

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

Par : Grégory Bokobza

Titre du mémoire : Ajustement des taux de mortalité aux grands âges.

Confidentialité : Non Oui (Durée : 1 an 2 ans)

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

*Membres présents du jury de la
filière :*

Signatures :

Entreprise :

Nom : Apécialis

Signature :



*Membres présents du jury de
l'institut des Actuaires :*

Signatures :

Directeur de mémoire en entreprise :

Nom : Philippe DELERIVE

Signature :



Invité :

Nom :

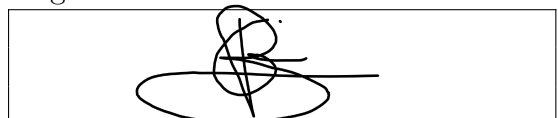
Signature :

.....
*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

L'évolution du niveau de vie moyen en France et les avancées de la médecine conduisent à des augmentations de l'espérance de vie. Il s'agit pour les assureurs vie, caisses de retraite et, plus généralement, pour tout organisme acceptant le risque de longévité, d'une augmentation non négligeable de ce risque. Cette dérive de la longévité nécessite une attention particulière étant donné la très forte volatilité, et donc les larges intervalles de confiance autour des bruts de mortalité obtenus à partir d'un certain âge.

L'objectif de ce mémoire est de confronter les principaux modèles de fermeture de tables et de challenger la table réglementaire aux grands âges à travers le provisionnement d'un portefeuille de rentes, dont les assurés sont tous des personnes âgées.

Pour cela, de nombreux modèles ont été développés, dont les modèles des P-splines, de Whittaker-Henderson ou encore le modèle de Brass mais il en existe beaucoup d'autres. Chacun de ces modèles a ses propres particularités et, selon la méthode utilisée, les résultats obtenus peuvent s'avérer très différents.

Mots clés : Tables de mortalité, Fermeture de tables, Grands âges, Risque de longévité, modèle des P-splines, modèle de Lindbergson, modèle de Denuit & Goderinaux, modèle de Kannistö, modèle de Whittaker-Henderson, modèle de Brass, Rentes viagères, Provisions.

Abstract

The evolution of the average standard of living in France and the advances in medicine are leading to increases in life expectancy. For life insurers, pension funds and, more generally, for any organization that accepts longevity risk, this represents a significant increase in this risk. This drift of longevity requires a particular attention given the very high volatility, and thus the large confidence intervals around the crude mortality figures obtained from a certain age.

The objective of this thesis is to compare the main table closure models and to challenge the regulatory table at high ages through the provisioning of a portfolio of annuities, whose insureds are all elderly.

For this purpose, many models have been developed, including the P-splines, the Whittaker-Henderson model and the Brass model, but there are many others. Each of these models has its own particularities and, depending on the method used, the results obtained can be very different.

Key words: Mortality tables, Closing of tables, Large ages, Longevity risk, P-splines model, Lindbergson model, Denuit & Goderniaux model, Kannistö model, Whittaker-Henderson model, Brass model, Life annuities, Provisions.

Remerciements

J'aimerais tout d'abord remercier Philippe Delerive, directeur général d'Apprecialis et tuteur de ce mémoire pour m'avoir permis d'effectuer une année d'apprentissage au sein de son cabinet, au cours de laquelle j'ai pu réaliser ce mémoire, ainsi que pour ces recommandations. J'aimerais également remercier Safia Lekehal Le Calvez pour sa relecture du mémoire, ses conseils et son temps.

J'aimerais remercier Olivier Lopez, directeur de l'ISUP et tuteur de ce mémoire, pour les pistes de réflexion amenées.

Je remercie également ma famille et mes amis pour leur soutien. J'aimerais notamment remercier Aurélia pour sa relecture attentive du mémoire.

Enfin, j'aimerais remercier toute autre personne ayant pu contribuer, d'une quelconque manière, à ce mémoire.

Sommaire

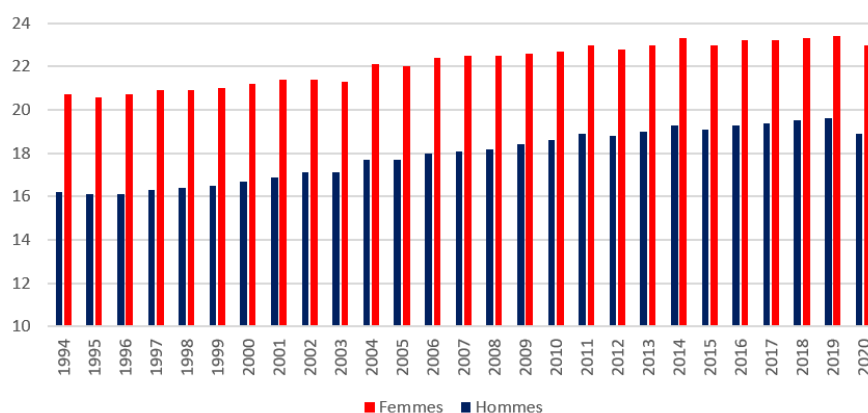
Résumé	2
Abstract	3
Remerciements	4
Introduction	8
1 Cadre réglementaire	11
1.1 Réglementation	11
1.1.1 Solvabilité 2	12
1.1.2 IFRS 17	12
1.2 Historique	13
2 Risque de longévité	16
2.1 Généralités	16
2.2 Facteurs de risque	19
2.3 Couverture du risque	21
2.3.1 Réassurance	21
2.3.2 Swap de longévité	21
2.3.3 Titrisation	22
3 Base de données	24
3.1 Présentation des données	24
3.2 Traitements	25
3.3 Statistiques descriptives	25
3.4 Qualité des données	31
3.4.1 Généralités	31
3.4.2 Tests de sensibilité au nombre de décès pour les supercentenaires	31

4	Construction de la table	35
4.1	Diagramme de Lexis	35
4.2	Durée de vie et notations actuarielles	36
4.3	Taux bruts de mortalité	37
4.3.1	Estimateur binomial	38
4.3.2	Estimateur de Kaplan-Meier	41
4.4	Mortalité prospective	43
5	Ajustement aux âges avancés	46
5.1	Modèles paramétriques	47
5.1.1	Modèle des P-splines	47
5.1.2	Modèle de Lindbergson	50
5.1.3	Modèle de Denuit & Goderniaux	54
5.1.4	Modèle logistique	58
5.2	Modèles non-paramétriques	60
5.2.1	Modèle de Whittaker-Henderson	61
5.2.2	Modèle de Whittaker-Henderson vu comme un ajustement bayésien	63
5.2.3	Modèles relationnels	66
6	Erreur de modèle	70
6.1	Tests de cohérence	70
6.2	Standardized Mortality Ratio (SMR)	71
6.3	Mean Absolute Percentage Error (MAPE)	73
6.4	Tests de sensibilité	75
7	Application à un portefeuille de rentes	84
7.1	Généralités	84
7.2	Présentation des données	86
7.3	Statistiques descriptives	87
7.4	Impacts financiers des ajustements	88
7.5	Tests de sensibilité	90
	Conclusion	92
	Table des figures	94
	Références	97
	Annexes	99
A	Modèle de Lee-Carter	99

B	Autres modèles d'ajustement aux grands âges	100
B.1	Modèle de Kimeldorf-Jones	100
B.2	Modèle à noyau	101

Introduction

Le record de longévité dans le monde est aujourd’hui officiellement détenu par une française décédée à l’âge de 122 ans et 164 jours. La doyenne française actuelle est âgée de 116 ans. Au-delà de ces deux exemples et, de manière plus générale, avec les évolutions du niveau de vie et les progrès de la médecine, une récente étude de l’INSEE¹ montre que l’espérance de vie résiduelle moyenne en France à 65 ans augmente assez rapidement. En effet, en 26 ans, celle-ci est passée de 20,7 ans en 1994 à 23 ans en 2020 pour les femmes, soit une augmentation de plus de 11%. La même étude montre que, pour les hommes, cette espérance de vie résiduelle est passée de 16,2 ans à 18,9 ans, soit une augmentation de plus de 16% en 26 ans. En se fondant sur les mêmes données, nous pouvons remarquer sur la Figure suivante que l’augmentation est assez régulière.



Évolution de l’espérance de vie résiduelle à 65 ans en France

La stabilité de l’évolution de l’espérance de vie résiduelle à 65 ans peut amener à penser que cette tendance puisse se prolonger dans le temps.

Cette dérive de longévité est étudiée de près par les démographes, statisticiens, actuaires – et les scientifiques de manière plus générale – qui tentent de modéliser la tendance future que la mortalité va suivre, ainsi que les conséquences que cela pourrait engendrer. Plusieurs opinions ont alors émergé : certains pensent que l’Homme pourrait

1. Institut National de la Statistique et des Études Économiques.

bientôt atteindre un âge qu'il ne pourra pas dépasser d'un point de vue biologique. La courbe du nombre de survivants en fonction de l'âge tendrait donc à se « rectangulariser ». D'autres pensent, au contraire, que la médecine, et des méthodes de plus en plus innovantes permettraient de vivre bien plus longtemps. Ces hypothétiques expériences et avancées médicales soulèvent cependant de nombreuses questions éthiques.

L'environnement actuel de taux bas, permet moins de limiter l'impact du risque de longévité qui prend, par conséquent, plus d'ampleur – notamment pour les assureurs vie, fonds de pension et caisses de retraite –, ce qui induit un montant de provision plus élevé. Il est donc essentiel pour les organismes proposant des prestations sous forme de rentes viagères, de prêter une attention particulière à ce risque. Il est primordial de pouvoir évaluer ces engagements sous les différents référentiels réglementaires en vigueur (principalement Solvabilité 2 et IFRS 17²). Pour cela, il faut être en mesure d'estimer la mortalité future du portefeuille d'assurés.

Afin de comprendre et de gérer le risque de longévité, les organismes concernés utilisent des tables de mortalité, décrites par le cabinet Aprealis comme étant la « présentation pour chaque âge du nombre de personnes survivantes. À partir de ces dernières, nous pouvons calculer le taux de décès propre à chaque âge. Les tables de mortalité permettent de calculer les tarifs et le provisionnement des risques décès. Elles peuvent être établies par génération, avec une table pour chaque année de naissance. Il s'agit donc de tables prospectives qui font l'hypothèse de gains de mortalité futurs. Elles peuvent aussi être valables pour l'ensemble des générations, elles reposent alors sur l'observation du passé : il s'agit de tables rétrospectives qui ne tiennent pas compte des gains de mortalité à chaque génération ».

Afin de pouvoir obtenir le nombre de survivants à chaque âge, il est essentiel d'au moins prendre en compte les deux éléments suivants :

- La longueur de récoltes des données et la fenêtre d'observation relativement courte favorisent les données incomplètes. Il faut donc corriger ce biais, notamment via la prise en compte de la censure et de la troncature. De nombreux mémoires sur la construction de tables traitent du sujet, voir par exemple WERTEL (2015) qui développe les estimateurs de Kaplan-Meier et d'Hoem pour pallier ce problème.
- Les faibles échantillons aux grands âges. En effet, le faible nombre de personnes vivantes à partir d'un certain âge provoque une forte volatilité sur les taux de mortalité qui deviennent alors assez incertains. Il convient alors de les ajuster via différents modèles.

