisir la méthode qui surestime le moins les taux de décès mais compte tenu des résultats précédents et des bons résultats graphiques qu'offre Coale-Kisker, c'est cette méthode qui est retenue pour fermer la table.

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	306,526	269,18	181,98	182,34
MAPE	33,65	21,09	24,85	21,67
Déviante	1521	2784	1510	1891

Tableau 9 : Comparaison des différentes fermetures de table pour un CBD sur la population féminine

Chez les femmes, les résultats sont assez partagés entre les quatre méthodes. En comparant avec les résultats du groupe masculin, les taux de décès sont globalement mieux prédits, avec des erreurs de prédictions plus faible ; en particulier la déviance résiduelle.

Il faut trouver le bon compromis entre une fermeture prudente et qui reste assez proche de la réalité. C'est pour cette raison que l'analyse graphique intervient à nouveau.

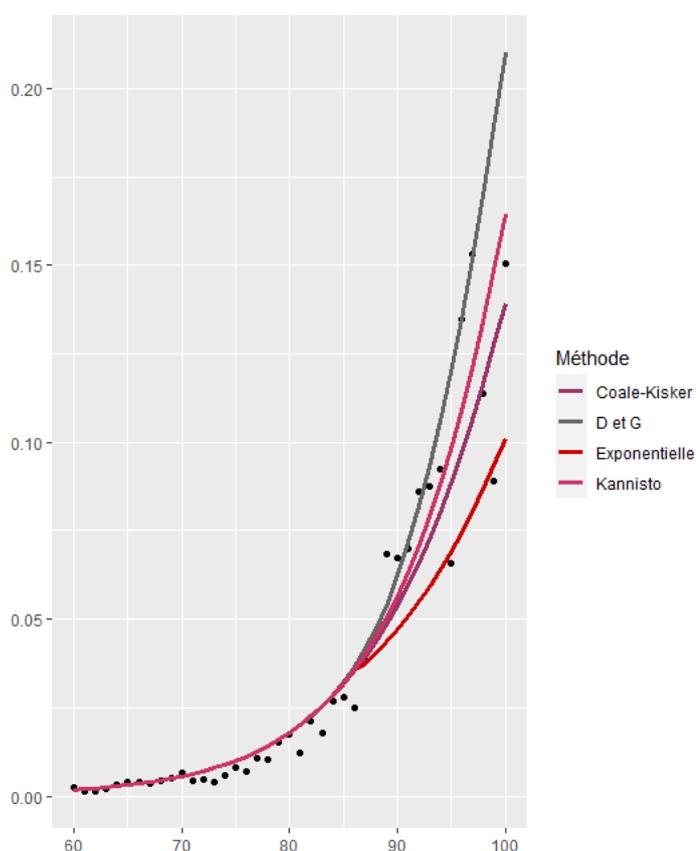


Figure IV-25 : Taux bruts et estimés par différentes fermetures sur les femmes en 2014

Le résultat obtenu ressemble sensiblement à celui des hommes. La fermeture par la méthode de Coale-Kisker est retenue pour les mêmes raisons que celles de la population masculine.

Les graphiques obtenus pour une modélisation de la mortalité du portefeuille par la méthode Cairns-Blake-Dowd clôturée par une fermeture de table de type Coale-Kisker donnent les résultats suivants :

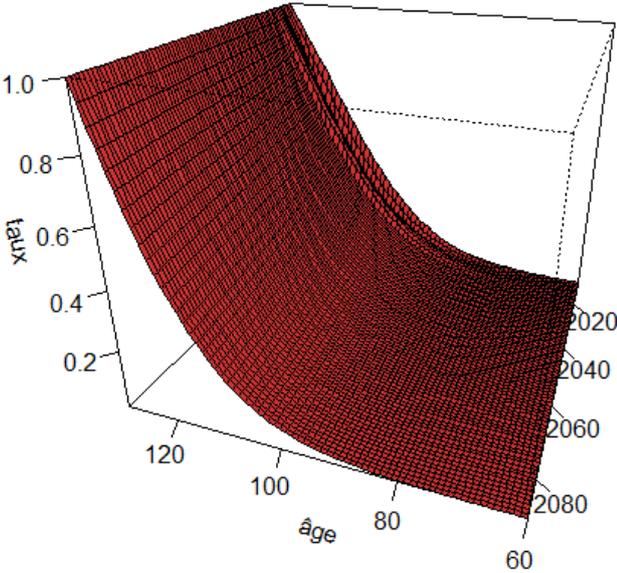


Figure IV-26 : $\tilde{q}_x(t)$ estimés et projetés par Cairns-Blake-Dowd avec une extrapolation Coale-Kisker pour les hommes

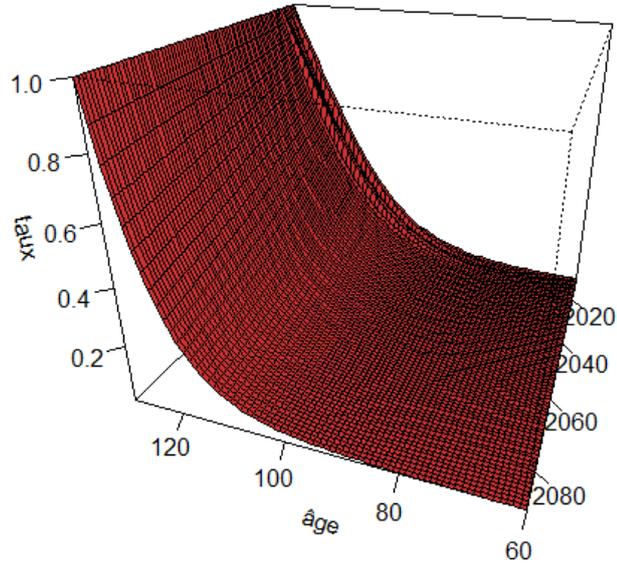


Figure IV-27 : $\tilde{q}_x(t)$ estimés et projetés par Cairns-Blake-Dowd avec une extrapolation Coale-Kisker pour les femmes

D. Méthodes de positionnement à partir d'une table existante

Il a été démontré que lorsque la taille du portefeuille d'observation s'avère être trop faible, les approches par positionnement sont plus fiables. Les méthodes de positionnement sont faites à partir de deux éléments :

- Le portefeuille client et les observations d'entrées/sorties/décès qui y sont liées ;
- Une table de mortalité importée qui servira de référence dans cette partie de l'étude. C'est la table de mortalité générationnelle femme TGF05 qui sera utilisée sur les femmes et la TGH05 pour les hommes.

$q_x^{réf}(t)$ sera à partir de maintenant la notation utilisée pour définir le taux de mortalité d'un individu d'âge x sur l'année t affilié à notre table de référence.

La tranche d'âge qui suscite notre intérêt est l'ensemble des âges supérieurs à 60 ans. On va donc modéliser notre table prospective sur la tranche d'âges $x = [60, 130]$. Nous souhaitons donc effectuer une table de mortalité prospective allant du début de l'observation jusqu'en 2089. On note donc $t = [2010, 2089]$ l'ensemble des années calendaires de la table de mortalité prospective.

Cette partie a été réalisée en partie à l'aide du package R **ELT** et des méthodologies existantes (Planchet et Tomas).

1. Positionnement par *Standardized Mortality Ratio* (SMR)

a) *Théorie et application*

La première méthode utilisée est assez simple. Elle consiste à appliquer à la table de référence un coefficient de majoration ou de réduction. Ce coefficient est quant à lui calculé à partir des observations de notre jeu de données.

Les probabilités de décès du portefeuille \tilde{q}_x sont obtenues à partir de la formule suivante :

$$\tilde{q}_x(t) = SMR * q_x^{réf}(t) \text{ avec } SMR = \frac{\sum_{(x^*, t^*)} D_{x^*}(t^*)}{\sum_{(x^*, t^*)} E_{x^*, t^*} * q_{x^*}^{réf}(t^*)}$$

x^* et t^* sont des variables importantes puisqu'elles représentent respectivement la plage d'âge commune entre le portefeuille et la table de référence et la période d'observation commune. Ces deux paramètres sont à sélectionner assez sensiblement (en particulier x^*) puisque le SMR peut fluctuer fortement en fonction de ces paramètres.

- La période commune t^* se restreint donc aux années 2010 à 2019 : $t^* = [2010, 2019]$
- Concernant la tranche d'âge commune x^* , la tranche la plus large va de 60 à 100 ans, en adéquation avec les résultats de validité du critère de Cochran. Outre l'importance du choix

de x^* , il faut également prendre en considération le caractère volatile des taux de décès au niveau des grands âges. Il peut donc être intéressant de restreindre la taille de la tranche commune pour plus de précision.

Sexe	[60,85]	[60,90]	[60,95]	[60,100]
Homme	71,04%	69,76%	68,83%	68,12%
Femme	76,19%	74,46%	73,84%	72,91%

Tableau 10 : Valeurs des coefficients SMR pour différentes tranches d'âge et sexes

La dépendance du SMR à la tranche d'âge commune est d'autant plus visible en avec les résultats obtenus. Concernant le choix du x^* , $x^* = [60, 100]$ sera retenu pour la suite de l'étude. Le manque de précision aux grands âges sera pallié par une méthode de fermeture de table.

Une fois le modèle estimé, il est important qu'il satisfasse les critères de validation.

b) Validation

i. Validation par critère de proximité :

Cette partie consiste à contrôler l'écart entre les résultats obtenus et les observations faites en amont. Concernant les tests, le risque retenu s'élève à 5%.

Parmi ce groupe de critères, les tests et critères donnent les p-valeurs (p-v) suivantes :

Population	p-v du test de vraisemblance	p-v du test SMR	p-v du test de Wilcoxon
Masculine	0,00	0,499	0,00
Féminine	0,005	0,498	0,013

Tableau 11 : Tests de validation par critère de proximité, méthode 1 sur la tranche [60, 100]

- Le test du rapport de vraisemblance est appliqué en opposant les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 : \text{l'ajustement effectué correspond à la loi de mortalité sous-jacente} \\ H_1 : \text{l'ajustement ne correspond pas à la loi sous-jacente} \end{array} \right.$$

En prenant pour hypothèse que les décès suivent une loi de Poisson de paramètre $E_{x,t}q_{x,t}$, la statistique de test du rapport de vraisemblance s'écrit :

$$\varepsilon^{VR} = \sum_{x,t} (D_x(t) \ln \left(\frac{D_x(t)}{E_x(t) \tilde{q}_x(t)} \right) - D_x(t) - E_x(t) \tilde{q}_x(t))$$

Si $\varepsilon^{VR} > X^2_{1-\alpha}(n)$, le quantile $1 - \alpha$ d'une distribution X^2 à n degrés de liberté : on rejette H_0

La p-valeur de ce test étant inférieur à 0,05, l'hypothèse nulle est rejetée pour les hommes et pour les femmes. Le rejet de H_0 avec un risque de 5% est une mauvaise nouvelle puisqu'il souligne un ajustement biaisé comparé à la loi initiale.

- Le test du SMR est quant à lui basé sur le rapport entre nombre de décès observés et nombre de décès qui font suite à l'ajustement. Le test confronte les hypothèses suivantes :

$$\left[\begin{array}{l} H_0 : \text{le SMR n'est pas significativement différent de 1} \\ H_1 : \text{le SMR est significativement différent de 1} \end{array} \right.$$

Le test donne pour affirmation que l'ajustement a bien été réalisé si l'hypothèse H_0 n'est pas rejetée. Nous souhaitons que le SMR ne soit pas significativement différent de 1 puisque si le SMR est inférieur à 1, les taux de décès sont surestimés et si au contraire il est supérieur à 1, une sous-estimation a été faite sur les taux de décès.

La statistique de ce test se calcule de la façon suivante :

$$\left[\begin{array}{l} \varepsilon^{SMR} = 3 \times D^{\frac{1}{2}} \left(1 - 1/9D - (D/E)^{\frac{1}{3}} \right) \text{ si } SMR > 1 \\ \varepsilon^{SMR} = 3 \times D'^{1/2} \left(-1 + 1/9D' + (D'/E)^{1/3} \right) \text{ si } SMR < 1 \end{array} \right.$$

Avec D , la somme totale des décès ; D' la somme des décès plus 1 ; et $E = \sum_{x,t} -E_x(t) \tilde{q}_x(t)$.

La statistique du test ε^{SMR} , doit suivre une loi normale centrée réduite pour accepter H_0 . Dans le cas où la statistique de test dépasse un quantile à 95% de la loi normale, H_0 est rejeté.

$\leftrightarrow \varepsilon^{SMR} \sim N(0,1) : H_0$ est acceptée

$\varepsilon^{SMR} > N_{1-\alpha}(0,1) : H_0$ est rejetée

Avec des p-valeurs bien supérieurs à 0,05, l'hypothèse H_0 n'est pas rejetée. C'est un critère positif pour la validation de notre ajustement car nous ne sommes ni dans un cas de sous-estimation ni de surestimation.

- Le test de Wilcoxon regroupe les taux de décès observés et les taux de décès ajustés par paire. La différence de chaque paire est ensuite calculée puis ordonnée par ordre croissant (en valeur absolue), ce qui permet de comparer les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 : \text{la médiane de la différence entre chaque paire est nulle} \\ H_1 : \text{la médiane de la différence entre chaque paire n'est pas nulle} \end{array} \right.$$

Avec la statistique de test :

$$\varepsilon^{WIL} = \frac{(w - 0.5 - n(n + 1))/4}{\sqrt{\frac{n(n + 1)(2n + 1)}{24}}}$$

La validation est hypothèses se fait de la façon suivante :

$\leftrightarrow \varepsilon^{WIL} \sim N(0,1) : H_0$ est acceptée

$|\varepsilon^{WIL}| > N_{1-\alpha/2}(0,1) : H_0$ est rejetée

Avec des p-valeurs assez faibles, on rejette l'hypothèse des différences nulles.

Un seul test sur les trois montre que notre ajustement est bon. Cette observation peut être creusée à partir de quantités qui se présentent comme des critères de validation du modèle :

Population	Résidus standardisés > 2	Déviance	χ^2	MAPE
Masculine	34	1203	611	33
Féminine	27	975	492	27

Tableau 12 : Quantités du critère de proximité pour la méthode 1 sur [60,100]

Ces quantités donnant une indication sur le bon ajustement du modèle. Le nombre de résidus standardisés supérieurs à 2 permet de se rendre compte de la qualité de l'ajustement par les résidus. Nous souhaitons bien entendu que cette quantité soit la plus faible possible. 34 et 27 sont des quantités un peu trop élevées au vu de la tranche d'âge sur laquelle l'étude se fait.

Globalement, les résultats obtenus ne permettent pas de dire que l'ajustement soit très fiable par rapport aux données d'observation. Il conviendra, suite à l'analyse graphique, de se placer sur un intervalle d'âge plus petit afin d'avoir plus de précision.

Par ailleurs, une observation intéressante qui découle de ces résultats est que la population féminine offre globalement une meilleure prédiction aux vues des critères de proximité de

l'ajustement. C'est en contradiction avec le fait que le portefeuille client n'est composé que de 40,00% de femmes. La volatilité des taux observés étant donc supposée être plus faible, l'exactitude de l'ajustement n'était pas supposée être meilleur que celui de la population masculine.

En comparant pour une année les taux estimés aux taux bruts, un écart est visible en particulier aux grands âges ; comme l'atteste le graphique ci-dessous représentant les \tilde{q}_x de 2018 pour la population féminine.

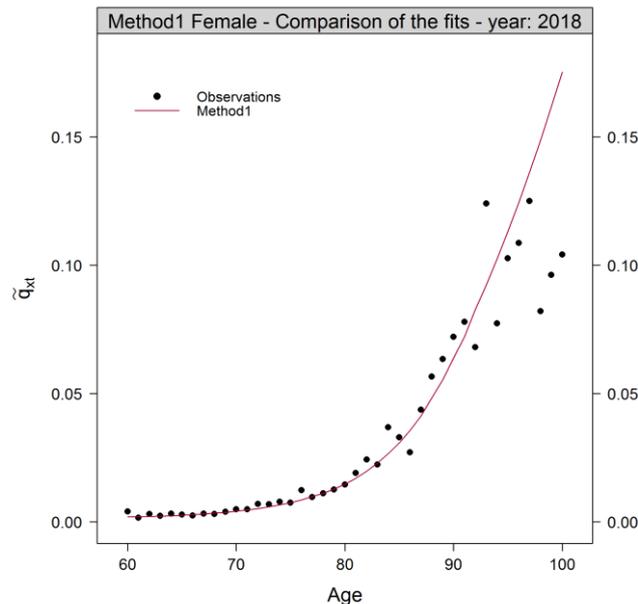


Figure IV-28 : Taux de décès bruts (Hoem) et calculée avec la méthode SMR pour les femmes en 2018

La proximité entre les taux observés et calculés à partir du SMR ne semble pas être très élevée sur les premiers âges. En avançant dans les âges, l'écart semble être de plus en plus élevé, avec un début de dérivation à 85 ans.

Sur cette même année, nous pouvons également observer les résidus de la réponse, de Pearson et de la déviance pour la population féminine.

Method1 Female - Plot of the residuals - year: 2018

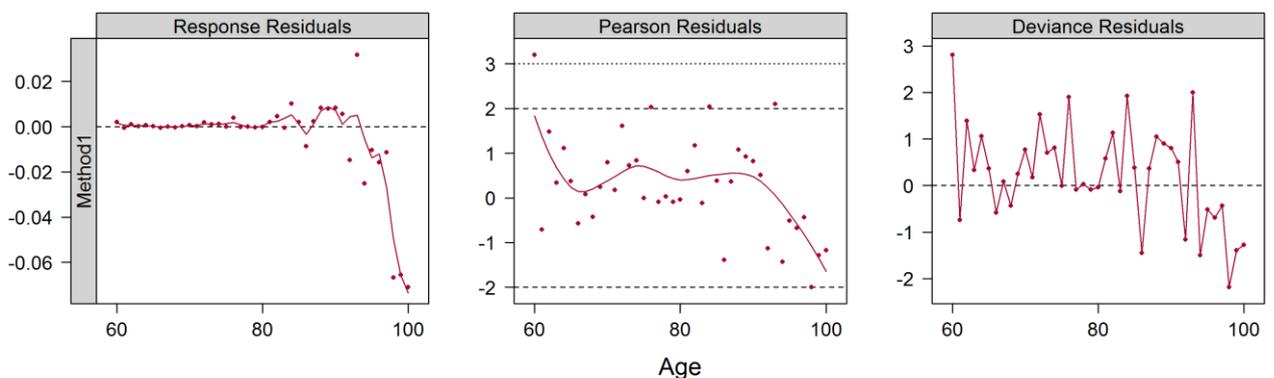


Figure IV-29 : Résidus issus de cette première méthode pour les femmes de 2018

Les résidus de la réponse augmentent fortement au fil des âges, atteignant presque -0.06 à 100 ans. Concernant les résidus de Pearson et de la déviance, la distribution est plus équilibrée. Par ailleurs, les fortes chutes de résidus de déviance après un certain âge mettent en avant le sur-lissage des taux. 85 ans peut être considéré comme l'âge à partir duquel la justesse du modèle comparé aux observations chute.

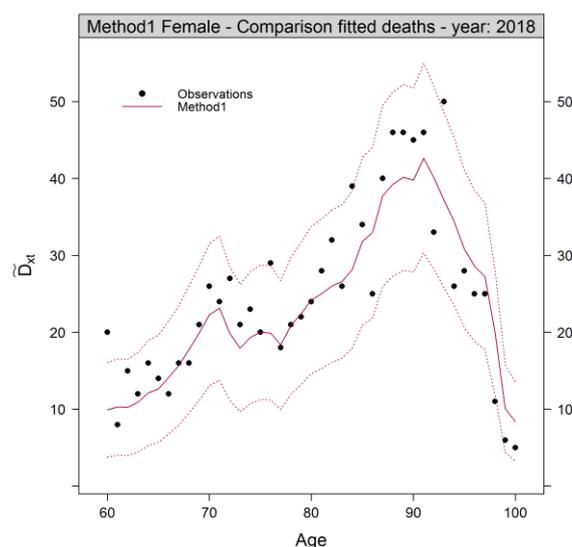


Figure IV-30 : Courbe des décès estimés par la méthode SMR et décès bruts sur la population féminine de 2018

Ce graphique permet de mettre en comparaison la courbe des décès estimés à partir des taux ajustés et les décès observés en 2018 pour les femmes. Les deux courbes qui englobent les résultats correspondent à l'intervalle de confiance à 95%. Les observations semblent globalement être en accord sur les premiers âges. Peu de points d'observations sont à une forte distance de la courbe des estimations. Néanmoins, à partir de 85 ans, le nombre de décès semble être sous-estimé puis surestimé dès 90 ans.

L'analyse graphique semble être en phase avec le résultat des tests ; nous laissant penser qu'il faudrait restreindre la tranche d'âge sur laquelle l'ajustement est fait. C'est pour cette raison que l'étude a été restreinte à 85 ans. Les âges suivants seront tout de même estimés à partir d'une méthode de fermeture de table.

En réitérant l'ajustement sur la nouvelle tranche d'âge [60, 85], les résultats sont plus satisfaisants. Les résultats sont visibles dans l'annexe.

ii. Validation par critères de régularité de l'ajustement :

L'ajustement que nous allons tester sur cette seconde partie a été fait sur la nouvelle tranche d'âge allant de 60 à 85 ans.

Voici les résultats obtenus aux critères de régularité de l'ajustement :

Population	p-v du test des runs	p-v du test des signes
Masculine	0,80	0,55
Féminine	0,20	0,24

Tableau 13 : Récapitulatif des p-valeurs des tests de régularité d'ajustement

Le second groupe de critères de validation se traduit en 2 tests :

- Le test des *runs* est un test non paramétrique. Ce test se base sur des *runs* ; un *run* est défini comme étant le segment d'une séquence construit avec des éléments avoisinants qui sont égaux. Il a pour objectif de déterminer l'indépendance des éléments d'une même séquence.

Les hypothèses sont les suivantes :

- H_0 : sur une séquence de n éléments, le nombre de *runs* de cette séquence est une variable aléatoire dont la distribution conditionnelle connaissant les nombres n_+, n_- (avec $n = n_+ + n_-$) est approximativement normale de paramètres μ, σ .
- H_1 : le nombre de *runs* n'est pas une variable aléatoire de distribution conditionnellement normale.

Avec comme statistique de test :

$$S^{RUN} = \frac{H - \mu}{\sigma}$$

- $\mu = \frac{2n_+n_-}{n_+ + n_-}$: le paramètre d'espérance de l'approximation normale
- $\sigma^2 = \frac{2n_+n_-(2n_+n_-(n_+ + n_-))}{(n_+ + n_-)^2(n_+ + n_- - 1)}$: la paramètre de variance de l'approximation normale
- H : le nombre de *runs* dans la séquence

H_0 est rejeté si $|S^{RUN}| > Q_{N(0,1)}(1 - \frac{\alpha}{2})$; donc si la valeur absolue de la statistique de test dépasse le quantile $1 - \frac{\alpha}{2}$ (0,975) d'une loi normale centrée réduite.

Le résultat obtenu ne rejette pas H_0 pour les deux populations, c'est un critère satisfaisant permettant de dire que l'ajustement est assez régulier.

- Le test des signes, de la même façon que le test de Wilcoxon, regroupe par paire taux décès observés et ajustés. Il analyse ensuite la fréquence des changements de signes des différences entre ces paires. La médiane est ensuite relevée puis confrontée aux hypothèses :

- H_0 : la médiane entre les signes des différences des paires est nulle

H_1 : la médiane entre les signes n'est pas nulle

De même que pour le test des *runs*, $n = n_+ + n_-$ avec n_+ le nombre de signes positifs et n_- le nombre de signes négatifs. La statistique de test est alors :

$$S^{SG} = \frac{|n_+ + n_-| - 1}{\sqrt{n}}$$

L'hypothèse nulle est rejetée à condition que $|S^{SG}| > Q_{N(0,1)}(1 - \frac{\alpha}{2})$; sous la même condition que le test des *runs*.

H_0 étant également rejeté au risque de 5% pour les deux sexes, les deux tests de ce groupe de critères de validation sont satisfaisants. La régularité de l'ajustement pour les deux sexes ne semble donc pas à revoir.

c) Fermeture de table

Les $\tilde{q}_x(t)$ ont finalement été estimés sur une tranche d'âge allant de 60 à 85 ans pour 2010 à 2089. Puisque le souhait est de construire une table de mortalité prospective allant jusqu'à 130 ans, des méthodes de fermeture de tables ont été appliquées. De la même façon que pour la méthode de Cairns-Blake-Dowd, 4 méthodes de fermeture de table ont été retenues : Exponentielle, Coale-Kisker, Kannisto et Denuit & Goderniaux.

Une problématique ressort de la fermeture par Coale-Kisker. En effet, les $\tilde{q}_x(t)$ ont été estimés à partir des $q_x^{réf}$ allant jusqu'à 2005. De ce fait, sur les dernières années et pour les premiers âges, aucune observation n'existait sur la table de référence. Des 0 apparaissent donc à partir de $\tilde{q}_{60}(2066)$ et se propagent de manière triangulaire, prenant en 2067 les $x = 60 ; 61$ et ainsi de suite. La méthode de fermeture par Coale-Kisker se base sur des taux estimés, pour chaque année, qui remontent à 65 ans.

Il est impossible d'utiliser cette méthode à partir de 2071 car $\tilde{q}_{65}(2071) = 0$. La méthode a tout de même été appliquée mais ne sera pas finalement retenue.

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
X^2	187,42	300,99	284,43	380,38
MAPE	22,52	28,42	27,73	30,18
Déviance	2422	2232	2203	2461

Tableau 14 : Quantités des différentes fermetures de table sur le modèle SMR pour les hommes

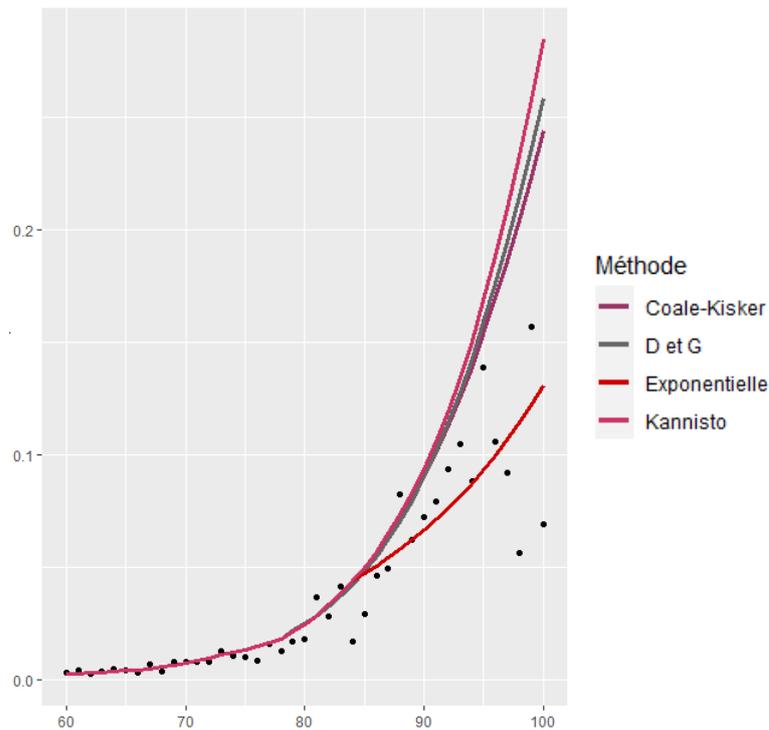


Figure IV-31 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014

L'extrapolation par la méthode exponentielle est assez clairement la plus adaptée à fermer la table pour la population masculine. Outre le fait qu'elle répond le mieux aux distances entre estimation et observation, le graphique des taux de 2014 montre que c'est la seule méthode qui ne surestime pas les probabilités de décès aux grands âges. De plus, la courbe colle assez bien aux observations et répond au principe de prudence. Puisque cette table sera utilisée sur les rentes viagères, il est indispensable que les probabilités de décès ne soient pas surestimées.

Les résultats des différentes fermetures de tables pour la population féminine se trouvent dans l'annexe. La méthode d'extrapolation retenue cette fois-ci est celle de Denuit & Goderniaux. En effet, malgré que Coale-Kisker fournit les meilleurs résultats, son extrapolation ne va que jusqu'à 2070 et c'est ce qui fait que Denuit & Goderniaux est préféré.

Voici le tracé final des \tilde{q}_x après fermeture sur cette première méthode de positionnement :

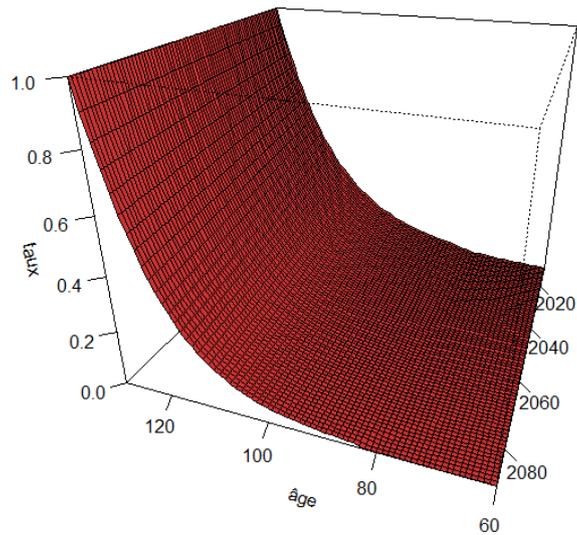


Figure IV-32 : Taux de décès estimés avec la méthode SMR et une fermeture de table exponentielle pour les hommes

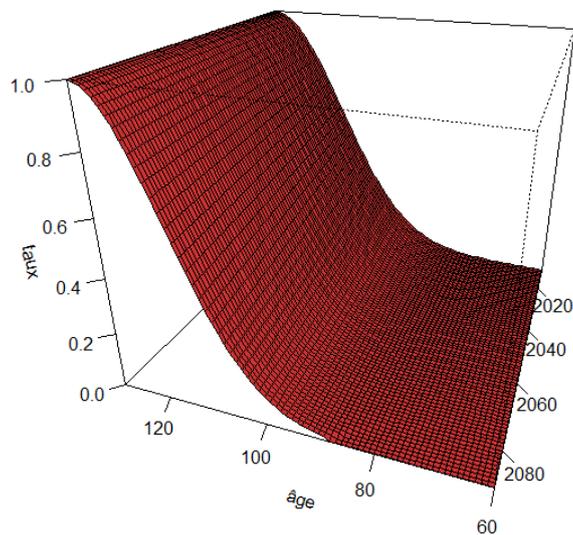


Figure IV-33 : Taux de décès estimés avec la méthode SMR et une fermeture de table Denuit & Goderniaux pour les femmes

La forme de la courbe n'est pas du tout la même. La fermeture par extrapolation exponentielle offre une courbe avec une forme exponentielle assez claire à l'instar de Denuit & Goderniaux qui se voit plus arrondie aux grands âges.

2. Modèle relationnel à deux paramètres

En 1971, Brass revisite le modèle de Cox en remettant en question l'aspect linéaire en fonction de l'âge de l'unique paramètre de son modèle. Brass réinterprète donc ce modèle en y ajoutant un second paramètre visant à répercuter la tendance dans le temps de la population de référence sur la population modélisée :

$$\text{logit}(q_x(t)) = \theta_1 + \theta_2 * \text{logit}(q_x^{\text{réf}}(t))$$

La fonction *logit* est utilisée car elle offre la possibilité de se placer sur \mathbb{R} et donc de ne pas ressortir comme résultat une quantité contrainte sur $[0, 1]$. $\text{logit}(b) = \ln\left(\frac{b}{1-b}\right)$

Le modèle utilisé dans cette seconde partie est une variante du modèle semi-paramétrique de Brass. Il consiste, en quelque sorte, à effectuer une régression linéaire sur les taux bruts :

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

Deux paramètres de régression sont à estimer (α, β) ; le premier est lié à la mortalité tandis que le second est un paramètre qui varie en fonction de l'âge. Les résidus du modèle $\varepsilon_{x^*}(t^*)$ sont supposés iid.

(α, β) sont estimés en minimisant la distance entre les probabilités de décès observées à l'aide de l'estimateur d'Hoem $\hat{q}_x(t)$ et celles estimées $\tilde{q}_x(t)$ à laquelle est pondérée l'exposition au risque :

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\text{argmin}} \left| \sum E_x(t) (\hat{q}_{x^*}(t^*) - \tilde{q}_{x^*}(t^*)) \right|$$

En utilisant la méthode des moindres carrés pour estimer ces éléments les résultats sont les suivants :

Sexe	$\hat{\alpha}$	$\hat{\beta}$
Homme	-0,8496	0,8785
Femme	-0,5552	0,9400

Tableau 15 : Coefficients de la méthode 2 pour la tranche d'âge [60, 100]

Une fois les paramètres estimés, nous pouvons calculer les taux de décès à partir de la table de référence pour $x = [60, 100]$ et $t = [2010, 2019]$:

$$\tilde{q}_x(t) = \frac{\exp(\hat{\alpha} + \hat{\beta} \text{logit}(q_x^{\text{réf}}(t)))}{1 + \exp(\hat{\alpha} + \hat{\beta} \text{logit}(q_x^{\text{réf}}(t)))}$$

Après exécution des critères de validation par proximité pour [60, 100] ans (voir annexe) ; les observations nous redirigent vers une nouvelle tranche d'âge. Cette fois-ci le modèle sera ajusté sur la tranche d'âge allant de 60 à 90 ans. Les nouveaux paramètres du modèle sont les suivant :

Sexe	$\hat{\alpha}$	$\hat{\beta}$
Homme	- 0,8485	0,8790
Femme	- 0,5066	+ 0,9527

Tableau 16 : Coefficients de la méthode 2 sur la tranche [60, 90] ans

Les indicateurs d'ajustement donnent de meilleurs résultats que pour la tranche d'âge précédente comme l'atteste le tableau ci-dessous.