2. « International Financial Reporting Standards ». Norme comptable internationale portant sur les contrats d'assurance datant de 2017 qui entrerait en vigueur en 2023.

Le but de ce mémoire est d'approfondir le deuxième sujet, le premier point reste toutefois traité dans les grandes lignes, l'objectif restant de challenger les différents modèles de fermeture de tables entre eux.

Pour réaliser cette étude, nous commencerons par décrire le cadre réglementaire autour des tables de mortalité en général, avant de s'intéresser au risque de longévité. Après avoir introduit les données, nous rappellerons et utiliserons certains estimateurs pour déterminer les taux de mortalité. Nous verrons également que ceux-ci restent très incertains au-delà d'un certain âge. Nous nous attarderons ensuite sur plusieurs modèles d'ajustement aux grands âges. Cela permettra, entre autres, de limiter la volatilité des taux de mortalité et le degré de confiance associé. Un risque d'erreur sera ensuite estimé par différentes méthodes statistiques. Nous obtiendrons ainsi des critères afin de comparer les modèles entre eux. Enfin, une étude sur les impacts financiers des différents ajustements sur le montant des provisions d'un portefeuille de rentes viagères sera réalisée.

Plusieurs mémoires d'actuariat et papiers de recherche traitent du sujet de la mortalité aux grands âges, ou de la mortalité pour des échantillons restreints, généralement, adaptables aux âges avancés. Certains d'entre eux, qui traitent de la construction de tables abordent également ce sujet. Ces articles et mémoires proposent deux types d'approches :

- L'approche endogène (cf. WENDJA (2007), CAPPE (2018) ou QUASHIE et DENUIT (2005) par exemple), consiste à utiliser l'information contenue dans les taux bruts de mortalité pour ajuster un modèle, permettant ainsi une projection réaliste par rapport aux attentes. Cependant, comme expliqué dans PLANCHET et LELIEUR (2007), dans le cas d'une population réduite, un ajustement endogène peut paraître « arbitraire » du fait de la forte volatilité et des attentes que l'on peut avoir.
- L'approche exogène consiste, au contraire, à se fonder sur une référence extérieure tout en modélisant l'écart entre la population étudiée et la population de référence. Celle-ci peut, par exemple, être une table réglementaire. Il existe, ici également, plusieurs méthodes comme la méthode de décalage d'âges, l'application d'un coefficient d'abattement ou encore des méthodes de positionnement (cf. CAPPE (2018)). Un GLM³ peut également être utilisé avec la mortalité de référence comme covariable.

3. Generalized Linear Model.

Chapitre 1

Cadre réglementaire

Du fait de l'inversion du cycle de production en assurance, les organismes assureurs connaissent l'engagement de l'assuré sans connaître le leur au moment de la signature du contrat. Il faut cependant que l'assureur puisse faire face à ses engagements en cas de sinistre et, c'est pourquoi il lui est essentiel d'estimer la charge future des sinistres.

Dans le cas du risque décès, cela passe notamment par l'utilisation d'une table de mortalité représentant (au mieux) la population assurée à chaque âge. Cela permet de connaître la durée de vie résiduelle des assurés, et ainsi, de pouvoir déterminer le montant de provisions, qui se calcule comme une somme actualisée des flux probables futurs que l'assureur devra verser. Cela représente donc une estimation précise de l'engagement de l'assureur. L'utilisation de ces tables primordiale et même réglementaire.

1.1 Réglementation

Il est fréquent pour les organismes assureurs d'utiliser suivant les cas, les tables réglementaires TGH05/TGF05 et TH 00-02/TF 00-02 fondées sur une population globale plutôt que des tables d'expérience. Les tables par générations sont utilisées pour des risques décès avec des engagements à long terme, comme pour des rentes viagères – garantie retraite par exemple. Cela permet de capter la dérive de la longévité au fil des générations. Les tables par moment seront utilisées, quant à elles, pour des risques liés au décès avec des engagements courts – garantie frais obsèques ou vie entière par exemple –, pour lesquels la dérive de longévité n'a pas de grandes conséquences.

L'utilisation des tables de mortalité pour le calcul des engagements d'assurance apparaît essentielle dans les deux principaux référentiels réglementaires : Solvabilité 2 et IFRS17.

1.1.1 Solvabilité 2

Le risque de mortalité est pris en compte dans la cartographie des risques sous Solvabilité 2. Avec la mise en vigueur de cette directive en 2016, les organismes d'assurance voient la légitimité de l'utilisation des tables de mortalité renforcée. Ainsi, l'article 77 de la directive Solvabilité 2 définit la meilleure estimation d'un engagement comme correspondant à « la moyenne pondérée par leur probabilité des flux de trésorerie futurs compte tenu de la valeur temporelle de l'argent estimée sur la base de la courbe des taux sans risque pertinente, soit la valeur actuelle attendue des flux de trésorerie futurs ». Dans ce contexte, il est important en assurance vie d'avoir des tables de mortalité représentant fidèlement la population assurée, notamment aux grands âges, pour les contrats prévoyant des prestations viagères (assurance vie, retraite...). Ne pas prendre cela en compte pourrait amener à un sur-provisionnement ou, au contraire, à un sous-provisionnement soit par excès soit par manque de prudence. Dans les deux cas, cela pourrait nuire à l'organisme assureur. En effet, comme évoqué par SALHI (2018), en cas de sous-provisionnement, l'organisme assureur pourrait rapidement se retrouver en situation d'insolvabilité. Il ne pourrait donc plus faire face à ses engagements. En cas de sur-provisionnement, l'organisme assureur aurait moins de fonds propres pour couvrir les exigences en capital (SCR¹) de Solvabilité 2, ce qui peut également rapidement devenir problématique. SALHI (2018) précise également que cela conduirait à une « augmentation de l'assiette servant au calcul du besoin en capital du risque décès » et donc, à une augmentation de ce besoin en capital.

1.1.2 IFRS 17

La norme IFRS 17 accorde également de l'importance aux tables de mortalité. En effet, cette norme précise que « l'ajustement pour risque non financier, Risk Adjustment, représente une prime de risque permettant de couvrir l'incertitude sur le montant et l'échéancier des flux de trésorerie contingents aux risques non financiers ». Cela signifie pour le risque décès – risque non financier – la prise en compte des probabilités de décès à chaque âge (et éventuellement par génération selon les produits) dans le calcul du Risk Adjustment.

Cependant, même si l'utilisation des tables de mortalité a pris encore plus d'importance avec la directive Solvabilité 2 et la norme IFRS17, celle-ci ainsi que leurs intérêts sont bien plus anciens. Elles permettent aux organismes assureurs, fonds de pension et caisses de retraite qui les utilisent, de pouvoir faire face à leurs engagements avec une très grande probabilité. Cela passe également par un ajustement précis des tables

1. Solvency Capital Requirement.

lorsque la taille de l'échantillon devient restreinte, en particulier aux grands âges.

1.2 Historique

En France, le Code des Assurances impose le choix de la table utilisée par rapport au type de produit. Les tables sont, en moyenne, renouvelées tous les 10 à 15 ans et sont généralement produites sur l'ensemble de la population française. Plusieurs tables réglementaires se sont succédé depuis les années 1960 :

Tables	Période d'utilisation	Population observée	Commentaires
PM 60-64	Avant 1984	Hommes	Utilisée pour les garanties en cas de décès.
PF 60-64		Femmes	Utilisée pour les garanties en cas de vie.
TD 73-77	1984 - 1993	Hommes	Utilisée pour les garanties en cas de décès.
TV 73-77		Femmes	Utilisée pour les garanties en cas de vie.
TD 88-90	1993 - 2006	Hommes	Utilisée pour les garanties en cas de décès.
TV 88-90		Femmes	Utilisée pour les garanties en cas de vie.
TPRV 93		Femmes	Utilisée pour le provisionnement des rentes viagères. Il n'y a pas de distinction entre les hommes et les femmes pour des questions d'égalité. Vivant plus longtemps, par prudence, la population observée est composée de femmes.
TH 00-02	Depuis 2007	Hommes	Utilisée pour les garanties en cas de décès.
TF 00-02		Femmes	Utilisée pour les garanties en cas de vie.
TGH 05		Hommes	Utilisée pour le provisionnement des rentes viagères des hommes.
TGF 05		Femmes	Utilisée pour le provisionnement des rentes viagères des femmes.

FIGURE 1.1 – *Historique des tables réglementaires*

Les tables de mortalité doivent considérer la mortalité de la population actuelle, les tables réglementaires se doivent donc d'évoluer pour prendre en compte la dérive de longévité. Illustrons ces propos par un exemple concret afin de montrer l'importance de ces changements de tables réglementaires et de la dérive de longévité.

Le modèle de Lee-Carter est l'un des plus utilisés pour estimer la mortalité future et, de fait, pour construire des tables de mortalité prospectives. Il dépend d'une série temporelle κ_t qui décrit l'évolution générale de la mortalité (cf. Annexe A). Prenons cette série comme indicateur de la mortalité en France.

Son évolution a été la suivante :

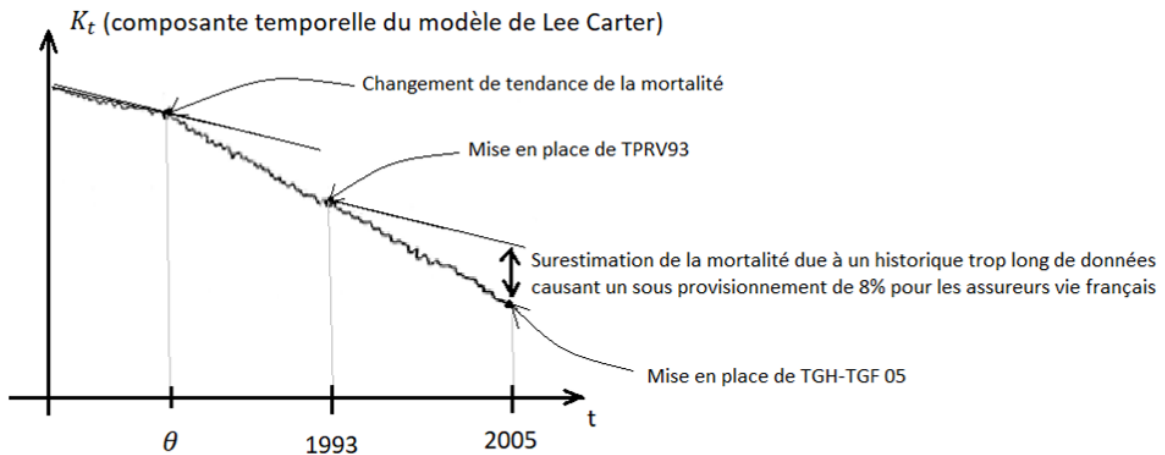


FIGURE 1.2 – *Évolution de la mortalité en France depuis les années 1970 selon le modèle de Lee Carter*

La Figure 1.2 montre qu'autour des années 1970 (au point θ), il y a eu une baisse significative de la mortalité en France. Cependant, en construisant la table TPRV93², l'historique pris allait au-delà des années 1970. Cela a conduit à surestimer la mortalité projetée dans cette table prospective. Le secteur de l'assurance vie en France, utilisant cette table pour le provisionnement des rentes viagères, a donc sous-provisionné ses engagements. Après avoir construit les tables TGH05 et TGF05, les assureurs vie français ont pu estimer l'écart entre le montant à provisionner avec la table TPRV93 et celui obtenu avec les tables TGH05 et TGF05. Cet écart s'élevait, en moyenne, à près de 8%. Cela signifie donc qu'en 2007 - moment où les tables réglementaires ont changé -, les assureurs vie ont dû provisionner un supplément de 8% pour avoir un montant de provisions suffisant. De plus, la table TPRV93 ayant été construite sur une population de femmes, elle aurait dû être plus prudente que les nouvelles tables réglementaires ventilées par sexe. La dérive de longévité aurait donc eu un impact encore plus grand sur le provisionnement, si elle n'avait pas été atténuée par cette ventilation hommes/femmes des nouvelles tables.