Population	Résidus standardisés > 2	Déviance	X^2	MAPE	R^2
Masculine	22	937	400	21,18	0,88
Féminine	16	831	334	25,14	0,92

Tableau 17 : Quantités des critères d'ajustement pour la méthode 2 sur la tranche [60, 90]

Le R^2 des deux sexes s'approche de 90%, ce qui voudrait dire que le modèle explique à 90% les observations ; un résultat satisfaisant. Les déviances et distances du X^2 ont également baissés suite au réajustement sur une nouvelle tranche d'âge.

Les tests du critère de régularité donnent également de bons résultats (c.f. annexe). Il convient maintenant de comparer les modèles de fermeture de tables pour cette méthode et pour des âges allant de 90 à 130 ans. Le modèle de fermeture retenu suite à cette comparaison est celui de Kannisto pour les hommes et celui de Denuit & Goderniaux pour les femmes.

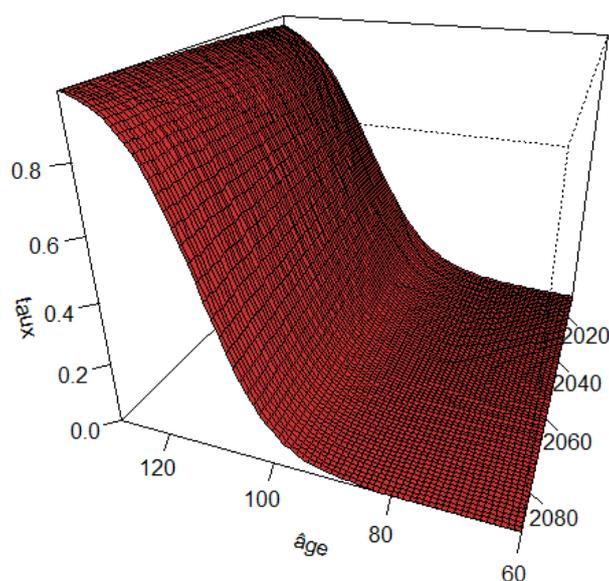


Figure IV-34 : Taux de décès estimés à partir de la méthode 2 et d'une fermeture de table de Kannisto pour les hommes

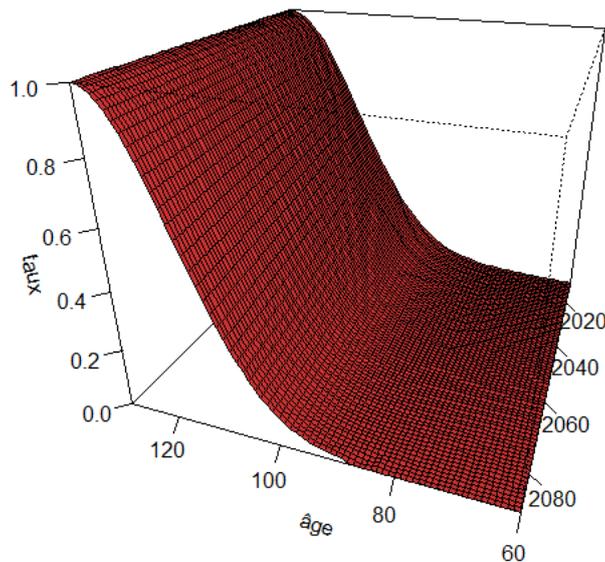


Figure IV-35: Taux de décès estimés à partir de la méthode 2 et d'une fermeture de table de Denuit & Goderniaux pour les femmes

La méthode de Kannisto ne fournit pas des taux de décès à 100% pour les individus atteignant 130 ans. C'est une condition qui ne figure pas dans la méthode. Néanmoins, les fermetures semblent être de la même forme.

3. Modèle linéaire généralisé de Poisson

Le modèle linéaire généralisé (GLM) de Poisson utilisé consiste à faire suivre aux décès du portefeuille une loi de Poisson ayant pour paramètre le produit entre l'exposition au risque et le taux de risque instantané. Cette méthode est utilisée comme un GLM de Poisson avec pour variable explicative la référence de mortalité. x^* et t^* sont également les plages d'âges et périodes d'observations communes à notre table de référence.

Soit :

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

Le taux de risque instantané est défini avec la formule suivante :

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

La formule utilisée met en avant une interaction entre âge et période d'observation. Cette interaction est d'autant plus importante du fait que sans sa présence, les forces de mortalité auront des mouvements parallèles pour chacune des dimensions.

Par la relation entre μ et les taux de mortalité :

$$\tilde{\mu}_{x^*}(t) = -\log(1 - \tilde{q}_x(t))$$

$$\Leftrightarrow \tilde{q}_x(t) = 1 - \exp[-(\beta_0 + \beta_1 \log(q_x^{ref}(t)) + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^*)]$$

Après application de la méthode et calcul des différents paramètres voici les résultats obtenus pour chacun des estimateurs :

Sexe	Population masculine				Population féminine				
	Paramètre	Estimation	Erreur Std	z value	p value	Estimation	Erreur Std	z value	p value
	$\widehat{\beta}_0$	-78,59	62,72	-1,22	0,22	85,99	66,01	-1,27	0,21
	$\widehat{\beta}_1$	1,18	0,18	5,87	0,00	1,16	0,18	10,75	<2E-16
	$\widehat{\beta}_2$	0,81	0,78	1,00	0,32	0,69	0,82	0,81	0,42
	$\widehat{\beta}_3$	0,05	0,03	1,25	0,21	0,04	0,03	1,30	0,19
	$\widehat{\beta}_4$	0,00	0,00	-1,04	0,30	0,00	0,00	-0,84	0,40

Tableau 18 : Coefficients obtenus pour la 3^{ème} méthode sur la tranche d'âge [60, 100] ans.

Les critères de proximité entre les estimations de cette méthode régressive sur [60,100] et les observations poussent à se restreindre sur la tranche d'âge d'observations allant de 60 à 90 ans.

Malgré des résultats satisfaisants sur la première tranche, le réajustement donne de meilleurs résultats notamment avec les quantités suivantes :

Population	Résidus standardisés > 2	Déviance	χ^2	MAPE	R ²
Masculine	21	781	394	21,41	0,88
Féminine	19	639	321	24,71	0,93

Tableau 19 : Quantités d'ajustement de la méthode 3 sur [60, 90]

Voici également les nouveaux coefficients du modèle régressif :

Population	Masculine	Féminine
$\widehat{\beta}_0$	-33,40	-33,70
$\widehat{\beta}_1$	0,81	1,07
$\widehat{\beta}_2$	0,39	0,01
$\widehat{\beta}_3$	0,02	0,02
$\widehat{\beta}_4$	-1,90E-04	-1,40E-05

Tableau 20 : Coefficients de la méthode 3 sur [60, 90]

Ce modèle prédit est donc plutôt bien ajusté aux observations de notre portefeuille, avec des quantités assez bonnes (déviance plutôt faible, R^2 proche de 90%). A noter que les résultats sont meilleurs une fois de plus sur la population féminine. Ce qui veut dire que les observations de femmes du portefeuille sont bien ajustées par la méthode de lissage.

Les critères de validation de groupe 2 sont également validés pour cette méthode. Il faut maintenant compléter cette méthode régressive avec une fermeture de table sur la tranche allant de 90 à 130 ans. Suite aux comparaisons, la méthode de fermeture de table choisie est l'extrapolation exponentielle pour les hommes et Denuit & Goderniaux pour les femmes.

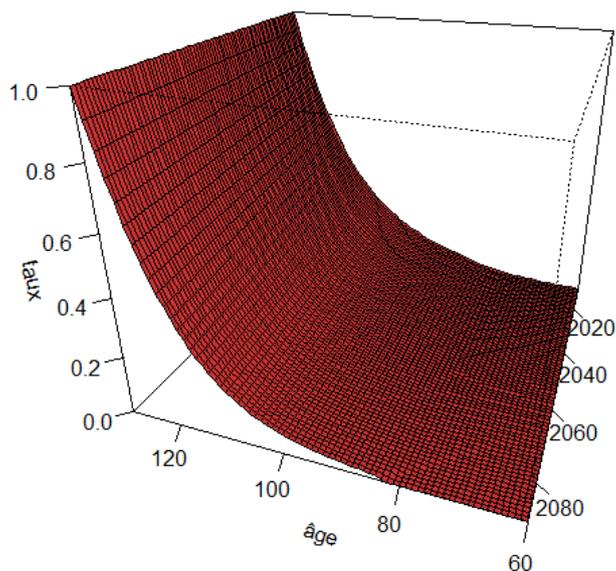


Figure IV-36 : Taux de décès estimés à partir de la méthode 3 avec une extrapolation exponentielle pour les hommes

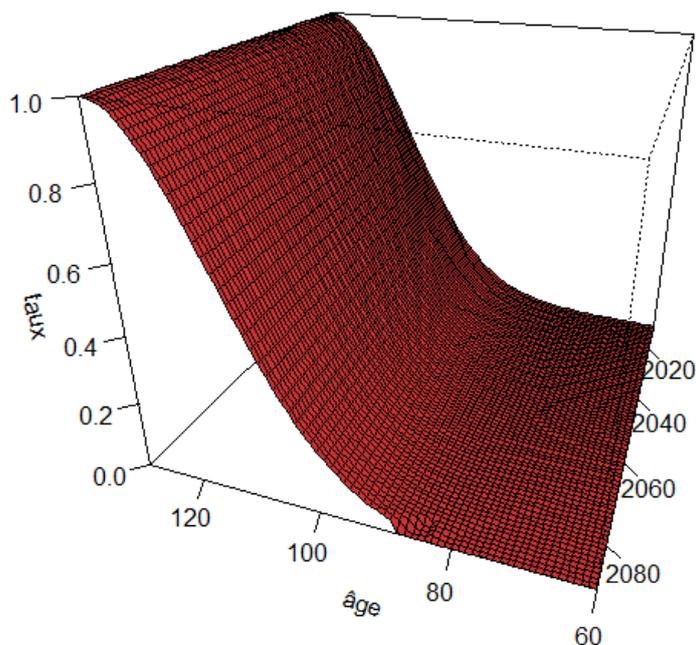


Figure IV-37 : Taux de décès modélisés par la méthode 3 pour les femmes avec une fermeture de table Denuit & Goderniaux

4. Lissage non-paramétrique et dérive de mortalité

Il existe une quatrième méthode de positionnement. Cette méthode se divise en 2 étapes :

- Les taux bruts calculés avec l'estimateur d'Hoem du moment sont le point de départ de la méthode. La première étape consiste à lisser ces taux bruts de façon non paramétrique.
- S'ensuit l'application de dérivées calculées à partir de la table de référence. Cette dérivée de mortalité s'applique directement sur les taux lissés et n'impose donc pas de relation directe entre les taux bruts et les taux de référence.

L'utilisation du lissage non-paramétrique au cours de la première étape constitue un avantage lors de la conception de cette méthode. Cette méthode est fortement sensible et requiert la validation d'un expert. Le modèle sera tout de même appliqué afin de se rendre compte des résultats attendus.

Le modèle non-paramétrique utilisé sera le suivant :

$$D_x \sim P \left(E_x q_x^{réf} \exp(f(x)) \right)$$

Avec :

- $E_x q_x^{réf}$ le nombre de décès espérés à partir de la table de référence
- f est une fonction lisse qui sera appliquée aux taux bruts.

La première étape de la modélisation est d'estimer f , la fonction lisse qui sera appliquée sur la table du moment. Une méthode de vraisemblance locale est utilisée pour estimer cette fonction. f s'estime en fonction de paramètres de lissage (fenêtre d'observation et le degré du polynôme). Il faut donc faire attention aux irrégularités et avoir un jugement fiable pour cette étape.

Afin de mettre en avant ces paramètres, la stratégie est d'évaluer le nombre d'ajustements candidats et de sélectionner le plus adapté à partir d'un critère de sélection : l'AIC. L'AIC, ou critère d'information d'Akaike est un critère visant à pénaliser les modèles en fonction du nombre de paramètres. C'est, en quelque sorte, une représentation du compromis entre l'augmentation du modèle faisant suite à l'ajout d'un paramètre et la satisfaction du critère de parcimonie. Le modèle choisi est celui avec le critère d'information d'Akaike le plus faible.

Les graphiques ci-dessous permettent de se rendre compte de l'AIC obtenu en fonction des degrés du polynôme et du degré de liberté utilisé. Le degré obtenant le score le plus faible sera finalement préféré :

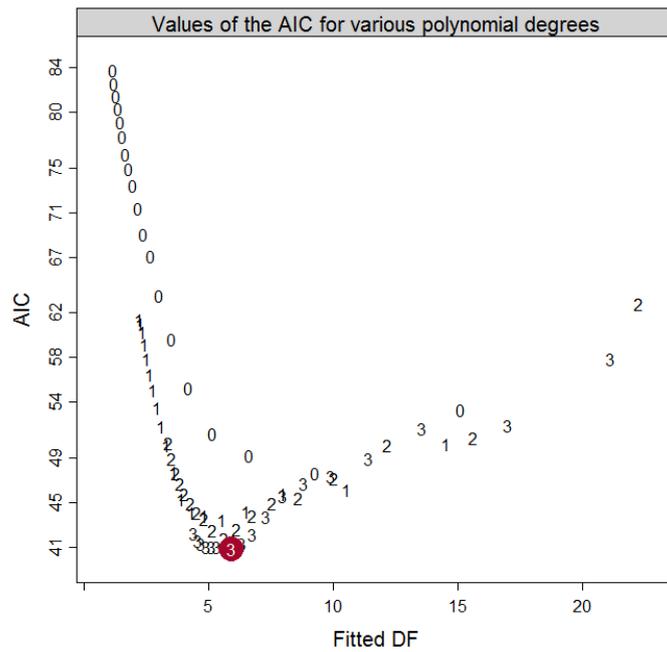


Figure IV-38 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population féminine

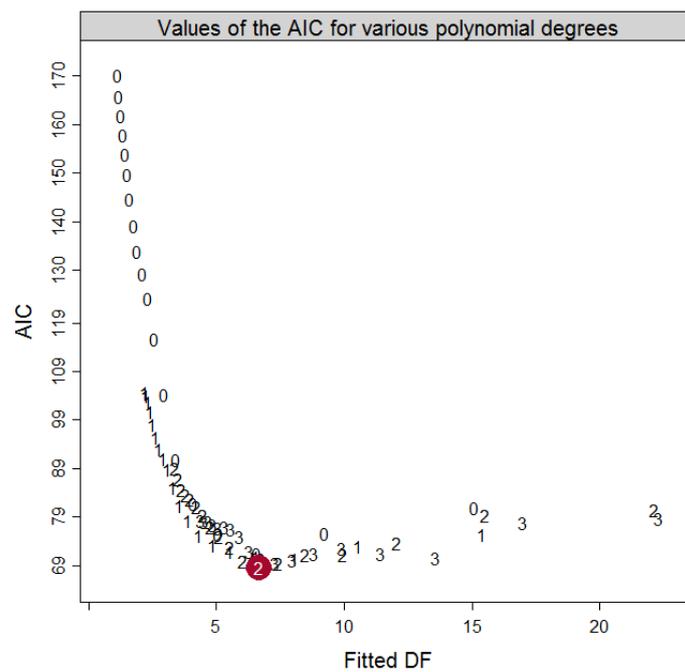


Figure IV-39 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population masculine

La documentation liée à ce modèle laisse penser qu'il faut choisir le paramètre lorsque l'AIC atteint un minimum après une forte descente. Dans ce cas, le degrés 2 serait logique pour les individus de sexe masculin et le degrés 3 serait préférable pour les femmes.

Nous sélectionnons donc pour les hommes un ajustement quadratique $u_2 = 6,660$ et $v_3 = 5,935$ avec pour AIC respectifs 68,572 et 40,873.

Pour résumer l'ensemble des paramètres des fonctions de lissage des taux bruts, voici un tableau récapitulatif :

Sexe	Degrés du polynôme	Degrés de liberté	AIC
Hommes	2	6,660	68,572
Femmes	3	5,935	40,873

Tableau 21 : Paramètres de fonctions de lissage des taux bruts pour la méthode 4 [60,100] ans

Une fois les taux bruts ajustés, les taux de décès se calculent avec la formule suivante :

$$\tilde{q}_x(t) = q_x^{réf}(t) \exp(\hat{f}(t))$$

Cette méthode laisse une grande place à l'interprétation ; notamment à travers le choix des sous-paramètres du modèle. C'est pourquoi un recours aux critères de validation par proximité est important (résultats disponibles dans l'annexe). Les résultats sont plutôt bons au vue de la complexité du modèle. Par ailleurs, cet aspect concave à partir de 90 ans souligne une mauvaise estimation aux grands âges liée à la volatilité de nos observations. La nouvelle tranche d'âge est donc [60, 90] pour cette méthode.