Cet exemple met en évidence l'importance de trouver des méthodes efficaces de détection de changements de tendance afin de pouvoir les traiter dès que possible, voire, dans le cas idéal, les anticiper. Avec ce type de méthode, les assureurs auraient donc pu se rendre compte du sous-provisionnement avant que le différentiel ne prenne d'aussi grandes proportions. Pour plus de détails sur cette problématique, se référer à EL KAROUI *et al.* (2015) par exemple.

2. Table Prospective pour Rentes Viagères.

Cet exemple illustre parfaitement l'importance d'utiliser des tables de mortalité en phase avec la réalité et qui représentent la population assurée. Il montre également la nécessité de bien connaître cette population, ainsi que les conséquences que peuvent avoir une mauvaise connaissance du risque de longévité en assurance vie et, plus particulièrement, pour les organismes proposant des rentes viagères. Cette connaissance est d'autant plus importante dans une période où les taux d'intérêt restent bas, ce qui ne permet pas de limiter cette dérive de longévité sur le montant des provisions. Il est alors essentiel d'analyser ce risque.

Chapitre 2

Risque de longévité

La notion de risque abordée ici est la notion de risque ERM¹. Elle se définit comme l'incertitude qu'a une entreprise sur ses objectifs du fait des risques qu'elle doit gérer. D'autres approches de la notion de risque existent, comme la notion de risque de Knight² par exemple, mais ne seront pas abordées ici. La notion de risque ERM se veut plus prudente que les autres car elle considère tous les changements possibles, même l'incertitude liée aux risques que l'on ne sait pas mesurer.

Le risque de longévité, risque long terme, nécessite un accroissement du capital provisionné, phénomène amplifié par Solvabilité II, et IFRS 17 dans une moindre mesure. Il peut alors être intéressant pour les assureurs, caisses de retraite et fonds de pension de transférer ce risque, tout ou en partie.

2.1 Généralités

En 2010, le CRO³ Forum a mesuré l'exposition (directe) au risque de longévité dans les pays de l'OCDE⁴. Celle-ci s'élevait alors à plus de 16 milliards de dollars américain, ce qui représentait, avec le taux de change de l'époque, environ 11.65 milliards d'euros. Le risque de longévité ayant pris encore plus d'ampleur du fait, entre autres, de l'évolution des encours des contrats d'assurance vie en France, il est tout à fait raisonnable de penser que l'exposition à ce risque est, en 2021, plus importante.

1. Enterprise Risk Management.

2. Cette approche fait la distinction entre les risques dont les probabilités de réalisation des événements peuvent être mesurées, et l'incertitude pour laquelle il nous ne savons pas déterminer une probabilité d'occurrence.

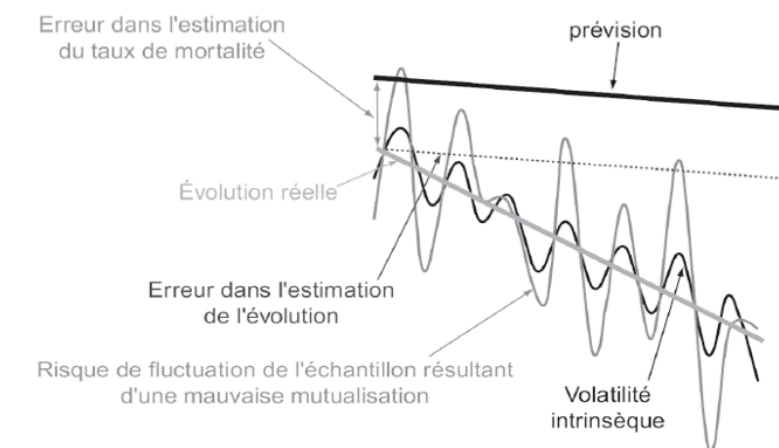
3. Chief Risk Officer. Le CRO Forum est un groupe spécialisé dans la gestion de risque en assurance, qui encourage les bonnes pratiques dans ce domaine.

4. Organisation de Coopération et de Développement Économique.

Le risque de longévité est un risque dont la modélisation et l'évaluation financière à long terme sont particulièrement difficiles. Cela est notamment dû – comme pour le risque CatNat – à la grande incertitude qui l'entoure et à sa sensibilité à certains phénomènes (pandémie, canicule, pics de pollution...) de plus en plus fréquents et difficilement maîtrisables. Ce risque est un enjeu majeur, en particulier pour les retraites et autres risques viagers en période de taux bas qui n'atténuent plus le montant des provisions mathématiques, comme cela a pu être le cas dans le passé, où les taux d'intérêt étaient plus élevés.

Pour mieux l'appréhender, il est utile de le diviser en plusieurs composantes : le risque biométrique⁵ et les autres risques qui en découlent comme le risque de taux ou de contrepartie.

Le risque biométrique se décompose de la manière suivante :



Source : Boumezoued (2016).

FIGURE 2.1 – Les différentes composantes du risque biométriques de longévité et leurs effets sur le logarithme du taux de mortalité en fonction du temps

- Le risque de niveau : risque d'estimation du taux de mortalité actuel.
- Le risque de tendance : risque d'estimation de la longévité future.
- Le risque d'oscillation saisonnière : risque d'estimation dû à la période d'étude, certaines étant plus propices aux décès que d'autres.
- Le risque d'échantillonnage : risque que l'échantillon observé n'est pas représentatif de la population étudiée.

Il peut alors arriver que les projections de l'évolution du risque de longévité comportent une erreur globale assez importante, comme cela a pu être le cas avec la table TPRV93 (cf. Chapitre 1.2).

5. Il s'agit du risque que la population vive, en moyenne, plus longtemps que prévu, sans prendre en compte les risques connexes.

Il est complexe d'estimer le risque de tendance du risque biométrique puisque les experts ne s'accordent déjà pas entre eux. D'une part, certains défendent l'hypothèse d'un âge limite biologique que l'Homme ne pourrait dépasser, prédisant ainsi une augmentation de l'effet de « rectangularisation » des courbes de survie. Les décès interviendraient alors, à terme, presque tous aux alentours de l'âge biologique limite. Un âge limite autour de 125 ans ou 130 ans est aujourd'hui crédible, même si cette approche paraît fragilisée. En effet, par le passé, de nombreux scientifiques ont défendu cette hypothèse en estimant qu'il était impossible de vivre au-delà de 100 ans, puis de 120 ans, ce qui est manifestement faux. D'autre part, certains démographes comme James Vaupel, considèrent que les progrès médicaux et l'évolution du niveau de vie vont permettre de poursuivre l'évolution croissante et, (presque) linéaire, de l'espérance de vie dans les pays les plus développés. Enfin, les adhérents du mouvement transhumaniste, comme Aubrey de Grey, pensent qu'il sera bientôt possible de vivre 500 ans (voire plus), et que le ralentissement du vieillissement dû, entre autres, aux avancées de la médecine, avec des techniques de réparation cellulaire, de prothèses bioniques, de thérapie génique ou autres évolutions médicales vues au cinéma, feront partie des grandes évolutions de la médecine dans un futur assez proche.

En actuariat, les hypothèses de mortalité futures généralement utilisées produisent un ralentissement de l'augmentation de l'espérance de vie, et sont donc en contradiction avec les idées de James Vaupel. Cela est en particulier développé dans DEBONNEUIL *et al.* (2015) où les auteurs donnent des explications mathématiques quant à ce phénomène. Il est aussi expliqué que, du fait des impacts des hypothèses non-décéléatrices de l'espérance de vie, les gouvernements pourraient les rejeter, mettant ainsi en danger les assureurs vie et caisses de retraite (dans un futur plus ou moins proche), qui seraient amenés à surestimer la mortalité des assurés, ce qui aurait, entre autres, pour conséquence un sous-provisionnement des rentes viagères.

Le risque biométrique a une importance capitale, relayant les risques de taux, d'inflation et de contrepartie au second plan. Il ne faut cependant pas les négliger car leurs prises en compte peut être très différente d'un portefeuille d'assurés à l'autre, et donc, avoir des conséquences plus ou moins importantes. En effet, la prise en compte de ces risques dépend, en particulier, de la politique de gestion de risques prévue par l'organisme assureur.

2.2 Facteurs de risque

Le risque de longévité se mesure sur deux niveaux :

- Au niveau d'une population nationale.
- Au niveau des individus.

Comme pour tous risques, certains facteurs favorisent la longévité, et d'autres, la réduisent. Dans le cas du risque de longévité, ces facteurs s'appréhendent principalement au niveau individuel, mais certains le sont également au niveau national. Il faut donc privilégier une approche mêlant ces deux niveaux.

De nombreux facteurs jouent un rôle déterminant sur la longévité au niveau individuel :

- Le sexe : nous savons que les femmes vivent, en général, plus longtemps que les hommes. Il s'agit d'une cause essentiellement biologique.
- Le niveau de richesse : les personnes riches vivent généralement plus longtemps que les personnes financièrement plus modestes.⁶
- Le style de vie : hygiène, nutrition, consommation d'alcool et de drogues, fumeur ou non-fumeur.
- L'environnement : pollution, climat.
- La facilité d'accès aux soins.
- Les interactions sociales et intellectuelles

De plus, les personnes vivant seules ont une espérance de vie réduite par rapport aux autres du fait de l'assistance mutuelle que peuvent se porter deux personnes au sein d'un même foyer. Le statut marital est donc un autre facteur de risque individuel important.

Certaines des notions jouant un rôle dans l'évolution de la longévité ont évolué dans le temps, du fait d'un contexte social différent. C'est, par exemple, le cas du niveau d'études ou de la situation maritale qui n'ont plus la même signification que par le passé. Afin de pouvoir évaluer l'impact d'un de ces facteurs sur ce risque, il convient alors de définir des classes par niveaux fixes, ou d'avoir un indicateur, comme un quantile, qui prend en compte l'évolution de ces facteurs en fonction du temps.

De nouveaux facteurs de risque peuvent également apparaître. On peut notamment penser à des risques de terrorisme de masse, des risques écologiques, auxquels nous étions moins confronté par le passé, ou d'autres risques qui n'ont pas encore émergés, et, qu'il est même difficile d'imaginer. Ces facteurs sont essentiellement à appréhender au niveau national.

6. Cela est vrai jusqu'à un certain niveau de richesse au delà duquel l'espérance de vie se réduit.

Les changements dans les politiques publiques pourront aussi avoir des conséquences non-négligeables sur l'évolution de la longévité des populations. On retient par exemple campagnes de prévention contre des maladies (cancer, MST, diabète. . .) invitant la population à se faire dépister, ainsi que les campagnes de promotion pour une meilleur hygiène de vie : prévention contre le tabac et l'alcool ; l'encouragement pour une pratique sportive ; pour une alimentation saine ; les campagnes de sécurité routière ou autre. Cependant les politiques publiques ne favorisent pas toujours un accroissement de la longévité, c'est par exemple le cas des politiques publiques favorisant l'usage d'opioïdes aux États-Unis.

Certaines questions éthiques peuvent également se poser.

Du point de vue de l'assurance, ces évolutions, notamment dans le domaine de la santé, posent des questions par rapport à l'asymétrie d'informations entre l'assuré et l'assureur. En effet, l'assuré aurait connaissance de son état de santé, tandis que le respect de la vie privée empêcherait l'assureur d'en prendre connaissance. Il est alors légitime de se demander si, dans ce contexte, et compte tenu d'un aléa réduit induit par les progrès de la médecine, le risque de longévité resterait assurable (cf. EL KAROUI et LOISEL (2017)).

Afin de résumer les différents facteurs de risques⁷, un tableau récapitulatif (non-exhaustif) peut aider à mieux comprendre.

	Augmentation de la longévité	Diminution de la longévité
Style de vie	<ul style="list-style-type: none"> - Moins de fumeurs - Meilleure alimentation - Pratique sportive régulière - Environnement 	<ul style="list-style-type: none"> - Stress accru - Mauvaise alimentation
Médecine	<ul style="list-style-type: none"> - Nouvelles techniques ou nouveaux antibiotiques - Prévention accrue des maladies 	<ul style="list-style-type: none"> - Résistance des virus et bactéries aux antibiotiques - Nouvelles maladies
Autre	<ul style="list-style-type: none"> - Politique publique 	<ul style="list-style-type: none"> - Terrorisme de masse - Politique publique

FIGURE 2.2 – *Facteurs du risque de longévité pour les années à venir*

Face à ce risque qui prend de plus en plus d'importance, il peut être intéressant de le transférer – ou du moins une partie – vers des réassureurs, ou vers les marchés financiers.

7. Les sources sont en particulier le CRO Forum et EL KAROUI et LOISEL (2017) assez riches sur le sujet.

2.3 Couverture du risque

Il existe plusieurs façons de se couvrir face aux risques. Il est par exemple possible de transférer le risque vers une autre entité ou vers les marchés financiers. Les trois façons principales de se couvrir face au risque de longévité sont les suivantes :

- La réassurance.
- Le swap de longévité.
- La titrisation.

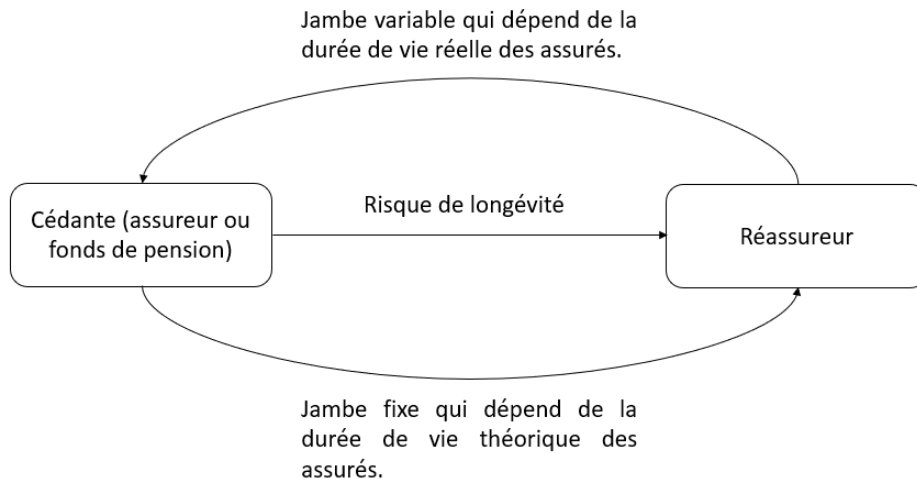
2.3.1 Réassurance

La réassurance consiste à transférer une partie du risque vers un réassureur, en contrepartie d'une prime. Du fait de sa simplicité, il s'agit de la forme de couverture la plus utilisée en France. Il existe deux types de réassurance : le Buy-In et le Buy-Out. Dans le premier cas, il s'agit d'un simple échange de flux de trésorerie. L'assureur reste responsable du versement des prestations auprès des assurés. Le Buy-Out, quant à lui, est similaire à un rachat du portefeuille. Il permet à l'assureur, contrairement au Buy-In, de ne plus porter le risque qui est transféré vers le réassureur, qui devient alors responsable du versement des prestations.

2.3.2 Swap de longévité

Le swap de longévité est assez peu utilisé en France où les assureurs préfèrent un produit de réassurance mais il est assez populaire dans les pays anglo-saxon.

Il s'agit de paiements périodiques fixes, versés à la partie acceptant le swap de longévité en échange d'un paiement périodique fondée sur la différence entre la mortalité attendue et la mortalité réelle du portefeuille. Un swap est composé de deux jambes : une jambe fixe et une jambe variable. La cédante paie la jambe fixe à l'organisme réassureur et celui-ci paie les prestations de la cédante, qui représentent la jambe variable.

FIGURE 2.3 – *Schéma simplifié d'un swap de longévité*

L'organisme réassureur a tout intérêt à construire ses propres tables d'expérience car la durée de vie théorique sera plus représentative de la population assurée. Le fait d'utiliser une table nationale pourrait biaiser la durée de vie théorique ce qui aurait pour effet que la transaction ne serait plus équitable.

En cas de forte dérive de la mortalité prévue par l'organisme assureur sur son portefeuille, un swap de longévité est une option intéressante à envisager.

2.3.3 Titrisation

La titrisation est une opération consistant à transformer des actifs peu liquides en des valeurs mobilières. Le risque est alors porté par les investisseurs, qui pourront diversifier leurs portefeuilles avec d'autres risques non-corrélés. La démocratisation de cette méthode de couverture pourrait permettre de transférer le risque de longévité vers les marchés financiers, et ainsi le rendre plus liquide. Les investisseurs y sont cependant assez réticents, du fait de l'asymétrie d'informations avec la cédante qui connaît mieux la population assurée qu'eux. Cette méthode a été fragilisée par la crise des « subprimes » en 2008 au cours de laquelle la titrisation a joué un rôle important dans la chute des marchés financiers.

La titrisation est très peu utilisée en France, ce qui est moins le cas à l'international. Elle a, par exemple, été utilisée par Swiss Re, avec notamment l'exemple de Kortis Capital⁸.

8. Cf. EL KAROUÏ et LOISEL (2017) pour plus de détails sur cette opération

Le risque de longévité est un risque essentiel à prendre en compte lorsque l'on étudie les tables de mortalité, ou lors de l'étude de garanties viagères. Une bonne projection de la mortalité permet aux assureurs d'estimer leurs engagements de manière plus précise et plus rigoureuse. Cependant il existe de nombreuses divergences d'avis parmi les experts. C'est pourquoi il peut être intéressant de se couvrir par rapport à ce risque.

Il nous faut à présent introduire les données de mortalité sur lesquels nous allons travailler.

Chapitre 3

Base de données

Pour réaliser cette étude, les données utilisées sont des données issues de la Human Mortality Database. Il s'agit d'une base de données qui fournit les données de mortalité de nombreux pays. Elle recense, en particulier, des estimations de l'exposition, du nombre de décès et du nombre de naissances qui ont lieu dans ces pays.

Une base de données assurantielle sera également utilisée et introduite dans le dernier chapitre de ce mémoire (cf. Chapitre 7.2). Elle servira à l'application des résultats obtenus.

3.1 Présentation des données

Les données utilisées sont déjà, en partie, traitées. Les traitements déjà réalisés sont expliqués dans WILMOTH *et al.* (2021). Ils diffèrent selon le pays, par exemple, les traitements effectués sur les données du Japon ne sont pas les mêmes que ceux réalisés sur les données françaises. Il est donc nécessaire de faire preuve de vigilance, dans le cas où l'on veut étudier et comparer les données de différents pays à l'aide de cette base. Dans notre cas, nous nous concentrerons uniquement sur les données relatives à la France.

Les données recueillies recensent des estimations du nombre de décès dans la population française (civile et militaire) depuis 1816 ainsi qu'une estimation de la population et, de fait, de l'exposition. Elles sont ventilées par sexes, âges et cohortes. Cela nous permettra donc de construire une table générationnelle, utile pour l'évaluation du montant de provisionnement des rentes viagères.

Les données ne sont pas disponibles pour les âges supérieurs à 110 ans (inclus). Cela est dû au faible nombre de données à ces âges. Il s'agit d'une des limites des données à prendre en compte, étant donné notre sujet.

L'article de l'Institut des Actuaire DEVINEAU *et al.* (2017), permet de comprendre que la Human Mortality Database comporte des anomalies pour certaines générations (1919 et 1920 par exemple). Ce même article démontre, néanmoins, que la calibration des taux de mortalité via des méthodes de projection de tables est peu affectée par cela. Nous ne prendrons donc pas ces anomalies en compte, ni ne tenterons de corriger les données. Les résultats obtenus pourront alors contenir un léger biais que nous considérerons comme négligeable vu le faible nombre de cohortes pour lesquelles ce cas se présente et les conclusions de cet article.

3.2 Traitements

Les données pour les âges supérieurs à 110 ans sont disponibles sur le site « [Les centenaires français](#) ». Il s'agit d'un site internet qui recense, entre autres, la mortalité et le nombre des femmes vivantes de plus de 109 ans et des hommes vivants de plus de 107 ans en France. La Human Mortality Database fournissant certaines de ces données, pour plus de stabilité, seule la partie des âges supérieurs à 110 ans est exploitée à partir de ce site internet.

Une assez bonne cohérence entre les deux sources de données peut être constatée. Il faut cependant faire preuve de vigilance par rapport à cette deuxième source, étant donné qu'il s'agit d'un forum, et non d'une source officielle ou qui travaille avec des institutions officielles. Des tests de sensibilité ont donc été mis en place dans la section 3.4.2. Ils permettront de mesurer les conséquences d'une erreur au niveau de ces données sur les taux de mortalité obtenus.

3.3 Statistiques descriptives

Afin de mieux connaître les caractéristiques des données utilisées, il est important de commencer par une étude statistique. En effet, cela peut permettre, d'orienter le choix des modèles par la suite et d'analyser si le portefeuille de rentes utilisé a les mêmes caractéristiques que la population générale, ou si possible de mesurer le décalage entre les deux populations, et d'en déduire son importance.