Le nouveau modèle est donc réalisé avec de nouvelles fonctions de lissages respectant les critères suivants :

Sexe	Degrés du polynôme	Degrés de liberté	AIC
Hommes	0	4,58	37,657
Femmes	1	2,091	29,439

Tableau 22 : Paramètres de fonctions de lissage des taux bruts pour la méthode 4 [60,90] ans

Les courbes retenant les valeurs de l'AIC pour différents degrés de polynômes et de liberté se trouvent en annexe. Pour les hommes, les taux seront lissés avec un polynôme de degrés 0 (donc une droite) et pour les femmes une fonction affine.

Les résultats obtenus avec la nouvelle tranche d'âge et les nouveaux paramètres de lissage sont bons, avec tous les critères de proximité validés et les quantités ci-dessous :

Population	Résidus standardisés > 2	Déviance	χ^2	MAPE	R ²
Masculine	19	785	375	21,23	0,90
Féminine	17	647	327	25,73	0,92

Tableau 23 : Critères d'ajustement de la méthode 4 sur la nouvelle tranche d'âge [60, 90] ans

Les critères de régularité étant validés, des modèles de fermetures sont également appliqués.

Suite aux comparaisons de fermetures, la fermeture retenue pour ce modèle est l'extrapolation exponentielle pour les deux populations.

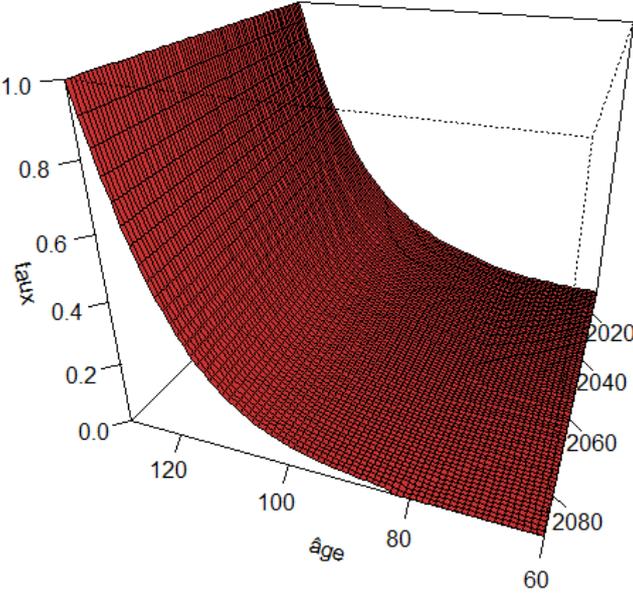


Figure IV-40 : Taux modélisés par la 4^{ème} méthode de positionnement avec une extrapolation exponentielle, population masculine

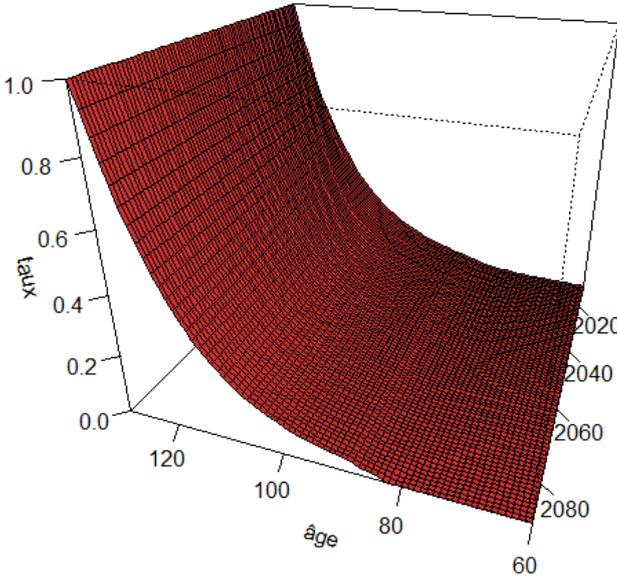


Figure IV-41 : Taux modélisés par la 4^{ème} méthode de positionnement avec une extrapolation exponentielle, population féminine

E. Estimation par régression logistique

1. Théorie

La régression logistique binomiale est un cas particulier de Modèle Linéaire Généralisé (GLM). Tandis qu'en considérant Y la variable à expliquer et X le vecteur composé des variables explicatives, un modèle de régression linéaire classique serait de la forme :

$$Y = X^T \beta$$

La régression logistique binomiale est un cas particulier de la régression binomiale. Puisque dans le cas d'une régression binomiale, la distribution de Y est binaire, les relations diffèrent. Le modèle linéaire se « généralise » donc à partir d'une fonction de lien g et avec la formule suivante :

$$g(E(Y)) = X^T \beta$$

En l'occurrence, la régression logistique binomiale utilise la fonction de lien correspondante à la fonction logit (où $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$). La fonction logit permet de transformer une valeur comprise entre 0 et 1 en une valeur comprise entre $-\infty$ et $+\infty$. Il existe d'autres modèles de régression binomiaux mais la régression logistique est souvent préférée pour des raisons pratiques.

La fonction de lien g est alors $g(p) = \text{logit}(p) = \log\left(\frac{p}{1-p}\right)$ et $g^{-1}(p) = \frac{e^p}{1+e^p}$.

La régression logistique binomiale, à l'instar de la régression logistique binaire, est utilisée lorsque la variable dépendante n'est pas une variable binaire en elle-même mais un compte basé sur une variable binaire.

Dans le cadre de cette étude, la variable à expliquer Y est le décès. La variable Y prend comme valeur 0 dans le cas où l'individu ne décède pas et 1 s'il décède. La finalité de cette régression serait de modéliser la probabilité de survie en fonction des variables explicatives : $P(Y = 0|X)$.

Le modèle s'écrit alors de la façon suivante :

$$\begin{aligned} \text{logit}(P(Y = 0|X)) &= X^T \beta \\ \Leftrightarrow P(Y = 0|X) &= \frac{e^{X^T \beta}}{1 + e^{X^T \beta}} \end{aligned}$$

La finalité du modèle serait de modéliser l'évènement où D , le nombre de décès. Afin d'apprendre au mieux les informations fournies par les observations, mais aussi pour éviter de ne pas considérer les phénomènes de censure, la prise en compte de l'exposition au risque dans la modélisation est indispensable. L'exposition au risque E est pris en compte sous forme d'*offset* dans la régression. Ajouter cet *offset* à notre modèle permet de corriger le poids de chacune des observations.

La probabilité de survie modélisée n'est plus $P(Y = 0|X)$ mais $P(Y = 0| (X, E))$:

$$P(Y = 0| (X, E)) = \left(\frac{e^{X^T \beta}}{1 - e^{X^T \beta}} \right)^E$$

2. Prérequis

Le portefeuille assuré est composé d'une ligne d'observation pour chacun des Assurés. Puisque la régression logistique se fera sur la totalité de ce portefeuille et sur l'ensemble des années d'observations, il faudra bien distinguer l'évolution de chacun des individus chacune des années observées. De ce fait, chaque ligne d'observation s'est vue dupliquée 10 fois, correspondant au nombre d'années. Le résultat donne le suivi de chaque client sur les 10 années observées.

La seconde complexité qui rentre en jeu est liée à notre souhait de modéliser par période. Chaque individu se voit attribuer deux âges dans l'année ; avec pour frontière sa date d'anniversaire. Pour avoir une modélisation des probabilités de survie par période, chaque observation ligne d'observation a été une fois de plus dupliquée deux fois.

Le calcul de l'exposition au risque ne se fait pas en nombre de jours comme précédemment mais en pourcentage de temps passé sur une année complète. L'exposition est calculée pour chaque ligne d'observation de la façon suivante :

- Pour les lignes correspondantes à l'âge en début d'année :

$$E = \frac{(\text{Min}(\text{Date de sortie}, \text{Date d'anniversaire}) - \text{Max}(\text{Date d'entrée}, 01/01/\text{Année}))}{\text{Nombre de jours dans l'année}}$$

- Pour les lignes correspondantes à l'âge après anniversaire :

$$E = \frac{(\text{Min}(\text{Date de sortie}, 31/12/\text{Année}) - \text{Max}(\text{Date d'entrée}, \text{Date d'anniversaire}))}{\text{Nombre de jours dans l'année}}$$

Les lignes avec des expositions nulles peuvent ensuite être retirées, elles correspondent à des observations qui n'ont pas eu lieu dans l'année et pour cet âge.

L'objectif est de construire le meilleur modèle à partir des variables explicatives à disposition :

Les variables originaires de la base extraite :

- Le sexe de l'assuré ;
- L'année d'observation déterminée suite à la duplication des observations ;
- L'âge de l'assuré observé ;
- Le code produit sur lequel le contrat a été souscrit ;
- Le segment d'origine du contrat ;
- Le mode de gestion choisi par l'assuré.
- Le code postal du lieu de résidence ;
- La provision mathématique (PM), ou encours total, disponible sur le contrat le mois avant la date de sortie.

Les variables externes :

- Le revenu moyen affilié à sa ville de résidence ;
- Le pourcentage d'occupation parmi l'espace disponible dans la ville de résidence ;
- L'APL ou accessibilité potentielle localisée : c'est un indicateur local par commune qui met en avant les disparités d'offre de soins. Dans ce cas précis, c'est l'APL aux médecins généralistes qui est d'intérêt ;
- Le nombre de sites pollués dans la commune.

La base finale regroupe 2 896 856 observations de 14 variables dont 12 explicatives, 1 à expliquer et l'exposition. Cette base est divisée en deux bases de façon aléatoire : une base d'entraînement correspondant à 60% des observations de la base initiale et une base de tests sur le reste des observations. L'apprentissage se fait donc sur une base ayant 1 738 113 observations au total.

3. Application

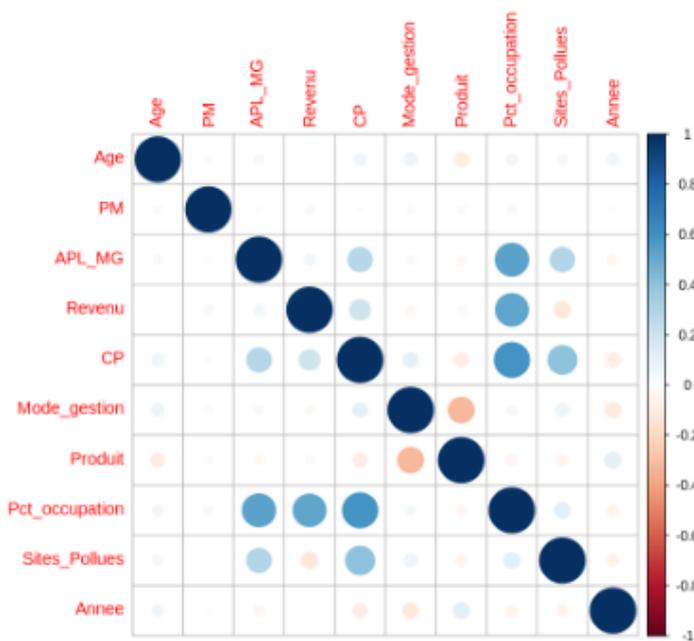


Figure IV-42 : Coefficients de corrélation entre variables explicatives

Afin d'obtenir le meilleur modèle, des algorithmes de sélection de variables explicatives sont utilisés. Il faut avant toute chose se rendre compte de la corrélation entre les différentes variables ; c'est ce que met en avant le graphique ci-dessus. Quelques variables semblent légèrement corrélées mais les coefficients restent globalement au-dessous de 50%. Le graphique laisse donc penser que les variables sont toutes utilisables pour la modélisation.

Les algorithmes effectuent les calculs de significativité du modèle en fonction d'un indicateur choisi au préalable. Concernant notre étude, c'est l'AIC qui sera l'indicateur retenu pour départager la significativité des variables et du modèle.

- La méthode Backward débute avec un modèle contenant toutes les variables, donc les 12 disponibles, et retire au fur et à mesure les variables qui pèchent à la significativité du modèle. Ce en se basant sur le critère d'Akaike.
- La méthode Forward au contraire part d'un modèle presque vide, avec comme pour seule variable explicative une constante. Les variables qui permettent d'améliorer le modèle sont ajoutées au fur et à mesure.
- Le Stepwise, où à chaque étape le modèle est impacté par l'ajout ou la soustraction de variable explicative.

Ces méthodes donnent chacune un modèle réalisé à partir de l'ensemble des variables explicatives de l'étude.

Méthode	Nombre de variables	AIC	BIC	Déviante
Backward	11	95049	95234	95019
Forward	10	95880	95224	95023
Stepwise	11	95049	95234	95019

Tableau 24 : Quantités de comparaisons pour les modèles obtenus avec les méthodes de sélection

Les méthodes Backward et Stepwise ont pour résultat un modèle équivalent composé de 11 variables. La méthode Forward quant à elle est composée d'une variable de moins mais possède un AIC plus élevé. Puisqu'il convient de retenir l'AIC le plus faible, le modèle résultat des deux autres méthodes sera retenu malgré que leur BIC semble être au contraire plus élevé. La déviance rejoint les résultats de l'AIC en désignant comme meilleur modèle celui à 11 variables.

Finalement, les variables explicatives retenues pour cette régression logistique sont les suivantes : le sexe, l'année, l'âge, l'accès aux médecins généralistes, le revenu moyen de la commune, le code postal, le mode de gestion, le code produit, la segmentation, le pourcentage d'occupation et le nombre de sites pollués.

En appliquant la régression à la base test, des probabilités sont obtenues. Celles-ci varient de $7,95 \times 10^{-5}$ et 0,65. En considérant que les probabilités supérieures ou égales à 0,5, 253 décès sont prédits dans la base test.

Afin de se rendre compte de la qualité du modèle, il est possible de faire appel à la matrice de confusion. L'intérêt de la matrice de confusion est de pouvoir comparer les étiquettes prédites et réelles de chaque individu. La comparaison se fait donc entre les prédictions issues du GLM et la variable binaire de décès issue du portefeuille.

La matrice de confusion se présente de cette manière :

Matrice de confusion	Positif	Négatif
Testé positif	Vrai positif	Faux positif
Testé négatif	Faux négatif	Vrai négatif
Total	Positifs	Négatifs

Cette matrice de confusion permet d'obtenir des indicateurs de validité du modèle :

- L'*Accuracy* ou proportion d'individus bien prédits par le modèle. Elle est à utiliser avec attention puisqu'elle donne une précision globale du modèle et si l'une des classes est disproportionnée par rapport à une autre, la classe la moins représentée sera peu prise en considération.

$$Accuracy = \frac{Vrai\ positif + Vrai\ négatif}{Total}$$

- La sensibilité est la probabilité que l'individu soit testé négatif (et donc décédé par le modèle) sachant qu'il est négatif.

$$Sensibilité = \frac{Vrai\ positif}{Positifs}$$

- La spécificité au contraire désigne la probabilité qu'un individu positif (en vie) ne soit pas testé négatif.

$$Spécificité = \frac{Faux\ positif}{Négatifs}$$

L'*Accuracy* obtenu pour la régression est de 99,5%, qui est une très bonne valeur. Cette valeur est à prendre avec attention car l'une des deux classes est très disproportionnée puisque le nombre de décès reste assez faible comparé à l'ensemble des observations.

La sensibilité s'élève à 99,5% et la spécificité à 91%. La sensibilité et la spécificité offrent également de bons résultats ; le modèle prédit donc bien le décès ou non en prenant pour seuil de décès 0,5. Ces résultats restent à prendre avec des pincettes ; en effet, les classes de nos données restent fortement disproportionnées. Cela limite certaines erreurs de prédictions.

V. Comparaisons et application à la tarification

Plusieurs modèles ayant été appliqués, une comparaison est nécessaire. Le but est de pouvoir se rendre compte de quels sont les modèles qui donnent les résultats les plus fiables par rapport aux données ; desquels offrent les résultats les moins risqués pour l'assureur mais aussi de pouvoir évaluer l'écart entre les résultats des différents modèles et ceux de la table de mortalité générationnelle en vigueur pour les rentes viagères.

Les tables de mortalités obtenues seront également appliquées à la tarification afin de directement se rendre compte de l'impact qu'il y aura sur les rentes viagères.