Pour cela, nous avons plusieurs indicateurs, comme la répartition femmes/hommes pour l'exposition, ainsi que pour la population décédée. Il est également possible de ventiler les données par âge. Nous pouvons, de plus, nous intéresser aux taux de mortalité, ce qui nous permettra de constater la dérive de longévité au fil des générations ainsi que la forte volatilité aux grands âges. Il est aussi pertinent de tracer les taux de mortalité en échelle logarithmique, afin de mettre en évidence leurs fortes variations

aux grands âges.

Pour cette étude, le portefeuille de rentes retenu est seulement composé de personnes âgées de plus de 80 ans. Les graphiques suivants se concentreront uniquement sur ces âges. Pour limiter l'erreur de tendance on se focalisera seulement sur les années postérieures à 1990.

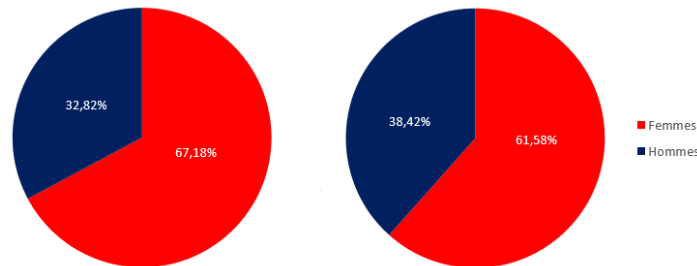


FIGURE 3.1 – Répartition hommes/femmes de l'exposition (à gauche) et des décès (à droite) pour les personnes de plus de 80 ans depuis les années 1990

Sur cette tranche d'âge, la population est largement constituée de femmes. De plus, il y a une surreprésentation des décès pour les hommes étant donné qu'en représentant 32,82%, ils constituent plus de 38% des décès. Cela reflète, entre autres, la longévité accrue des femmes par rapport aux hommes.

S'intéresser à la répartition des âges dans les bases de données est aussi une bonne manière de se représenter les données.

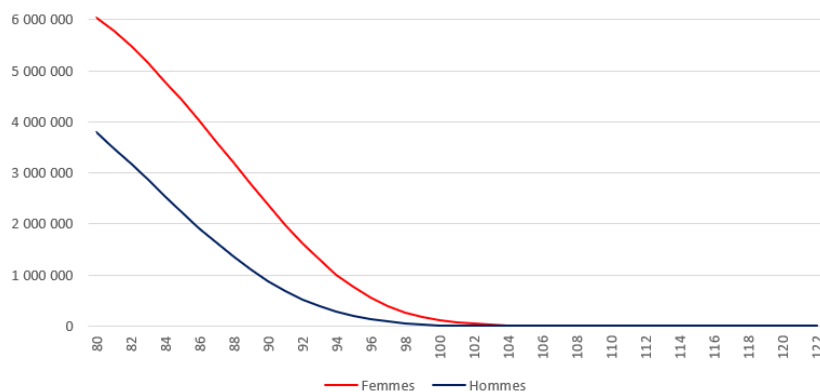


FIGURE 3.2 – Répartition de l'exposition par âge depuis les années 1990

Comme nous pouvions nous y attendre, l'exposition baisse rapidement jusqu'à 90 ans pour les hommes, et 95 ans pour les femmes. Au delà, cette baisse est énormément ralentie. Cela est notamment dû au faible nombre de survivants à ces âges. Cette chute semble légèrement plus forte chez les femmes et ralentit autour de 95 ans. L'exposition

devient nulle chez les hommes dès 112 ans tandis qu'elle devient nulle à partir de 123 ans pour les femmes.

Le nombre de décès se décompose de la manière suivante :

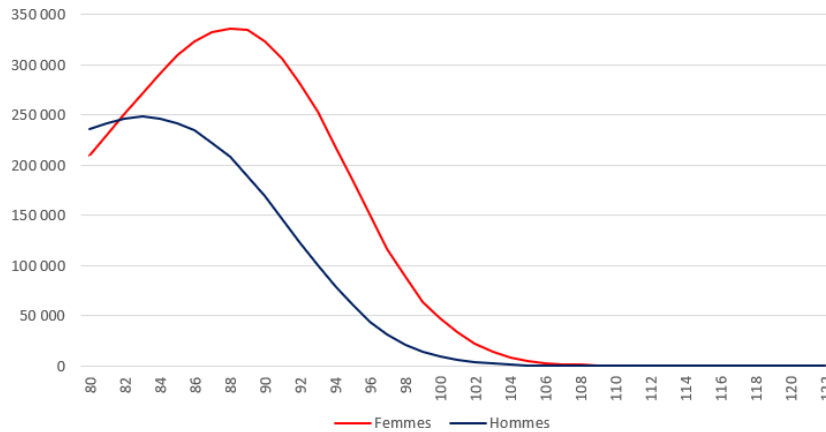


FIGURE 3.3 – Répartition des décès par âge depuis les années 1990

Nous constatons que le nombre de décès est, pour les femmes, en hausse jusqu'à 89 ans puis décroît, tandis que pour les hommes celui-ci décroît dès 83 ans. Les décroissances du nombre de décès sont notamment dues à une baisse de l'exposition. Le dernier décès a lieu à 111 ans chez les hommes et 122 ans chez les femmes.

Ces graphiques sont intéressants mais ne permettent pas de rendre compte du nombre de décès par rapport à l'exposition. Pour cela, il faut tracer les taux bruts de mortalité. Toujours pour les années calendaires postérieures à 1990 les résultats sont les suivants :

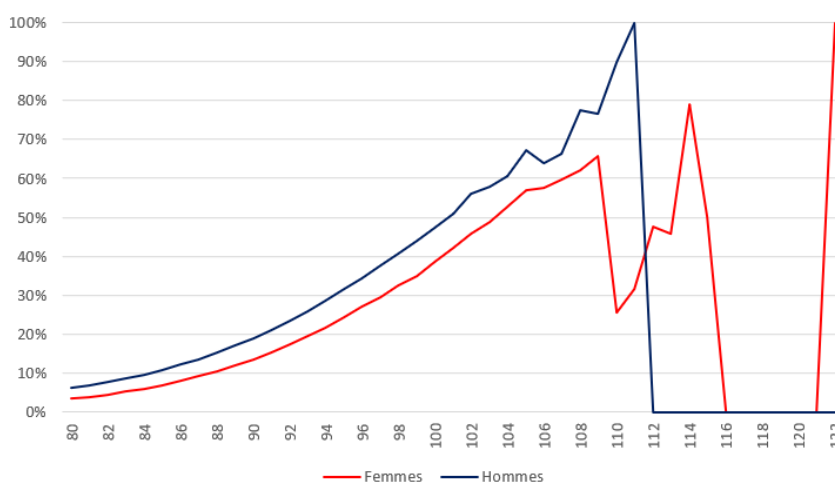


FIGURE 3.4 – Taux bruts de mortalité par âge et par sexe toutes cohortes confondues

Les taux bruts de mortalité montrent une très forte croissance après 80 ans ainsi

qu'une grande volatilité aux âges supérieurs à 100 ans, quelque soit le sexe, due à la faible taille des échantillons disponibles. La Figure ci-dessus présentant les taux de mortalité toutes générations confondues, la volatilité à ces âges est réduite par rapport à la volatilité que l'on aurait obtenu génération par génération du fait de la réunion des cohortes.

Il est intéressant de tracer cette courbe en échelle logarithmique.



FIGURE 3.5 – *Taux bruts de mortalité par âge et par sexe, en échelle logarithmique, toutes cohortes confondues*

Cette Figure montre une légère concavité pour les âges inférieurs à 100 ans impliquant, entre autres, que pour ces âges l'évolution n'est pas exponentielle contrairement à l'évolution du taux de mortalité pour les âges inférieurs à 80 ans. Cela va dans le sens d'un ajustement à ces âges étant donné que les modèles de mortalité les plus classiques – les modèles de Lee-Carter et de Gompertz-Makeham – proposent ce type d'évolution. Sur le graphe correspondant, nous remarquons également l'instabilité aux grands âges.

Il est également intéressant d'étudier la mortalité en fonction de la génération. En effet, nous avons vu au Chapitre 2 qu'il y avait une baisse de la mortalité au fil des générations à âges fixés. Celle-ci se retrouve dans les données traitées :

3.3. STATISTIQUES DESCRIPTIVES

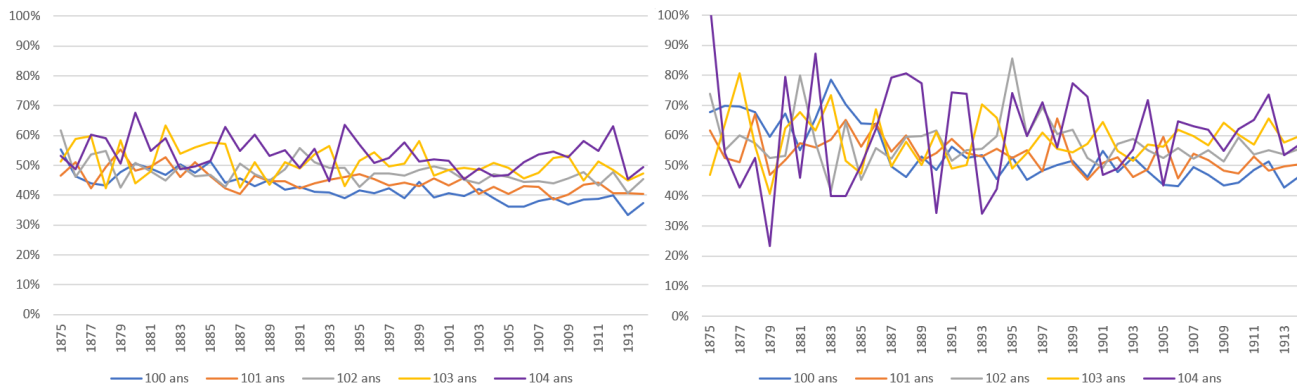


FIGURE 3.6 – *Évolution de l'exposition à âge fixe en fonction de la cohorte pour les femmes (à gauche), et pour les hommes (à droite)*

Une légère baisse, en moyenne, du taux de mortalité au fil des générations est à constater aussi bien chez les hommes que chez les femmes. En effet, le taux de mortalité pour les âges donnés est, pour les femmes, autour de 50% pour la génération 1875 contre 45% pour la génération 1914. Pour les hommes, le taux de mortalité passe de 60% pour la génération 1875 à 55% pour la génération 1914. La volatilité du taux de mortalité en fonction des générations est bien plus élevée chez les hommes que chez les femmes. Cela est probablement dû aux plus faibles effectifs à ces âges – que l'on peut qualifier de plus extrêmes – pour les hommes. La dérive de la longévité semble donc assez similaire selon le sexe, la baisse moyenne du taux de mortalité étant identique sur la même période.

Il est également possible de visualiser les taux de mortalité en 3 dimensions. Cela permet de voir, à la fois la dérive liée à la génération et l'évolution de la mortalité en fonction de l'âge.