Les modèles à comparer sont donc les suivants :

- Le modèle Cairns-Blake-Dowd appliqué à la population masculine du portefeuille et fermé par la méthode Coale-Kisker à partir de 85 ans, **CBDM**.
- Le modèle Cairns-Blake-Dowd appliqué à la population féminine du portefeuille et fermé par la méthode Coale-Kisker à partir de 85 ans, **CBDF**.
- Le modèle de positionnement SMR appliqué à la population masculine du portefeuille, avec la table de référence TGH05, fermé par une extrapolation exponentielle à partir de 85 ans, **SMRM**.
- Le modèle de positionnement SMR appliqué à la population féminine du portefeuille, avec la table de référence TGF05, fermé par une extrapolation Denuit & Goderniaux à partir de 85 ans, **SMRF**.
- Le modèle de positionnement relationnel à deux paramètres, appliqué à la population masculine du portefeuille, avec la table de référence TGH05, fermé par une extrapolation de type Kannistö à partir de 90 ans, **RelationnelM**.
- Le modèle de positionnement relationnel à deux paramètres, appliqué à la population féminine du portefeuille, avec la table de référence TGF05, fermé par une extrapolation de type Denuit & Goderniaux à partir de 90 ans, **RelationnelF**.
- Le modèle de positionnement par GLM Poisson, appliqué à la population masculine du portefeuille, avec la table de référence TGH05, fermé par une extrapolation exponentielle à partir de 90 ans, **PoissonM**.
- Le modèle de positionnement par GLM Poisson, appliqué à la population féminine du portefeuille, avec la table de référence TGF05, fermé par une extrapolation de type Kannistö à partir de 90 ans, **PoissonF**.
- Le modèle de positionnement par lissage paramétrique et dérive de mortalité, appliqué à la population masculine du portefeuille, avec la table de référence TGH05, fermé par une extrapolation exponentielle à partir de 90 ans, **LissageM**.
- Le modèle de positionnement par lissage paramétrique et dérive de mortalité, appliqué à la population féminine du portefeuille, avec la table de référence TGF05, fermé par une extrapolation exponentielle à partir de 90 ans, **LissageF**.
- La régression logistique appliquée sur le portefeuille, avec des variables explicatives internes et externes, **Logistique**.

A. Adéquation aux données

L'adéquation aux données a déjà servi de critère de validation et de sélection (pour le choix des fermetures de table notamment). Il est tout de même intéressant de pouvoir comparer les niveaux d'adéquation aux données brutes des différents modèles étudiés. La proximité aux données brutes sera donc étudiée sur la tranche d'âge allant de 60 à 100 ans. Les indicateurs qui sont retenus pour comparer ces modèles sont : la distance du χ^2 entre les taux de décès bruts et modélisés, la déviance des taux de décès et le nombre de décès total en appliquant les taux de décès à l'exposition réelle du portefeuille.

Concernant la régression logistique, il n'est pas possible d'en tirer une table de taux de décès, la comparaison portera donc sur le nombre de décès moyen uniquement. Le nombre de décès moyen sera prédit à partir de la base de données test.

Les modèles prédits à partir de la population féminine sont comparés aux taux bruts de décès des femmes calculés par la méthode d'Hoem du moment. Les modèles prédits à partir de la population masculine sont comparés aux taux bruts de décès des hommes par Hoem.

Méthode	χ^2	Déviance	Nombre de décès
CBD M	507,64	6 193	10 153
CBD F	584,52	5 992	9 016
SMR M	622,89	7 140	9 914
SMR F	654,74	5 841	9 142
Relationnel M	563,09	6 347	10 215
Relationnel F	589,87	5 737	8 825
Poisson M	546,83	6 519	10 089
Poisson F	576,96	5 638	8 920
Lissage M	519,27	6 399	10 099
Lissage F	530,71	6 054	8 504
Logistique	X	X	16 212

Tableau 25 : Critères de proximité entre modèles obtenus et valeurs brutes

Le nombre de décès bruts pour la population féminine est de 8 768 contre 10 347 décès pour la population masculine. Le nombre de décès total s'élève à 19 115 personnes.

L'ensemble des décès issus des taux modélisés sur la population masculine n'excède pas le nombre de décès brut de cette population. Les méthodes utilisées ont eu tendance à sous-estimer le nombre de décès au vu des résultats obtenus.

Les deux méthodes offrant les résultats les plus éloignés de la réalité sont les méthodes de SMR sur la population masculine et féminine. Les quantités de proximité obtenues et les nombres de décès estimés n'encouragent pas à dire que ces méthodes offrent une bonne fiabilité aux données brutes.

B. Niveaux de risques

Les comparaisons de cette partie ne se restreignent pas aux résultats allant de 60 à 100 ans mais sur la totalité des âges prédits (60 à 130 ans). Le but de cette partie est d'évaluer les tables de mortalités modélisées qui minimisent au mieux le risque assureur.

Un modèle prudent est un modèle qui surestime très peu les taux de décès. Au contraire, des taux de décès assez bas sont souhaités. L'espérance de vie liée à un modèle peu risqué doit être surestimée. Les individus doivent donc, en termes de prédiction, vivre le plus longtemps possible. L'assureur quant à lui portera finalement moins de risque en utilisant cette table de mortalité.

. Le nombre total de décès est cette fois-ci étudié sans différenciation entre sexe. Les taux de décès sont appliqués à la somme des expositions des femmes et des hommes pour les 10 années d'observations et pour les âges allant de 60 à 100 ans. L'évolution temporelle des taux bruts est calculée comme la différence, pour chaque modèle, des taux obtenus en 2010 et de ceux obtenus en 2089.

Méthode	Nombre de décès	Evolution temporelle
CBD M	21 315	4,55
CBD F	17 272	7,56
SMR M	20 788	4,44
SMR F	17 072	4,23
Relationnel M	21 592	2,07
Relationnel F	16 580	4
Poisson M	21 143	4,01
Poisson F	16 770	0,9
Lissage M	21 227	4,45
Lissage F	16 168	4,56
Logistique	16 297	X

Tableau 26 : Nombre de décès et évolution temporelle des différents modèles

Les modèles offrant le moins de décès sont les modèles réalisés à partir de l'échantillon de population féminine. Le modèle Lissage F semble notamment être assez peu risqué.

Les résultats issus de la régression logistiques sont inutilisables pour l'assureur au vu du code des assurances qui ne permet pas de différencier un individu par critère de sexe. Le sexe est une variable importante de notre régression, ce qui rend ce modèle obsolète.

Concernant l'évolution temporelle, c'est un critère difficilement interprétable. Le souhait principal est de pouvoir observer une certaine évolution de taux au fil des années ; dans le but de prendre en considération l'allongement de la durée de vie. Il est difficile d'interpréter les résultats issus de cette différence entre les taux de 2089 et les taux de 2010 ; la seule observation à faire serait que le modèle Poisson F semble ne pas contenir d'évolution de taux et donc de sous-estimer l'allongement de la durée de vie humaine.

Ces comparaisons ont permis de se rendre compte des caractéristiques propres à chacun des modèles. Il convient alors de retenir 3 modèles applicables à la tarification ; les modèles qui semblent être les plus cohérents sont retenus : CBD F ; Relationnel F et Lissage F

C. Application aux calculs de rentes et autres comparaisons

A. Application aux calculs de rentes

C'est en se basant sur la formule de base des rentes viagères que les modèles seront appliqués. La table de mortalité en vigueur pour calculer ce type de rente est la TGF05. La TGF05 est une table de mortalité générationnelle qui a été construite à partir de données de population féminine. Elle est utilisée pour les rentes viagères puisque son utilisation permet d'avoir des probabilités globalement plus faibles que les autres tables en vigueur.

Afin d'étudier les différents montants de rente obtenus à partir des tables à disposition, 4 cas de figures sont considérés :

- Individu 1 : Un assuré né en 1970 qui voit sa rente débiter à 60 ans ;
- Individu 2 : Un assuré né en 1970 qui voit sa rente débiter à 90 ans ;
- Individu 3 : Un assuré né en 2005 qui voit sa rente débiter à 60 ans ;
- Individu 4 : Un assuré né en 2005 qui voit sa rente débiter à 90 ans.

C'est sur ces 4 individus que sera calculé le montant de la rente viagère à partir d'un capital constitutif commun s'élevant à 100 000 €. La rente calculée sera trimestrielle.

Méthode	Relationnel F	Lissage F	CBD F	TGF-05
Individu 1	2 971,33 €	2 839,30 €	2 884,50 €	2 861,73 €
Individu 2	12 267,36 €	9 627,01 €	9 259,76 €	10 752,43 €
Individu 3	2 647,21 €	2 446,80 €	2 497,92 €	2 511,21 €
Individu 4	9 687,07 €	7 219,13 €	6 677,89 €	7 956,20 €

Tableau 27 : Montant annuel de la rente pour chaque modèle et individu

Le modèle qui offre les montants annuels de rente les plus élevés est le modèle Relationnel F. Celui-ci offre les montants les plus élevés des 4 modèles ; ce résultat laisse sous-entendre que les taux de décès sont surestimés sur ce modèle comparé aux 3 autres qui offrent des montants de rentes plus faibles et donc plus de sûreté à l'assureur. Ce modèle était l'un des plus fiables par rapport aux données brutes du portefeuille ; néanmoins, ce modèle a un niveau de risque trop élevé en offrant des montants de rentes plus élevés que la TGF05. C'est contradictoire avec le montant de décès total évalué au cours de la comparaison mais cet indicateur est très dépendant de l'exposition au risque.

Le second modèle étudié dans cette tarification offre les montants les plus faibles pour un individu qui débute sa rente à 60 ans. Cela signifie que la méthode de lissage appliquée sur les taux (qui consistait à lisser les taux bruts à partir d'une fonction puis à y appliquer une dérive de mortalité) donne des résultats satisfaisants en termes de risque pris par l'assureur. Le grand bémol de ce modèle est lié à sa construction. Comme précisé au cours de sa mise en place, c'est un modèle qui offre beaucoup de place à l'interprétation et qui dépendant du choix des paramètres, qui est très subjectif.

Le modèle CBD F offre quant à lui les rentes les moins élevées pour un individu âgé de 90 ans. Cela signifie que ce modèle ne surestime pas les taux de mortalité aux âges élevés.

Les résultats obtenus pour la table générationnelle femme en vigueur se situent dans la moyenne des autres modèles.

B. Espérance de vie

Outre l'application aux calculs de rentes, d'autres comparaisons peuvent être faites sur ces tables de mortalité. Pour s'intéresser de plus près aux espérances de vies, nous avons tracé la courbe des espérances de vie d'un individu né en 1958 en fonction des âges pour les 3 modèles retenus et les tables réglementaires :

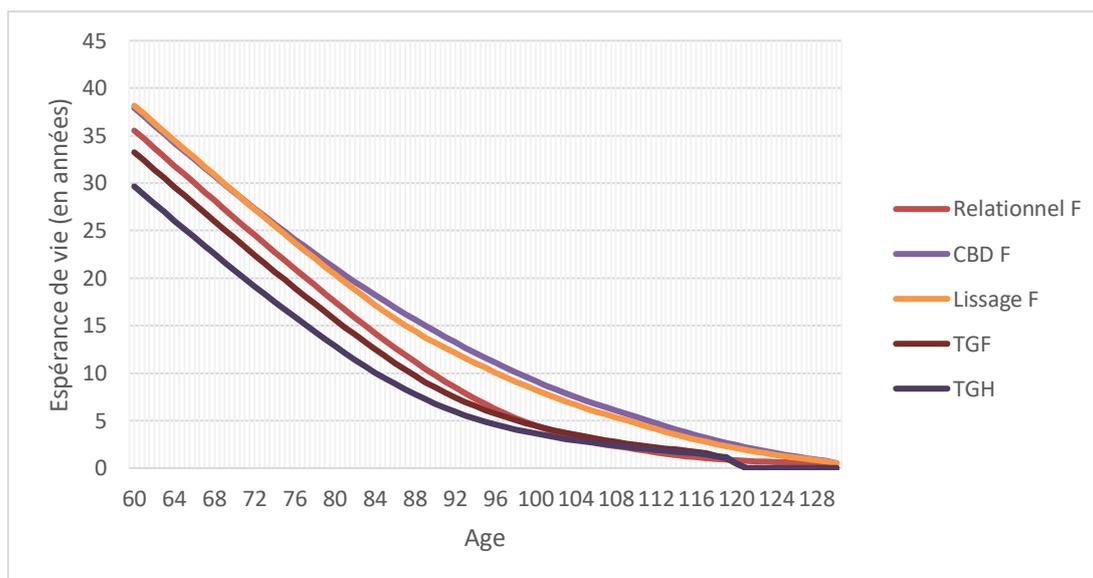


Figure V-1 : Espérance de vie pour l'année de naissance 1958 pour chacun des modèles

Les deux modèles qui offrent les espérances de vie les plus hautes dans ce cas de figure sont CBD F et Lissage F. Il est difficile de les distinguer aux premiers âges mais à partir de 75 ans, c'est la CBD F qui est au-dessus ; signifiant que c'est le modèle qui fait prendre le plus de précautions en termes d'estimation de décès.

A l'inverse, les tables réglementaires ont des espérances beaucoup plus basses. La TGH05 est la table offrant les espérances de vies les plus faibles. C'est également pour cette raison que dans le cadre de la mise en place d'une rente viagère, c'est la TGF05 qui est la table réglementaire généralement utilisée.

Aux grands âges, la différence est flagrante entre CBD F, Lissage F et les autres tables. En effet, si l'on se place à 110 ans, l'espérance de vie d'un individu né en 1958 est de 6 ans avec CBD F, 4,72 avec Lissage F et aux alentours de 2 ans pour les autres.

Si un seul des modèles étudiés devrait être retenu de cette étude, ce serait le CBD F. D'une part pour les bons résultats obtenus lors de l'application aux calculs de rentes ; mais aussi pour sa bonne adéquation aux données brutes.

Conclusion

Dans ce mémoire, nous avons présenté le viager immobilier. Le viager immobilier est une transaction comportant une multitude d'avantages pour les deux parties. Néanmoins, la dépendance des calculs de valeur du bien aux tables de mortalité comporte un risque pour l'acheteur : c'est le risque de longévité.

L'allongement de la durée de vie humaine a été présentée dans le cadre des données de la population française. Ce phénomène a une influence positive sur la longévité mais aussi des conséquences négatives pour les partis impactés par le risque de longévité.

La problématique actuarielle qui se dégageait était alors formulée de la façon suivante : « dans cet environnement de risque pour l'assureur, de quelle manière est-il possible de limiter le risque de longévité ? ».

Dans le cadre de la vente de viager, une réponse au risque de longévité a été abordée dans la première partie. Cette réponse était basée sur un nouveau mécanisme. Le viager immobilier n'étant pas contrôlé par un cadre réglementaire précis, les aménagements liés aux transactions restent possibles. C'est de cette façon qu'ont pu être présentées des conditions permettant de limiter ce risque tout en conservant les avantages du viager immobilier classique.

L'autre moyen de limiter le risque de longévité, qui s'affirmait comme l'ambition principale du mémoire, visait à construire une table de mortalité prospective d'expérience à partir des données du portefeuille épargne des CGI et grands comptes de Generali VIE. Puisque le risque de longévité est étudié, la table devait être construite sur les individus de plus de 60 ans uniquement. Nous avons pour souhait de garder une attention particulière aux grands âges qui sont de façon générale les plus difficiles à prédire.

Après avoir restreint la base de données à 10 années d'observations, pour les âges n'excédant pas 100 ans (par le critère de Cochran) ; des taux bruts de mortalité ont été calculés. Ces taux bruts nous ont permis d'appliquer plusieurs méthodes de modélisation. Les 5 méthodes qui suivent ont été appliquées sur la population féminine et masculine de façon distincte.

La première était de type stochastique ; projetant la mortalité aux années à venir, c'est le modèle Cairns-Blake-Dowd qui a été appliqué sur le portefeuille de base. Ce modèle régressif capte les évolutions tendanciennes de la mortalité en fonction du temps à travers deux paramètres.

Les 4 autres modèles qui ont été appliqués consistaient à positionner la mortalité de notre portefeuille par rapport à une table de référence donnée. La table de référence utilisée pour effectuer ces méthodes était la TGF05 pour les femmes et la TGH05 pour les hommes. La première méthode visait à appliquer un coefficient basé sur les données brutes aux tables de référence. La seconde méthode était basée sur une régression linéaire appliquée sur les taux bruts à partir de deux paramètres ; dont un qui capte les évolutions de la table de référence. La troisième méthode appliquée était de la famille des GLM puisque c'est un GLM Poisson qui a été appliqué aux taux instantanés de mortalité. La dernière méthode de positionnement appliquée visait à lisser nos taux bruts de mortalité à partir d'une fonction de lissage ; après avoir choisi les différents paramètres de cette fonction. Une dérive de mortalité issue des tables de référence a ensuite été appliquée.

Ces modèles ont été complétés aux grands âges par des méthodes de fermeture de table. Des analyses graphiques et basées sur des quantités observées nous ont permis de capter l'âge à partir duquel la fermeture doit être appliquée. Cet âge peut différer d'un modèle à un autre et d'un sexe à l'autre. Une seconde analyse a été appliquée à chaque fois afin de retenir la méthode de fermeture de table la plus adaptée aux données ; par critère de proximité aux données brutes. 4 fermetures de tables ont été comparées : l'extrapolation exponentielle, la méthode de Coale-Kisker, la fermeture de Denuit & Goderniaux et la méthode de Kannistö.

Nous avons également pour souhait de construire un modèle permettant de prendre en compte les variables explicatives que le portefeuille offrirait. C'est dans cette optique qu'une régression logistique a été réalisée. Cette méthode donnait des résultats convenables mais difficilement utilisables car elle ne correspond pas au format des tables habituellement certifiées.

L'ensemble de ces modèles (avec fermeture pour certains) ont ensuite été comparés. Les comparaisons se basaient sur deux critères : l'adéquation aux données brutes du portefeuille et le risque porté par l'assureur en utilisant ces tables. Dans l'ensemble, les modèles construits à partir de la population féminine du portefeuille ont offert de meilleurs résultats. A la suite de ces comparaisons, une application à la tarification d'une rente viagère classique a été faite. Les résultats ont montré que certains de nos modèles donneraient des rentes moins élevées que celles tarifées avec la TGF05.

L'utilisation de tables d'expérience peut être un réel plus pour les assureurs. Néanmoins, la contrainte de certification liée à l'utilisation de ces tables rend, dans certains cas, leurs utilisations compliquées en termes de temps et d'investissement.

D'autres mécanismes ont été pensés par les assureurs afin de limiter ce risque de longévité. Ces mécanismes de couvertures sont utilisés par les assureurs de manière assez récurrente avec deux catégories principales : la réassurance et les contrats dérivés.

Nous avons vu que l'allongement de la durée de vie humaine est un phénomène qui ne fait qu'accentuer la complexité d'estimation de l'espérance de vie et donc le risque longévité pris par l'assureur. Une question qu'il est possible de se poser est comment évoluera cet allongement ? Des théoriciens ont creusé la question et plusieurs hypothèses se dégagent : l'allongement de la durée de vie qui perdura de façon linéaire dans le temps (J. Waupel) et l'évolution limitée qui se résorbera dans peu de temps (J. Olshansky) sont notamment des hypothèses qui ont été émises. Aubrey de Gray suppose quant à lui que l'évolution de la longévité se fera de plus en plus rapidement et affirme que les Hommes vivront jusqu'à 1 000 ans.

Bibliographie

F.Abitbol, 2016, L'assurance de viager, Mémoire ISUP [1]

P. Saint Pierre, 2015, Introduction à l'analyse des durées de survie, cours Université Pierre et Marie Curie. [2]

O.Lopez, 2019, Survival analysis and other topics, cours ISUP. [3]

F.Planchet et P.Therond, 2006, Modèles de durée : applications actuarielles, *Economica*. [4]

Q.Guibert et F.Planchet, 2017, Utilisation des estimateurs de Kaplan-Meier par génération et de Hoem pour la construction de tables de mortalité, ISFA – Laboratoire SAF [5]

M.Boyer, C.Dorion et L.Stentoft, 2015, Les modèles factoriels et la gestion du risque de longévité, *L'Actualité économique, Revue d'analyse économique* [6]

E.Vanasse, 2019, Risque de longévité pour les régimes de retraite canadiens à prestations déterminées, Mémoire Université Laval [7]

Y.R. Safitri, S.Mardiyati et M.Malik, 2019, The Cairns-Blake-Dowd model to forecast Indonesian mortality rates, Université de Mathématiques et de Sciences Naturelles, Indonésie (FMIPA) [8]

A.M.Villegas, P.Millosovich et V.K.Kaishev, 2018, StMoMo : An R Package for Stochastic Mortality Modelling, Conférence à Brisbane, CEPAR, UNSW Sydney, CASS Business School, University of London [9]

J. Thomas et F.Planchet, Méthodes de positionnement : Aspects méthodologiques, ISFA – Laboratoire SAF. [10]

J. Thomas et F.Planchet, Construction d'une table de mortalité par positionnement : mode d'emploi, ISFA – Laboratoire SAF. [11]

G.Parat, 2016, Couverture assurantielle du risque de longévité dans le cadre d'un Prêt Viager Hypothécaire, Mémoire ISFA [12]

A.Quashie et M.Denuit, 2005, Modèles d'extrapolation de la mortalité aux grands âges, Institut des Sciences Actuarielles & Institut de Statistique Université Catholique de Louvain [13]

S.Guillaume, 2019, Modélisation de la mortalité dans le cadre du provisionnement de la garantie plancher en cas de décès : utilisation de méthodes de scoring, Mémoire ISFA [14]

I.Oirda, 2018, Construction de tables de mortalité prospectives pour les Organisations Professionnelles Agricoles, Mémoire ENSAE [15]

S.Gbari, M.Denuit et J.Trufin, 2017, Analyse de la mortalité aux âges extrêmes en Belgique, *Detralytics*. [16]

P.D.Thesis et M.D.Pascariu, 2018, Modelling and forecasting mortality, Université of Southern Denmark, SDU [17]

Annexes

1) Méthodes de positionnement : validation des modèles

a) Méthode n°2: Modèle relationnel à deux paramètres

i) Validation par critère de proximité sur la tranche [60,100]

Population	p-v du test de vraisemblance	p-v du test SMR	p-v du test de Wilcoxon	Résidus standardisés > 2	Déviante	χ^2	MAPE	R ²
Masculine	0,00	0,18	0,10	30	1171	512	27,83	0,64
Féminine	0,00	0,13	0,20	23	1126	469	27,08	0,82

Tableau 28 : Résultats des critères de proximité Méthode 2 sur la tranche [60,100]

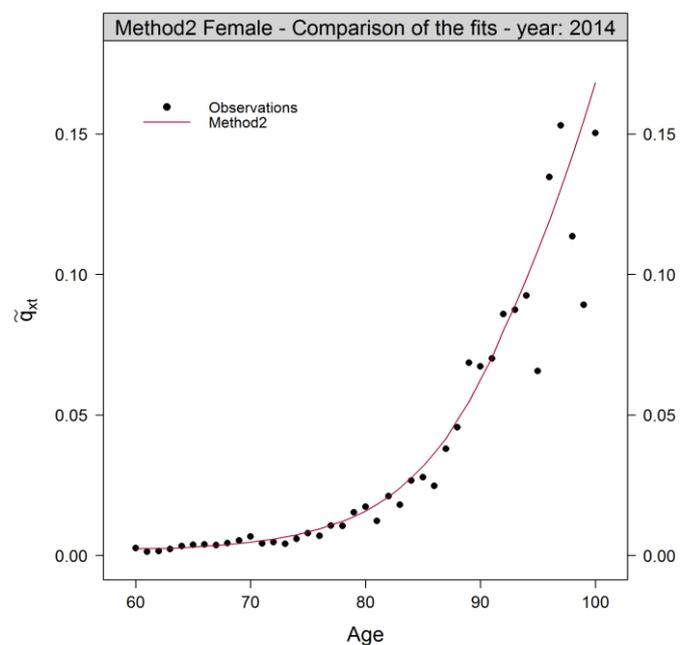


Figure 0-1 : Taux estimés par la méthode 2 (sur [60,100]) et observés pour la population féminine de 2014

Method2 Female - Plot of the residuals - year: 2014

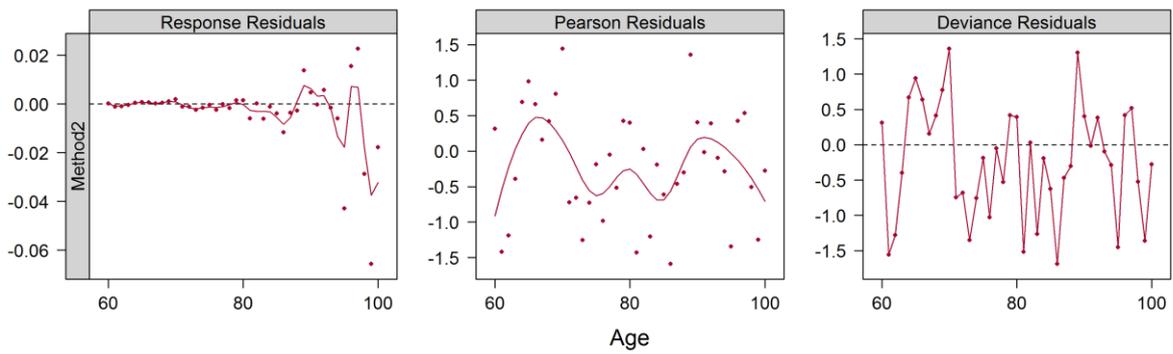


Figure 0-2 : Résidus de l'estimation de la méthode 2 ([60,100]) sur la population féminine de 2014

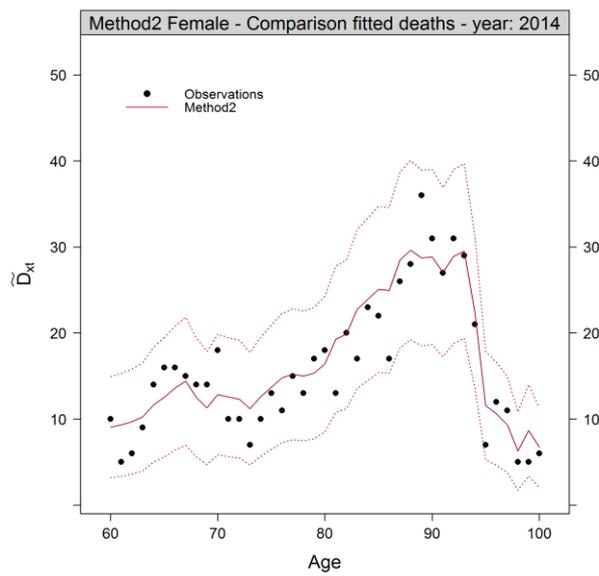


Figure 0-3 : Décès estimés par la méthode 2 ([60,100]) et observés sur la population féminine de 2014

ii) Validation par critère de régularité sur la tranche [60,90]

Population	p-v du test des runs	p-v du test des signes
Masculine	0,21	0,77
Féminine	0,31	0,61

Tableau 29 : P-valeurs des tests de régularité de la méthode 2 ([60,90])

b) Méthode n°3 : Modèle linéaire généralisé de Poisson

i) Validation par critère de proximité pour la tranche [60,100]

Population	p-v du test de vraisemblance	p-v du test SMR	p-v du test de Wilcoxon	Résidus standardisés > 2	Déviante	X ²	MAPE	R ²
Masculine	0,00	0,50	0,05	24	985	497	27,16	0,69
Féminine	0,08	0,50	0,09	20	898	450	26,15	0,83

Tableau 30 : Résultats des critères de proximité Méthode 3 sur la tranche [60,100]

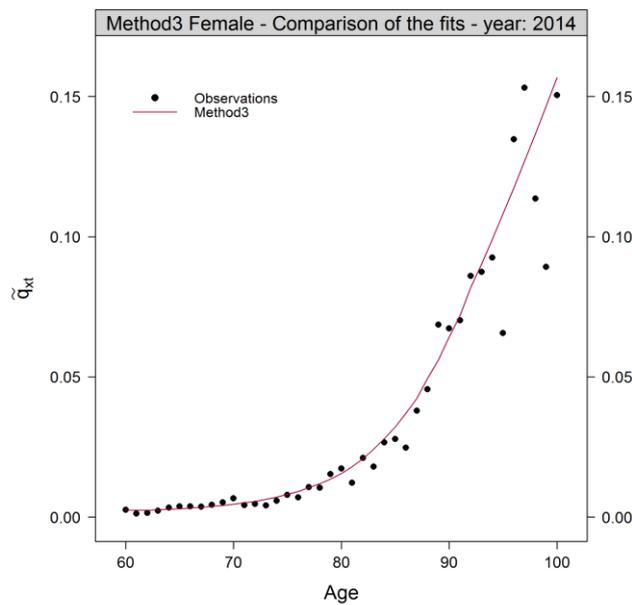


Figure 0-4 : Taux estimés par la méthode 3 (sur [60,100]) et observés pour la population féminine de 2014

Method3 Female - Plot of the residuals - year: 2014

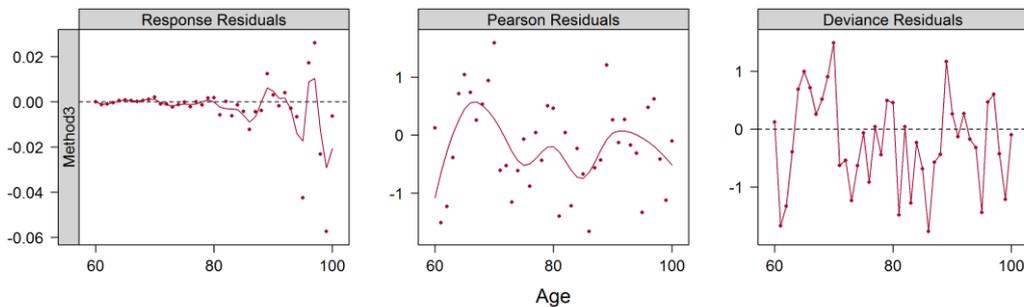


Figure 0-5 : Résidus de l'estimation de la méthode 3 ([60,100]) sur la population féminine de 2014

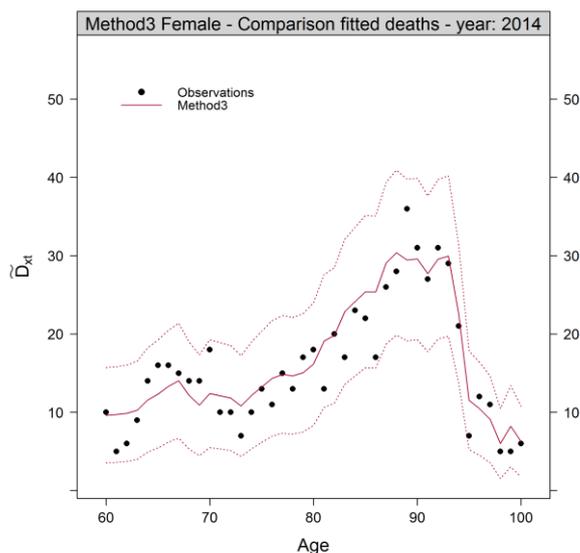


Figure 0-6: Décès estimés par la méthode 3 ([60,100]) et observés sur la population féminine de 2014

ii) Validation par critère de régularité pour [60,90] ans

Population	p-v du test des runs	p-v du test des signes
Masculine	0,15	0,28
Féminine	0,63	0,19

Tableau 31 : Résultats des critères de régularité Méthode 3 sur la tranche [60,90]

c) Méthode n°4: Lissage non paramétrique et dérive de mortalité

i) Validation par critère de proximité pour la tranche [60, 100]

Population	p-v du test de vraisemblance	p-v du test SMR	p-v du test de Wilcoxon	Résidus standardisés > 2	Déviante	χ^2	MAPE	R ²
Masculine	0,00	0,38	0,34	21	978	466	24,81	0,77
Féminine	0,14	0,45	0,10	21	879	449	26,22	0,84

Tableau 32 : Résultats des critères de proximité Méthode 4 sur la tranche [60,100]

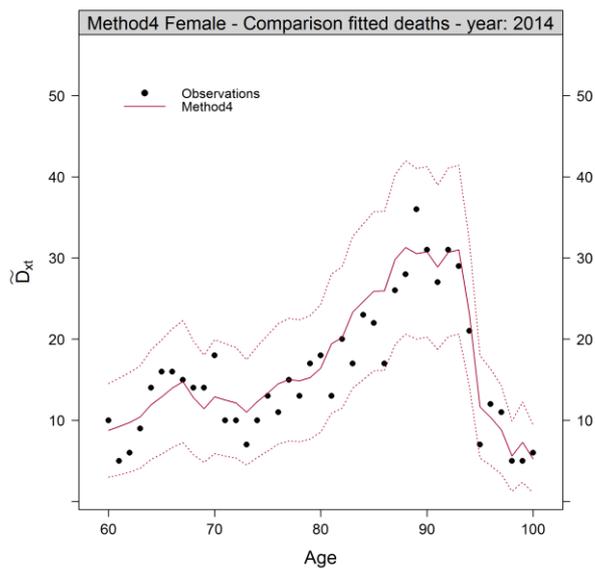


Figure 0-7 : : Décès estimés par la méthode 2 ([60,100]) et observés sur la population féminine de 2014