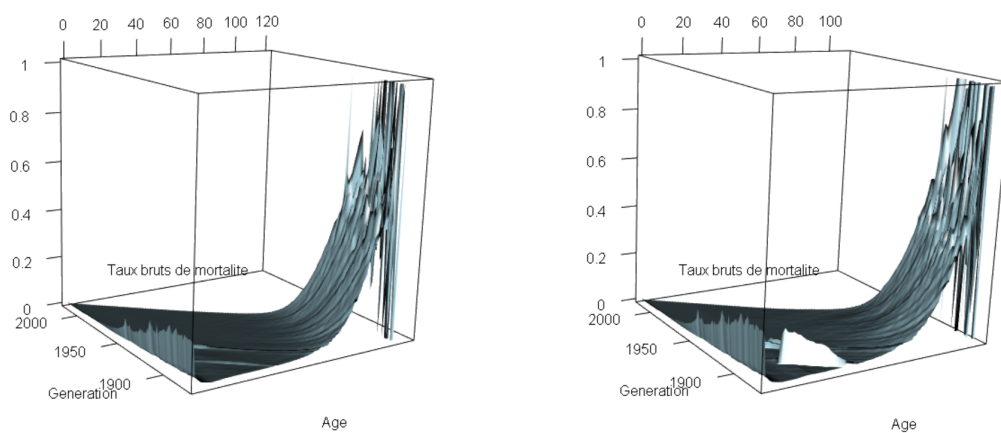


FIGURE 3.7 – *Représentation en 3 dimensions des taux bruts de mortalité des femmes (à gauche) et des hommes (à droite)*

Dans la Figure ci-dessus les taux de mortalité sont donnés à tout âge. Celle-ci montre les anomalies annoncées dans la section 3.1. En effet, pour les générations 1919 et 1920 il est possible de constater une forte hausse de la mortalité, notamment pour la population masculine. Nous constatons également des hausses de mortalité, principalement chez les hommes, autour des deux Guerres Mondiales. Nous remarquons enfin une baisse très importante de la mortalité infantile passant d'un taux de mortalité d'environ 15% pour les générations autour de 1900 à un taux de mortalité proche de 0% aujourd'hui¹.

La mortalité montre bien une évolution exponentielle pour l'ensemble des générations pour les âges allant de 30 à 75 ans. Aux grands âges, elle se révèle être très volatile peu importe la génération et le sexe. La forte volatilité commence légèrement plus jeune chez les hommes que chez les femmes. Ce phénomène est dû à une longévité accrue des femmes par rapport aux hommes.

Afin de mettre en évidence la forme exponentielle ainsi que les variations des taux bruts de mortalité, nous utilisons une échelle logarithmique.

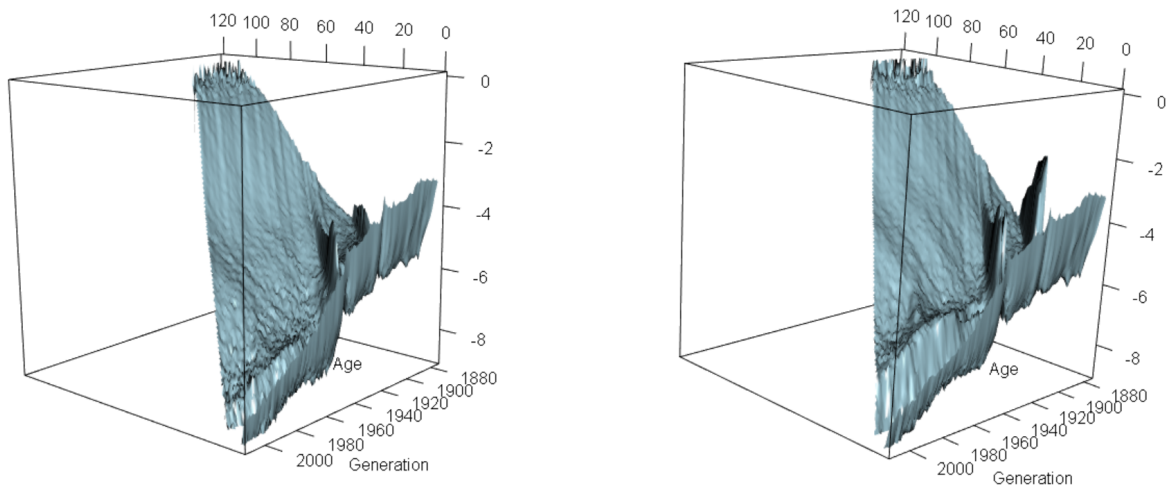


FIGURE 3.8 – Représentation en 3 dimensions des taux bruts de mortalité des femmes (à gauche) et des hommes (à droite) en échelle logarithmique

L'échelle logarithmique nous permet de mieux observer la baisse du taux de mortalité au fil des générations, phénomène moins marquée la Figure 3.7. En moyenne, la mortalité évolue linéairement de 30 ans jusqu'à 75 ans pour les femmes et 70 ans pour les hommes. Les taux de mortalité sont par la suite légèrement concaves, puis assez volatils.

1. Selon l'UNICEF, cette évolution est due à l'action menée par les gouvernements qui ont placé cette baisse comme l'une des priorités politiques. De manière plus concrète, l'UNICEF met principalement en avant les mesures suivantes : les soins prénatals, obstétricaux et postnatals ; la vaccination ; les antibiotiques.

3.4 Qualité des données

3.4.1 Généralités

À l'époque du big data et d'une quantité de plus en plus importante de données disponibles, la qualité des données est devenue un enjeu majeur. Une bonne qualité permet de pouvoir implémenter un modèle cohérent et réaliste, ce qui n'est pas forcément le cas avec des données de moindre qualité. C'est d'ailleurs une problématique abordée dans la directive Solvabilité 2, qui insiste fortement dessus. Elle met l'accent sur trois critères auxquels doivent répondre les données :

- L'exhaustivité : la profondeur d'historique doit être suffisante.
- L'exactitude : les données utilisées doivent être exactes et cohérentes.
- Le caractère approprié : les données utilisées doivent être utiles pour les calculs.

La réglementation aborde également, dans une moindre mesure, d'autres critères, comme la traçabilité, la documentation des limites...

Les données utilisées répondent à ces critères. En effet, l'historique de données est très important et celles-ci paraissent, de plus, cohérentes dans leur ensemble². Les données sont tracées et les méthodes de traitements ont été documentées par la Human Mortality Database.

La Human Mortality Database travaille pour la récolte des données françaises, avec des organismes comme l'INSEE et l'INED³. Le nombre de décès et l'exposition sont cependant des estimations, et peuvent donc présenter des écarts avec la réalité.

Nous considérons ainsi que les données utilisées sont donc de qualité, malgré quelques anomalies.

Pour les âges supérieurs à 110 ans, pour lesquels les données sont issues d'un forum, il est utile d'approfondir l'étude de la qualité des données en se posant des questions sur leur fiabilité. Pour cela il est utile de mettre en place des tests de sensibilité.

3.4.2 Tests de sensibilité au nombre de décès pour les supercentenaires

Le site des supercentenaires étant un forum et non une entité officielle, cela nous pousse à une rigueur accrue quant à la fiabilité de ces données. En effet, il est tout à fait possible que certains supercentenaires aient été oubliés et, dans ce cas, ne sont pas recensés dans cette base. Il est alors utile de faire des tests de sensibilité sur le nombre

2. Malgré quelques anomalies qui ne porteront pas à conséquences.

3. Institut National d'Études Démographiques.

de décès (et/ou le nombre de survivants) à ces âges, afin de rendre compte de l'impact que cela peut avoir sur les taux bruts de mortalité.

Nous n'allons pas illustrer, ni tester la sensibilité au nombre de décès des supercentenaires pour chaque génération. Nous la testerons sur quelques générations pour les femmes ainsi que pour les hommes. Nous généraliserons par la suite les conclusions à toutes les générations.

Les premiers tests effectués portent sur le nombre de décès, et sont les suivants :

Génération	1900	1901	1902	1903	1904	1905
Âge	110 ans	110 ans	110 ans	110 ans	110 ans	110 ans
Sensibilité	+1	+1	+1	+1	+1	+1

FIGURE 3.9 – Tests de sensibilité n°1 aux décès des supercentenaires

Par souci de lisibilité des graphiques, les résultats présentés ci-dessous illustrent seulement les sensibilités pour les générations 1900 et 1905 :

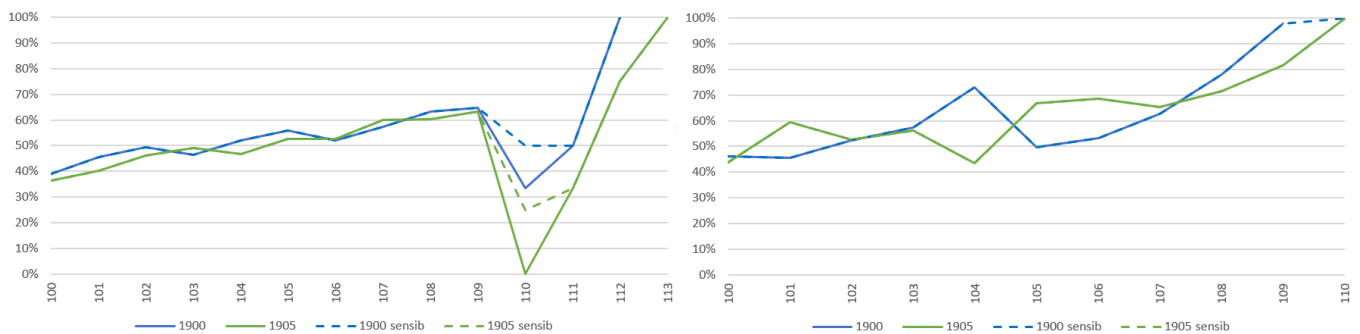


FIGURE 3.10 – Test de sensibilité n°1 sur le nombre de décès pour les supercentenaires femmes (à droite) et hommes (à gauche)

Pour les hommes, ces tests ne semblent pas avoir de grandes conséquences. Cela prolonge seulement le taux de mortalité de 96% à 109 ans à 100% à 110 ans pour la génération 1900, ce qui n'est pas vraiment significatif et n'aura pas un grand impact pour les futurs calculs de provisions. Les effets de ce test sont mêmes nuls pour la génération 1905, étant donné que le taux de mortalité à 110 ans était déjà de 100% ; rajouter un décès n'aurait donc pas vraiment de sens.

Les tests de sensibilité ont un impact plus important pour les femmes. En effet, le taux de mortalité à 110 ans passe, par exemple, de 0% à 25% pour la génération 1905 en ajoutant un seul décès. Cela est dû au fait que l'exposition est de 4 personnes seulement. Pour la génération 1900, le taux de mortalité augmente d'environ 15%. Ces variations ne sont pas négligeables et posent de réelles questions sur ces données.

Afin d'approfondir ces tests il est également utile de tester la sensibilité du taux au nombre de décès sur un autre âge que 110 ans.

Réalisons les tests suivants :

Génération	1900	1901	1902	1903	1904	1905
Âge	112 ans	112 ans	112 ans	112 ans	112 ans	112 ans
Sensibilité	+1	+1	+1	+1	+1	+1

FIGURE 3.11 – Tests de sensibilité n°2 aux décès des supercentenaires

Ces tests ne porteront que sur les femmes, étant donné qu'aucun homme n'a survécu plus de 111 ans. Faire ce test sur la population masculine n'aurait donc pas de sens.

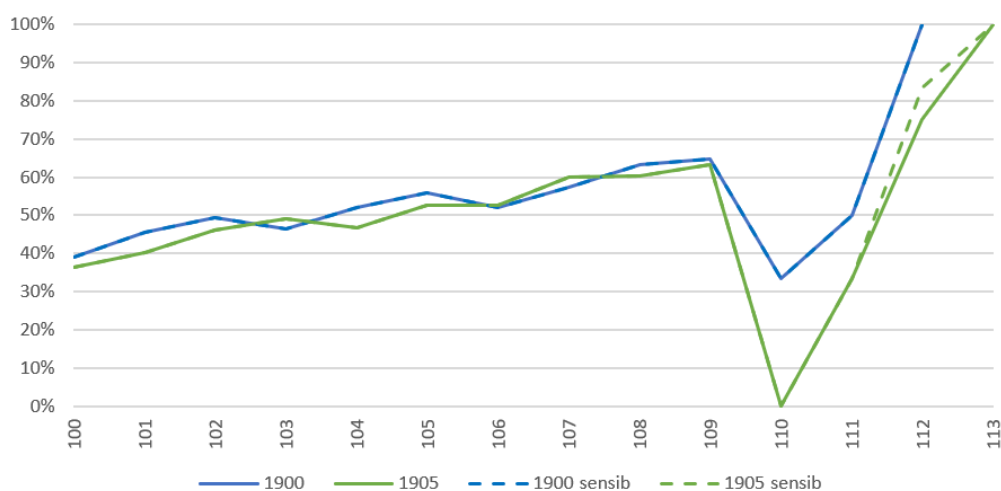


FIGURE 3.12 – Test de sensibilité n°2 sur le nombre de décès pour les supercentenaires femmes

Ce deuxième test permet de nous rassurer sur les données. En effet, ce test n'a pas de conséquences sur la génération 1900, dans laquelle toutes les femmes sont décédées avant 112 ans. L'impact sur le taux de mortalité pour la génération 1905 est, lui aussi, assez restreint par rapport au test précédent. Il n'est pas négligeable pour autant, étant donné que pour cette génération, celui-ci conduit à une hausse de 10% du taux de mortalité à 112 ans.

Des tests de sensibilité qui ont pour but d'augmenter l'exposition peuvent également être mis en place. Les conséquences de ceux-ci sont contraires à celles des deux précédents tests. En augmentant l'exposition, les taux bruts de mortalité baissent. Cela n'est cependant pas très significatif, et les provisions constituées par la suite sur le portefeuille de rentes seraient légèrement plus prudentes.

Maintenant que nous avons introduit, analysé les données et conclu sur leur bonne qualité dans l'ensemble, il faut s'intéresser aux taux bruts de mortalité, notamment afin de construire des intervalles de confiance. Cela permettra, entre autres, de déterminer un âge à partir duquel, en moyenne, l'incertitude liée à la faible taille de l'échantillon devient importante. C'est à partir de ces âges là que l'on pourra juger utile d'ajuster la table.

Chapitre 4

Construction de la table

Les tables de mortalité utilisées pour les risques viagers et qui nous intéressent ici, sont, de manière générale, des tables générationnelles. Ce sont des tables prospectives qui font des hypothèses sur la mortalité future, ce qui permet la prise en compte de la dérive de longévité. Pour faire ces projections, il faut travailler sur les taux bruts de mortalité pour obtenir un estimateur réaliste, prenant en compte la censure et la troncature, et offrant une certaine crédibilité statistique aux résultats obtenus.

4.1 Diagramme de Lexis

Le diagramme de Lexis permet d'opposer deux types d'approches : l'approche par période et l'approche par cohorte.

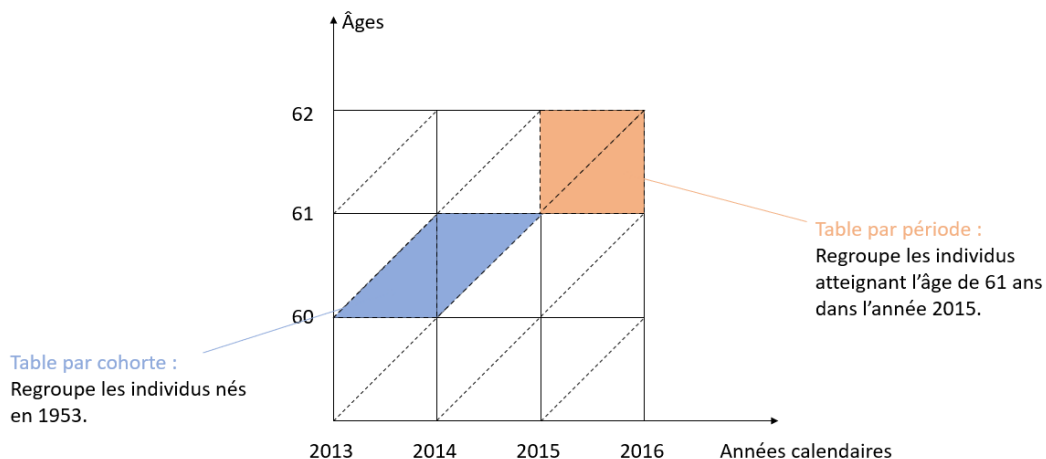


FIGURE 4.1 – *Diagramme de Lexis*

La Figure 4.1 montre qu'une table par période a l'avantage de donner la mortalité sur une année calendaire donnée. Elle permet de voir l'évolution de la mortalité par

rapport aux années calendaires, mais elle mélange des générations différentes. Sur la Figure 4.1, cela se lit à la verticale (en orange).

Une table par génération prend en compte la mortalité par cohorte. L'inconvénient est que les taux de mortalité mélangent deux années calendaires différentes. Elle a cependant l'avantage de voir l'évolution de la mortalité au fil des générations. Sur la Figure 4.1, cela se lit sur la diagonale (en bleu).

Nous rappelons que ce mémoire se place dans une approche par génération.

4.2 Durée de vie et notations actuarielles

En assurance, la durée de vie est perçue comme une variable aléatoire T . Celle-ci représente la durée de vie d'un individu.

La fonction de survie de cet individu, notée S , est définie de la manière suivante :

$$S(t) = \mathbb{P}(T \geq t)$$

Il s'agit de la probabilité pour cette personne de vivre au delà de l'âge t .

Pour un individu d'âge x , on peut modéliser sa durée de vie résiduelle comme étant :

$$T_x \sim T - x | T > x$$

De plus, nous pouvons montrer que, avec $t \in \mathbb{R}^+$:

$$\begin{aligned} {}_t p_x &= \mathbb{P}(T_x \geq t) \\ &= \mathbb{P}(T \geq x + t | T \geq x) \\ &= \frac{\mathbb{P}(T \geq x + t)}{\mathbb{P}(T \geq x)} \quad \text{car } \{T \geq x + t\} \subset \{T \geq x\}. \\ &= \frac{S(x + t)}{S(x)} \end{aligned}$$

La probabilité de vivre $x + t$ années pour un individu déjà âgé de x années est donc le ratio des deux probabilités de survie.

Le taux de décès instantané (ou force de mortalité) est la probabilité de décéder à un âge précis juste après l'âge t , sachant que l'individu est en vie à l'âge t .

Il se définit par :

$$\mu(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(T \leq t + dt | T \geq t)}{dt}$$

Il est alors possible de montrer que μ vaut :

$$\begin{aligned} \mu(t) &= \lim_{dt \rightarrow 0} \frac{\mathbb{P}(T \leq t + dt | T \geq t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + dt)}{dt \cdot \mathbb{P}(T \geq t)} \quad \text{par la formule de Bayes.} \\ &= \frac{1}{S(t)} \lim_{dt \rightarrow 0} \frac{S(t) - S(t + dt)}{dt} \\ &= -\frac{S'(t)}{S(t)} \end{aligned}$$

En résolvant cette équation différentielle on obtient :

$$S(t) = \exp\left(-\int_0^t \mu(s) ds\right)$$

Enfin, il peut être utile de rappeler quelques notations d'assurance vie :

- ${}_tq_x$: le taux de mortalité associé à ${}_tp_x$ (i.e. ${}_tq_x = 1 - {}_tp_x$). Il s'agit de la probabilité d'un individu d'âge x de décéder avant l'âge $x + t$.
Par abus de notation on pourra noter q_x (respectivement p_x) à la place de ${}_1q_x$ (respectivement ${}_1p_x$). Il s'agit de la probabilité pour un individu d'âge x de décéder dans l'année.
- L_x : le nombre de survivants à l'âge x .

4.3 Taux bruts de mortalité

Les taux de mortalité ont pour but de décrire l'évolution de la mortalité au sein de la population. Il s'agit d'une mesure importante liée au risque de longévité.

Il existe de nombreux estimateurs qui servent à les évaluer. Ici, nous nous intéressons à l'estimateur binomial ainsi qu'à l'estimateur de Kaplan-Meier. Contrairement à l'estimateur binomial, celui-ci permet de prendre en compte le fait que les observations des modèles de survie sont souvent incomplètes.

4.3.1 Estimateur binomial

Les taux de mortalité peuvent être estimés via la méthode du maximum de vraisemblance. De manière formelle, on définit une variable aléatoire Y valant 1 si l'individu i décède à l'âge (entier) x , et 0 sinon. Elle suit donc une loi de Bernoulli de paramètre q_x .

On rappelle que la probabilité d'occurrence d'une variable suivant une loi de Bernoulli de paramètre q_x s'écrit :

$$\mathbb{P}(Y_i = k) = q_x^k (1 - q_x)^{1-k}$$

Avec $k \in \{0, 1\}$

On se donne un échantillon $Y = (Y_1, \dots, Y_{L_x})$ de variables aléatoires indépendantes et identiquement distribuées suivant une telle loi. Il est alors possible d'explicitier la vraisemblance de ce modèle :

$$\begin{aligned} L(q_x | y_1, \dots, y_{L_x}) &= \prod_{i=1}^{L_x} \mathbb{P}(Y_i = y_i) \\ &= \prod_{i=1}^{L_x} q_x^{y_i} (1 - q_x)^{1-y_i} \\ &= q_x^{\sum_{i=1}^{L_x} y_i} (1 - q_x)^{L_x - \sum_{i=1}^{L_x} y_i} \end{aligned}$$

Où y_1, \dots, y_{L_x} sont les réalisations de Y_1, \dots, Y_{L_x} .

Afin de maximiser cette quantité, on calcule la log-vraisemblance :

$$l(q_x | y_1, \dots, y_{L_x}) = \log(q_x) \sum_{i=1}^{L_x} y_i + \left(L_x - \sum_{i=1}^{L_x} y_i \right) \log(1 - q_x)$$

Il s'agit d'une fonction concave en q_x , qui admet donc un maximum qui annule sa dérivée.