Method4 Female - Plot of the residuals - year: 2014

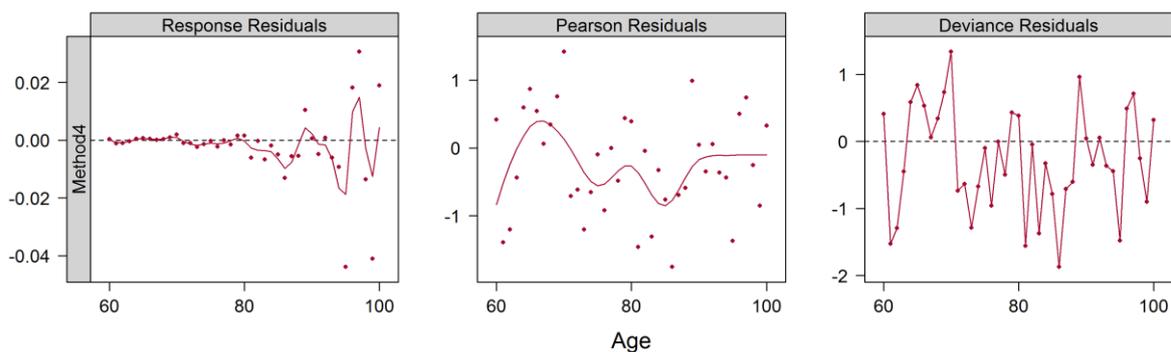


Figure 0-8 : Résidus de l'estimation de la méthode 4 ([60,100]) sur la population féminine de 2014

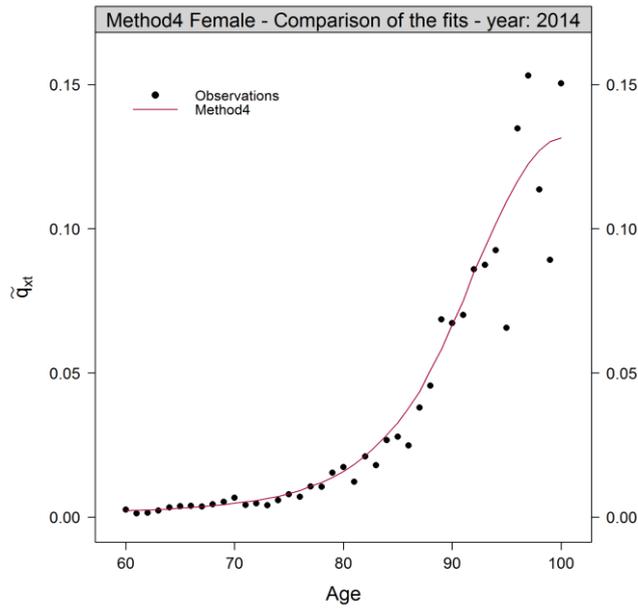


Figure 0-9 : Taux estimés par la méthode 4 (sur [60,100]) et observés pour la population féminine de 2014

ii) Choix des nouveaux paramètres de la fonction de lissage

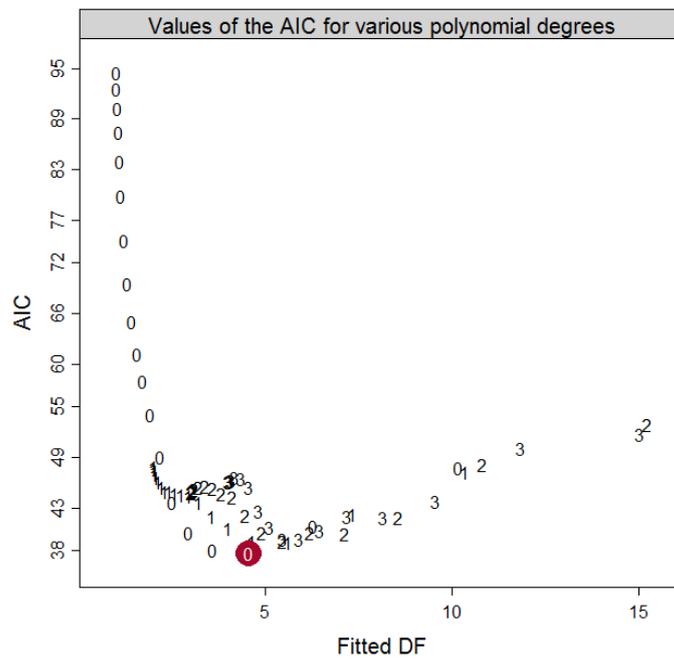


Figure 0-10 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population masculine nouvelle tranche

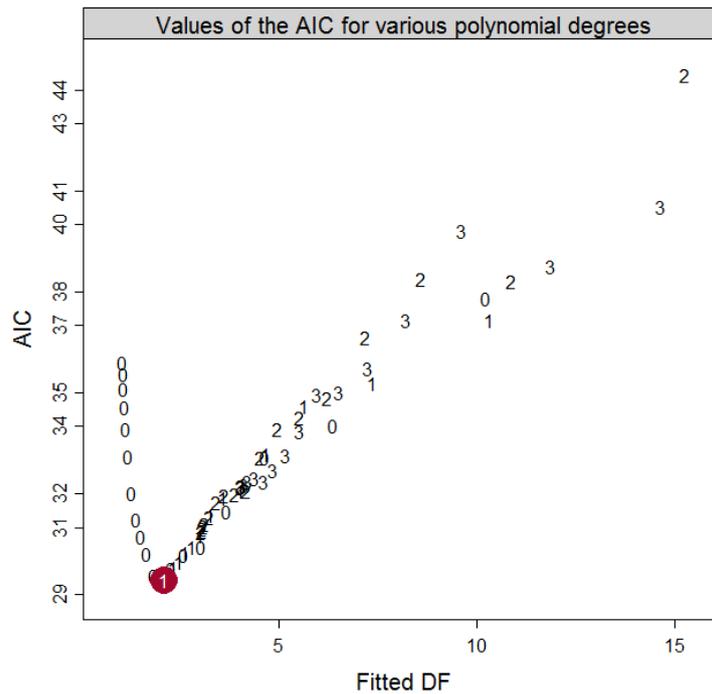


Figure 0-11 : Valeurs d’AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population féminine nouvelle tranche

iii) Validation par critère de proximité sur la nouvelle tranche [60,90]

Population	p-v du test des runs	p-v du test des signes
Masculine	0,43	0,61
Féminine	0,19	0,07

Tableau 33 : Résultats des critères de régularité Méthode 4 sur la tranche [60,90]

2) Comparaison des méthodes de fermeture de table pour chaque modèle

a. Méthode de positionnement n°1 : SMR

• Population féminine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	349,14	231,57	214,17	433,93
MAPE	30,33	22,55	22,05	23,93
Déviance	4576	2734	2671	2944

Tableau 34 : Quantités des différentes fermetures pour la Méthode 1 (femmes)

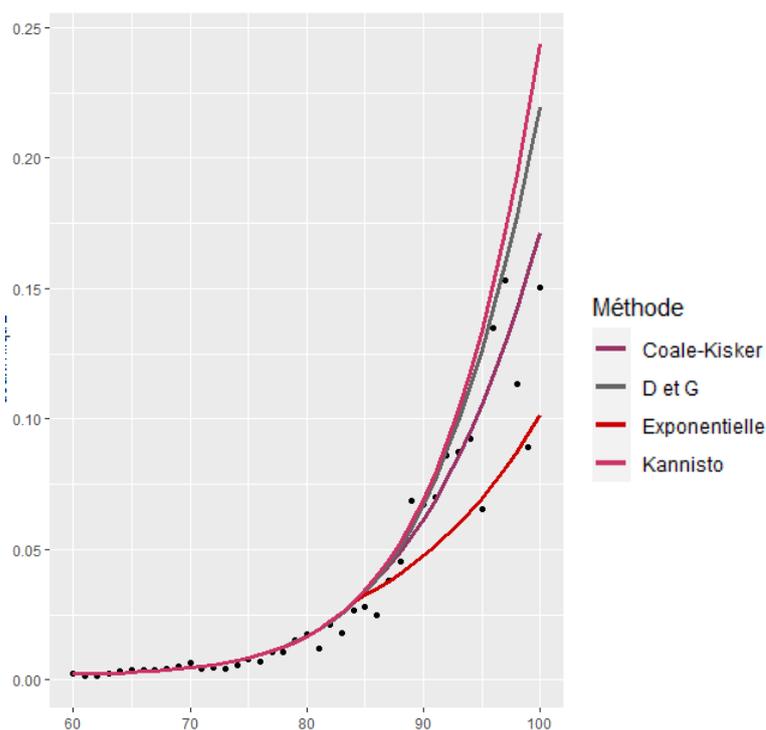


Figure 0-12 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 1

b. Seconde méthode de positionnement : modèle relationnel

- Population féminine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	143,96	210,5	164,67	229,54
MAPE	24,96	24,99	24,13	25,25
Déviante	2182,22	1773,30	1678,25	1795,50

Tableau 35 : Quantités des différentes fermetures pour la Méthode 2 (femmes)

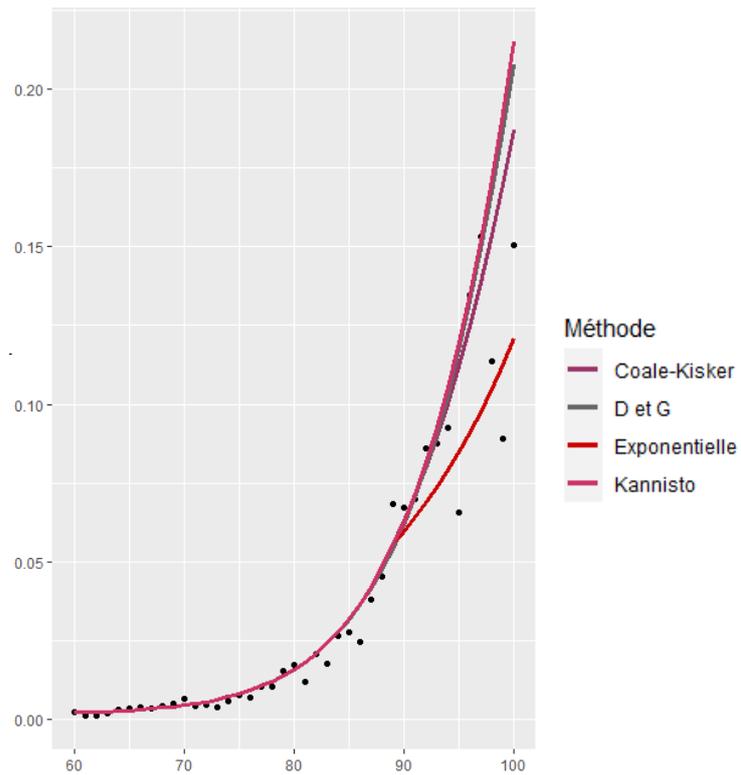


Figure 0-13 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 2

- Population masculine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	123,67	148,62	124,52	135,40
MAPE	30,85	32,45	31,95	31,81
Déviante	1450,59	1235,82	1232,45	1209,22

Tableau 36 : Quantités des différentes fermetures pour la Méthode 2 (hommes)

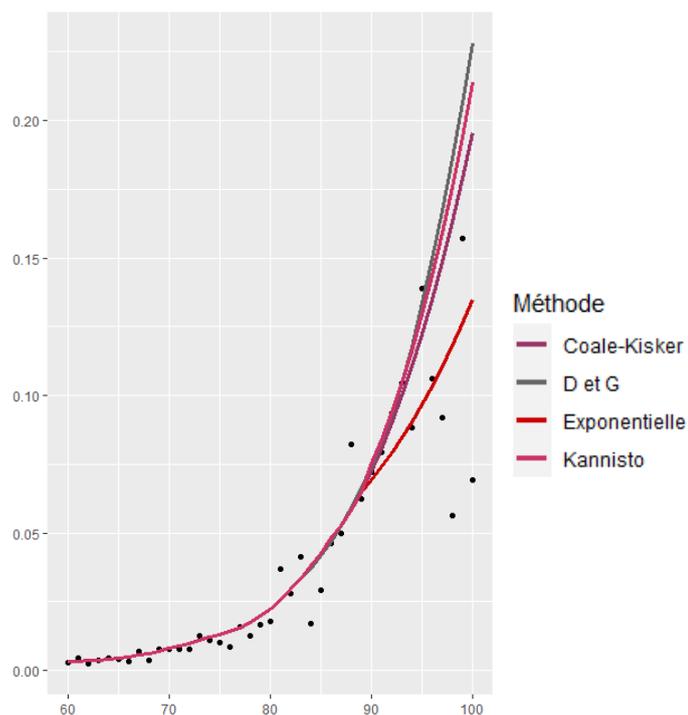


Figure 0-14 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 2

c. Méthode n°3 : Poisson

- Population masculine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	115,98	152,32	129,82	143,45
MAPE	28,48	32,2	31,47	31,57
Déviance	1362	1220	1198	1191

Tableau 37 : Quantités des différentes fermetures pour la Méthode 3 (hommes)

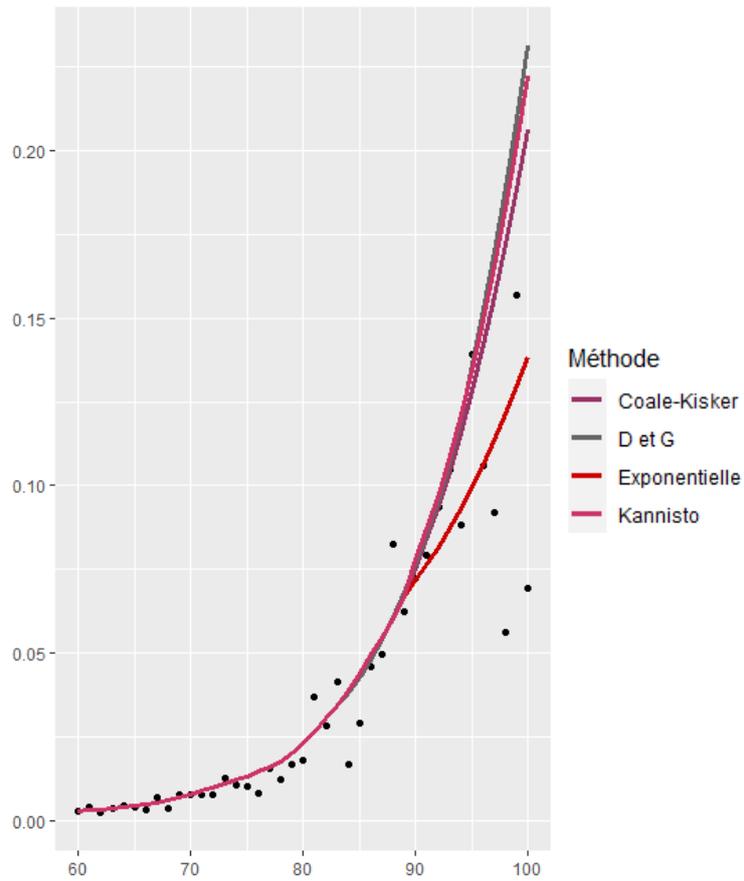


Figure 0-15 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 3

- Population féminine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	134,09	174,07	174,23	259,62
MAPE	24,34	24,99	24,27	25,47
Déviante	2068,66	1769,11	1686,36	1825,06

Tableau 38 : Quantités des différentes fermetures pour la Méthode 3 (femmes)

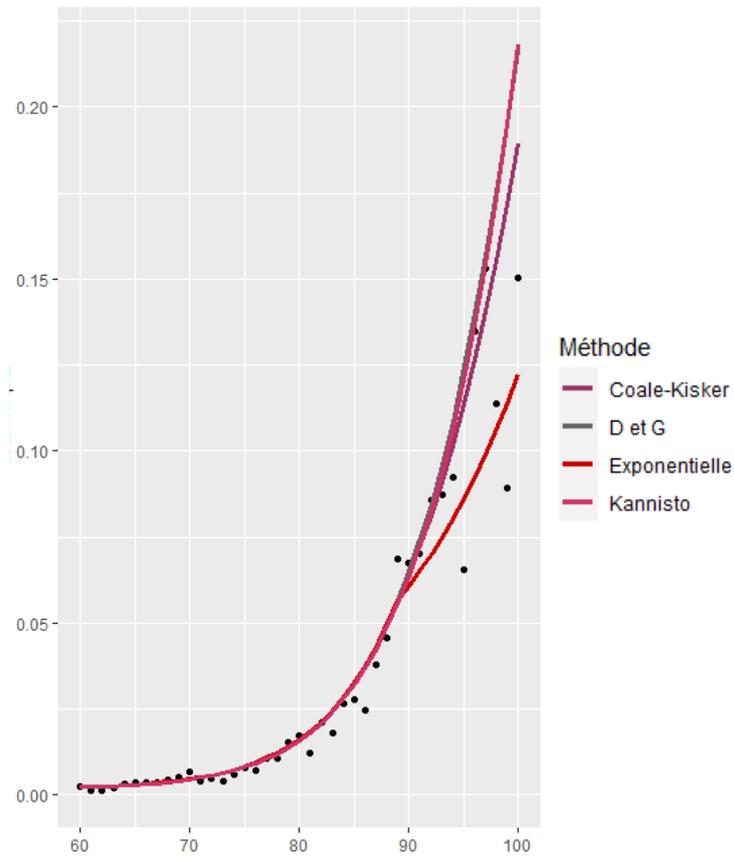


Figure 0-16 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 3

d. Méthode n°4 : Lissage

- Population masculine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	108,34	181,62	141,19	148,43
MAPE	26,27	31,15	31,17	31,67
Déviante	1274,71	1212,63	1176,68	1196,70

Tableau 39 : Quantités des différentes fermetures pour la Méthode 4 (hommes)

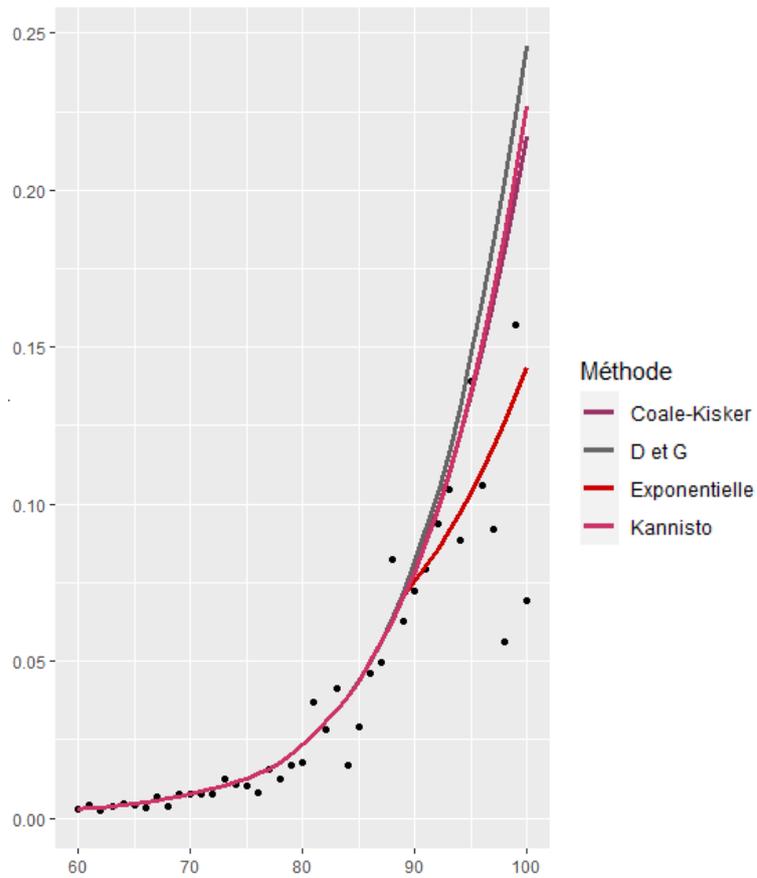


Figure 0-17 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 4

- Population féminine

Méthode	Exponentielle	Denuit & Goderniaux	Coale-Kisker	Kannisto
χ^2	138,96	243,22	169,82	244,04
MAPE	24,12	25,39	24,06	25,45
Déviante	2115,13	1802,31	1673,24	1810,05

Tableau 40 : Quantités des différentes fermetures pour la Méthode 4 (femmes)

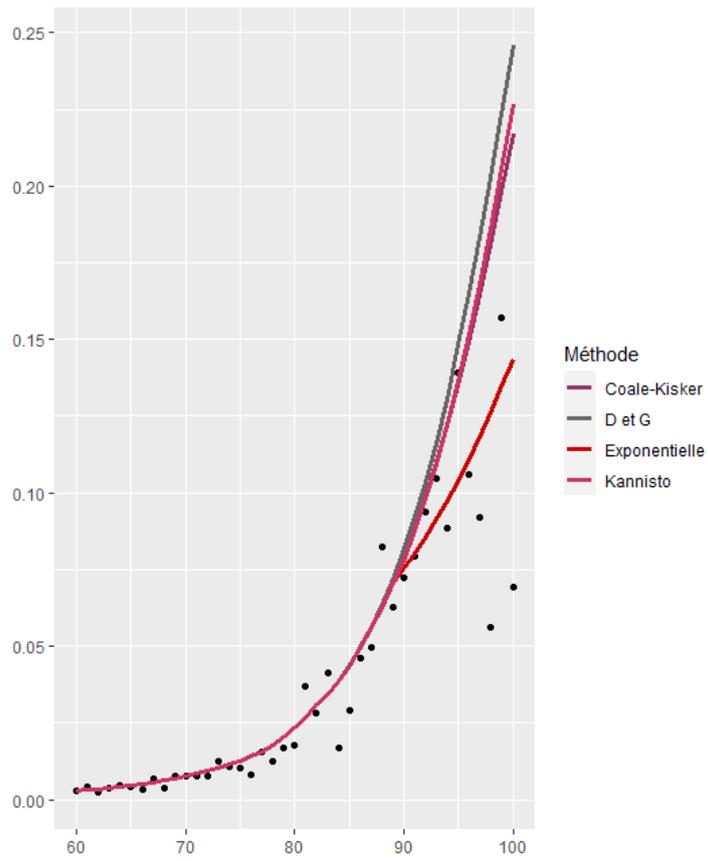


Figure 0-18 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 4

Table des illustrations

- Table des figures

Figure 0-1 : Evolution des espérances de vie à la naissance en France depuis 1950 à partir des tables de l'INSEE	0
Figure 0-2 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les hommes	0
Figure 0-3 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les hommes	0
Figure 0-4 : Evolution of life expectancy at birth in France since 1950, based on INSEE tables	0
Figure 0-5 : Death rates of the Hoem estimator by age and year for men	0
Figure 0-6 : Death rates of the Kaplan-Meier estimator by age and year for men	0
Figure II-1 : Causes de décès recensées en 2015 en France par sexe en milliers d'après les données de l'INSEEE	12
Figure II-2 : Taux de mortalité des hommes pour les années 2013 à 2015 basée sur la T68	12
Figure II-3 : Taux de mortalité infantile en France d'après les données de l'INSEEE.....	13
Figure II-4 : Evolution des espérances de vie à la naissance depuis 1950 à partir des tables de l'INSEE	14
Figure III-1 : Répartition des libellés de situation de contrats	20
Figure III-2 : Répartition des différents échelons de primes	21
Figure III-3 : Proportion des primes versées par segmentation	21
Figure III-4 : Répartition des âges d'assurés du portefeuille au 31/12/2019	22
Figure III-5 : Courbe retraçant les PM moyennes par âge à la sortie	23
Figure III-6 : Histogramme des âges de décès des individus du portefeuille	23
Figure III-7 : Proportion d'hommes et de femmes vivants par âge	24
Figure III-8 : Histogramme des répartitions de décès par âge et sexe.....	24
Figure III-9 : Exposition au risque de la population masculine par année	25
Figure III-10 : Exposition au risque par année sur la population féminine	26
Figure III-11 : Nombre de décès par année pour la population masculine.....	26
Figure III-12 : Nombre de décès par an pour la population féminine	27
Figure III-13 : Rapport entre décès et exposition par année chez les hommes	27
Figure III-14 : Rapport entre nombre de décès et exposition par année pour les femmes.....	28
Figure IV-1 : Zonier de validation des critères de Cochran pour les hommes (à gauche) et les femmes	30
Figure IV-2 : Exposition au risque en jours par âge et année pour les hommes du portefeuille	32
Figure IV-3 : Exposition en jours par âge et année pour les femmes du portefeuille	32
Figure IV-4 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les hommes	34
Figure IV-5 : Taux bruts de décès de l'estimateur d'Hoem par âge et année pour les femmes	34
Figure IV-6 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les hommes	35
Figure IV-7 : Taux bruts de décès de l'estimateur de Kaplan-Meier par âge et année pour les femmes	36
Figure IV-8 : Différence des taux de décès entre Hoem et Kaplan-Meier pour les hommes.....	36
Figure IV-9 : Différence des taux de décès entre Hoem et Kaplan-Meier pour les femmes.....	37
Figure IV-10 : Variance des taux de décès d'Hoem et de Kaplan-Meier par année pour les hommes..	39
Figure IV-11 : Variance des taux de décès d'Hoem et de Kaplan-Meier par année pour les femmes ..	39

Figure IV-12 : Kt1 estimé sur les 10 premières années pour la population masculine	42
Figure IV-13 : Kt2 estimé sur les 10 premières années pour la population masculine	43
Figure IV-14 : Kt1 estimé sur les 10 premières années pour la population féminine.....	44
Figure IV-15 : Kt2 estimé sur les 10 premières années pour la population féminine.....	44
Figure IV-16 : Cartographie des résidus du modèle estimé pour les hommes et les femmes (respectivement)	45
Figure IV-17 : qxt estimés pour la population masculine et féminine (respectivement)	46
Figure IV-18 : Estimation prédictive des Kt du modèle sur la population masculine.....	47
Figure IV-19 : Projections de scénarios aléatoires sur les paramètres Kt prospectifs pour les hommes	48
Figure IV-20 : Estimation prédictive des Kt du modèle sur la population féminine.....	49
Figure IV-21 : Projections de scénarios aléatoires sur les paramètres Kt prospectifs pour les femmes	50
Figure IV-22 : Taux bruts estimés et observés sur la population féminine de 2014	51
Figure IV-23 : Taux bruts estimés et observés sur la population masculine de 2014.....	52
Figure IV-24 : Taux bruts et estimés par différentes fermetures sur les hommes en 2014.....	53
Figure IV-25 : Taux bruts et estimés par différentes fermetures sur les femmes en 2014.....	54
Figure IV-26 : qxt estimés et projetés par Cairns-Blake-Dowd avec une extrapolation Coale-Kisker pour les hommes	55
Figure IV-27 : qxt estimés et projetés par Cairns-Blake-Dowd avec une extrapolation Coale-Kisker pour les femmes.....	55
Figure IV-28 : Taux de décès bruts (Hoem) et calculée avec la méthode SMR pour les femmes en 2018	60
Figure IV-29 : Résidus issus de cette première méthode pour les femmes de 2018	60
Figure IV-30 : Courbe des décès estimés par la méthode SMR et décès bruts sur la population féminine de 2018.....	61
Figure IV-31 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014	64
Figure IV-32 : Taux de décès estimés avec la méthode SMR et une fermeture de table exponentielle pour les hommes	65
Figure IV-33 : Taux de décès estimés avec la méthode SMR et une fermeture de table Denuit & Goderniaux pour les femmes	65
Figure IV-34 : Taux de décès estimés à partir de la méthode 2 et d'une fermeture de table de Kannisto pour les hommes.....	67
Figure IV-35: Taux de décès estimés à partir de la méthode 2 et d'une fermeture de table de Denuit & Goderniaux pour les femmes	68
Figure IV-36 : Taux de décès estimés à partir de la méthode 3 avec une extrapolation exponentielle pour les hommes	70
Figure IV-37 : Taux de décès modélisés par la méthode 3 pour les femmes avec une fermeture de table Denuit & Goderniaux	70
Figure IV-38 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population féminine	72
Figure IV-39 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population masculine.....	72

Figure IV-40 : Taux modélisés par la 4 ^{ème} méthode de positionnement avec une extrapolation exponentielle, population masculine.....	74
Figure IV-41 : Taux modélisés par la 4 ^{ème} méthode de positionnement avec une extrapolation exponentielle, population féminine	74
Figure IV-42 : Coefficients de corrélation entre variables explicatives.....	77
Figure 0-1 : Taux estimés par la méthode 2 (sur [60,100]) et observés pour la population féminine de 2014	89
Figure 0-2 : Résidus de l'estimation de la méthode 2 ([60,100]) sur la population féminine de 2014 .	90
Figure 0-3 : Décès estimés par la méthode 2 ([60,100]) et observés sur la population féminine de 2014	90
Figure 0-4 : Taux estimés par la méthode 3 (sur [60,100]) et observés pour la population féminine de 2014	91
Figure 0-5 : Résidus de l'estimation de la méthode 3 ([60,100]) sur la population féminine de 2014 .	91
Figure 0-6: Décès estimés par la méthode 3 ([60,100]) et observés sur la population féminine de 2014	92
Figure 0-7 : : Décès estimés par la méthode 2 ([60,100]) et observés sur la population féminine de 2014	93
Figure 0-8 : Résidus de l'estimation de la méthode 4 ([60,100]) sur la population féminine de 2014 .	93
Figure 0-9 : Taux estimés par la méthode 4 (sur [60,100]) et observés pour la population féminine de 2014	94
Figure 0-10 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population masculine nouvelle tranche.....	94
Figure 0-11 : Valeurs d'AIC par degrés de liberté et degrés de polynôme pour la méthode 4, population féminine nouvelle tranche	95
Figure 0-12 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 1	96
Figure 0-13 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 2	97
Figure 0-14 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 2	98
Figure 0-15 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 3	99
Figure 0-16 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 3	100
Figure 0-17 : Courbes des différentes fermetures de tables et des taux bruts pour les hommes en 2014 Méthode 4	101
Figure 0-18 : Courbes des différentes fermetures de tables et des taux bruts pour les femmes en 2014 Méthode 4	102

- Table des tableaux

Tableau 1 : Comparaison des montants de rente par individu et modèle (F=femme).....	0
Table 1 : Comparison of pension amounts per individual and model (F=woman).....	0
Tableau 2 : Barème des abattements octroyés aux rentes issues de vente de viager.....	2
Tableau 3 : Espérance de vie à la naissance par pays et sexe avec les données de l’OMS de 2015	11
Tableau 4 : Répartition des contrats par segmentations.....	21
Tableau 5 : Caractéristiques propres à chaque segment (âge moyen, produit principalement vendu, PM moyenne)	22
Tableau 6 : Récapitulatif d’âge moyen médian et modal par sexe, pour les décès ou au 31/12/2019.	25
Tableau 7 : Comparaison des différentes fermetures de table pour un CBD sur la population masculine	53
Tableau 8 : Comparaison des différentes fermetures de table pour un CBD sur la population féminine	54
Tableau 9 : Valeurs des coefficients SMR pour différentes tranches d’âge et sexes	57
Tableau 10 : Tests de validation par critère de proximité, méthode 1 sur la tranche [60, 100]	57
Tableau 11 : Quantités du critère de proximité pour la méthode 1 sur [60,100]	59
Tableau 12 : Récapitulatif des p-valeurs des tests de régularité d’ajustement	62
Tableau 13 : Quantités des différentes fermetures de table sur le modèle SMR pour les hommes	63
Tableau 14 : Coefficients de la méthode 2 pour la tranche d’âge [60, 100]	66
Tableau 15 : Coefficients de la méthode 2 sur la tranche [60, 90] ans.....	67
Tableau 16 : Quantités des critères d’ajustement pour la méthode 2 sur la tranche [60, 90]	67
Tableau 17 : Coefficients obtenus pour la 3 ^{ème} méthode sur la tranche d’âge [60, 100] ans.	69
Tableau 18 : Quantités d'ajustement de la méthode 3 sur [60, 90]	69
Tableau 19 : Coefficients de la méthode 3 sur [60, 90]	69
Tableau 20 : Paramètres de fonctions de lissage des taux bruts pour la méthode 4 [60,100] ans	73
Tableau 21 : Paramètres de fonctions de lissage des taux bruts pour la méthode 4 [60,90] ans	73
Tableau 22 : Critères d’ajustement de la méthode 4 sur la nouvelle tranche d’âge [60, 90] ans	73
Tableau 23 : Quantités de comparaisons pour les modèles obtenus avec les méthodes de sélection.	78
Tableau 24 : Critères de proximité entre modèles obtenus et valeurs brutes.....	81
Tableau 25 : Nombre de décès et évolution temporelle des différents modèles.....	82
Tableau 26 : Montant annuel de la rente pour chaque modèle et individu	84
Tableau 27 : Résultats des critères de proximité Méthode 2 sur la tranche [60,100]	89
Tableau 28 : P-valeurs des tests de régularité de la méthode 2 ([60,90])	90
Tableau 29 : Résultats des critères de proximité Méthode 3 sur la tranche [60,100]	91
Tableau 30 : Résultats des critères de régularité Méthode 3 sur la tranche [60,90]	92
Tableau 31 : Résultats des critères de proximité Méthode 4 sur la tranche [60,100]	92
Tableau 32 : Résultats des critères de régularité Méthode 4 sur la tranche [60,90]	95
Tableau 33 : Quantités des différentes fermetures pour la Méthode 1 (femmes).....	96
Tableau 34 : Quantités des différentes fermetures pour la Méthode 2 (femmes).....	96
Tableau 35 : Quantités des différentes fermetures pour la Méthode 2 (hommes)	97
Tableau 36 : Quantités des différentes fermetures pour la Méthode 3 (hommes).....	98
Tableau 37 : Quantités des différentes fermetures pour la Méthode 3 (femmes).....	99
Tableau 38 : Quantités des différentes fermetures pour la Méthode 4 (hommes)	100
Tableau 39 : Quantités des différentes fermetures pour la Méthode 4 (femmes).....	101