Calculons la dérivée de la log-vraisemblance :

$$\frac{\partial l}{\partial q_x}(q_x | y_1, \dots, y_{L_x}) = \frac{1}{q_x} \sum_{i=1}^{L_x} y_i - \frac{1}{1 - q_x} \left(L_x - \sum_{i=1}^{L_x} y_i \right)$$

Cette quantité s'annule pour :

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

Avec $d_x = \sum_{i=1}^{L_x} y_i$ représentant le nombre de décès à l'âge x .

L'estimateur du maximum de vraisemblance du taux de mortalité vaut donc le ratio entre le nombre de décès survenus entre les âges x et $x + 1$, et le nombre de survivants à l'âge x . Il s'agit ainsi d'un estimateur assez simple et intuitif. On peut montrer que c'est un estimateur sans biais. En effet, en supposant l'indépendance entre les individus d'un même âge, D_x suit une loi binomiale de paramètre (L_x, q_x) , où D_x est la variable aléatoire associée aux réalisations d_x . Donc :

$$\mathbb{E}[D_x] = L_x q_x$$

D'où,

$$\mathbb{E}[\hat{q}_x] = q_x$$

Par conséquent, il s'agit bien d'un estimateur sans biais.

D'autre part, on peut calculer la variance de cet estimateur qui vaut :

$$\mathbb{V}[\hat{q}_x] = \frac{\mathbb{V}[D_x]}{L_x^2} = \frac{q_x(1 - q_x)}{L_x}$$

La variance de cet estimateur augmente de manière assez importante lorsque la taille de l'échantillon est faible. C'est notamment le cas aux grands âges, intervalle qui nous intéresse tout particulièrement.

À partir de là, comme les individus sont supposés indépendants et identiquement distribués, il est possible d'appliquer le théorème central limite, et obtenir à l'aide du lemme de Slutsky et de la loi des grands nombres :

$$\sqrt{L_x} \frac{(\hat{q}_x - q_x)}{\sqrt{\hat{q}_x(1 - \hat{q}_x)}} \xrightarrow{L_x \rightarrow \infty} \mathcal{N}(0, 1)$$

Ainsi, en notant k_α le quantile de niveau α d'une loi normale centrée et réduite, on obtient :

$$\mathbb{P} \left(\left| \sqrt{L_x} \frac{(\hat{q}_x - q_x)}{\sqrt{\hat{q}_x(1 - \hat{q}_x)}} \right| \leq k_{1-\frac{\alpha}{2}} \right) \xrightarrow{L_x \rightarrow \infty} 1 - \alpha$$

Un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour le taux de mortalité d'un

individu d'âge x est donc :

$$IC(q_x)_{1-\alpha} = \left[\hat{q}_x - k_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{L_x}}; \hat{q}_x + k_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{L_x}} \right]$$

En pratique, pour la plupart des âges, le nombre de personnes observées est assez grand pour pouvoir appliquer l'intervalle de confiance obtenu ci-dessus. Cela nous permet de mettre un degré de confiance à accorder aux taux bruts obtenus via l'estimateur binomial. On peut donc tracer les taux bruts de mortalité avec leurs intervalles de confiance à 95% pour une cohorte donnée :

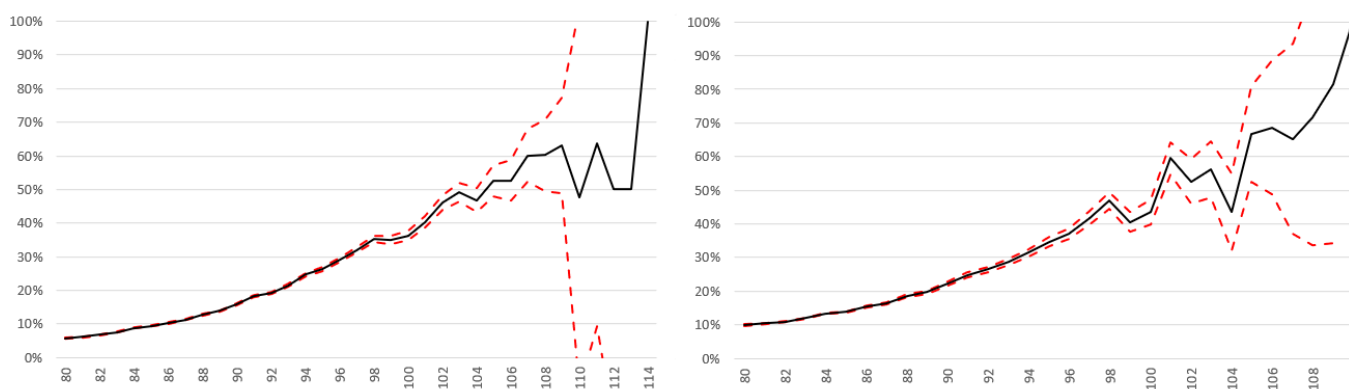


FIGURE 4.2 – Intervalles de confiance à 95 % pour l'estimateur binomiale à partir de 80 ans pour les femmes (à gauche) et les hommes (à droite) pour la cohorte 1905

Nous observons, sur ces deux graphiques, que la largeur de l'intervalle de confiance s'accroît rapidement avec l'âge, et dépasse 5% dès 100 ans chez les femmes et autour de 95 ans chez les hommes. Ce seuil de 5% a été pris arbitrairement, mais une telle variation n'est pas négligeable et peut avoir un impact sur le provisionnement de certains portefeuilles. On voit également, qu'après 105 ans, les taux de mortalité obtenus n'ont plus vraiment de sens, étant donné l'incertitude qui les entoure.

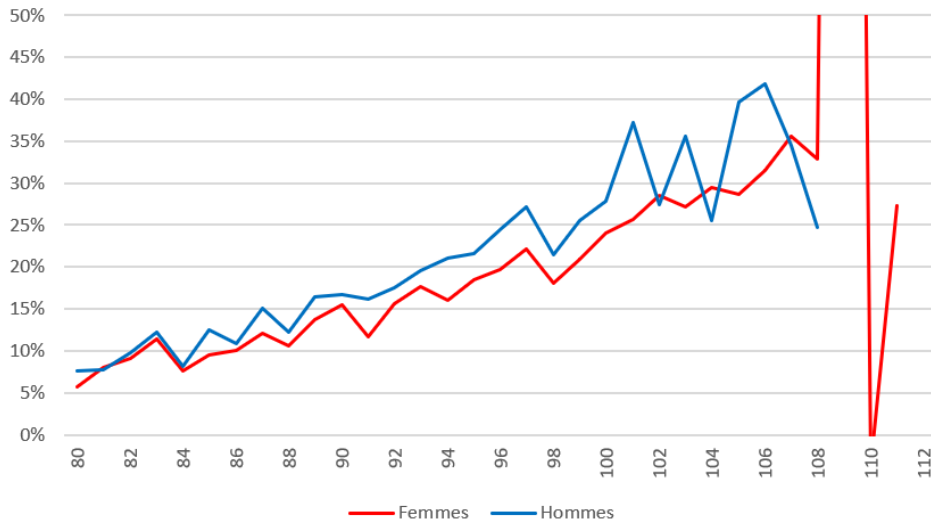


FIGURE 4.3 – *Taux d'accroissement de la largeur de l'intervalle de confiance par sexe pour l'estimateur binomial pour la cohorte 1905*

Le taux de mortalité obtenu est de plus en plus incertain avec l'âge, étant donné la croissance de la largeur de l'intervalle de confiance. Cette croissance est rapide et supérieure à 10% par âge dès 85 ans, ce qui représente une forte croissance.

Étant donné les intervalles de confiance et leurs évolutions en fonction de l'âge à génération fixée, nous proposons de fixer l'âge charnière à partir duquel nous ajustons les taux de mortalité autour de 85 ans. Cet âge pourra toutefois varier selon les méthodes d'ajustement et les générations considérées.

En résumé, nous retenons que l'estimateur binomial est un estimateur performant si les données proviennent d'observations faites sur la durée de vie entière et si l'échantillon est assez grand. En général, ce n'est pas le cas et cet estimateur peut s'avérer insuffisant car assez incertains. Les observations sont souvent incomplètes, ce qui entraîne un décalage avec la mortalité réelle. En effet, cet estimateur n'est pas sensible aux phénomènes de censure et de troncature fréquents pour les modèles de durée. Il reste cependant beaucoup utilisé pour sa simplicité et sa justesse lorsque L_x est grand.

4.3.2 Estimateur de Kaplan-Meier

L'estimateur non-paramétrique de Kaplan-Meier est un estimateur beaucoup utilisé pour la construction de tables de mortalité, car il permet de prendre en compte les observations incomplètes. Il s'agit de l'estimateur du maximum de vraisemblance de S , la fonction de survie. Il repose, en grande partie, sur le théorème de Glivenko-

Cantelli, qui montre que cet estimateur converge de manière uniforme vers la fonction de survie. Ainsi, on peut définir un taux de mortalité prenant en compte la censure et la troncature, assez réaliste lorsque la taille de l'échantillon observée est significative. Dans le cas contraire, cet estimateur aura également une forte volatilité.

En modèle de durée, les observations sont longues à obtenir à l'échelle humaine tandis que la fenêtre d'observation, relativement courte, favorise, par exemple, le fait que les individus étudiés puissent quitter l'observation en cours de route. Dans ce cas, on parle de censure à droite. Ne pas prendre en compte la censure conduirait à sous-estimer la mortalité réelle de la population. En effet, le fait qu'un individu ait quitté l'observation à un âge x contient l'information que cet individu a survécu au moins jusqu'à l'âge x .

L'estimateur de Kaplan-Meier¹ est un estimateur de la fonction de survie qui s'écrit comme suit :

$$\hat{S}_x(Y_i) = \prod_{Y_i < t} \left(1 - \frac{d_i}{n_i}\right)$$

Avec :

- d_i le nombre de décès d'individus d'âge x en Y_i .
- n_i le nombre de personnes survivantes d'âge x juste avant l'instant Y_i .

Une estimation de la variance du taux de mortalité obtenu via l'estimateur de Kaplan-Meier peut être donnée par l'estimateur de Greenwood, définit par :

$$\mathbb{V}[\tilde{q}_x] = \hat{S}_x(Y_i)^2 \sum_{Y_i < t} \frac{d_i}{n_i(n_i - d_i)}$$

Frédéric Planchet montre², qu'un intervalle de confiance asymptotique de niveau α pour q_x est :

$$IC_{1-\alpha}(q_x) = \left[1 - (1 - \tilde{q}_x) \left(1 \pm k_{1-\frac{\alpha}{2}} \sqrt{\frac{\mathbb{V}[\tilde{q}_x]}{\hat{S}_x(t)^2}} \right) \right]$$

À partir de ces formules, il est alors aisé de visualiser les estimations de mortalité obtenues à partir de l'estimateur de Kaplan-Meier, avec leurs intervalles de confiance pour une cohorte donnée :

1. Si le lecteur veut en savoir plus sur cet estimateur il peut, par exemple, se référer à LUZON (2019) qui détaille sa construction.

2. Cf. PLANCHET (2021b).

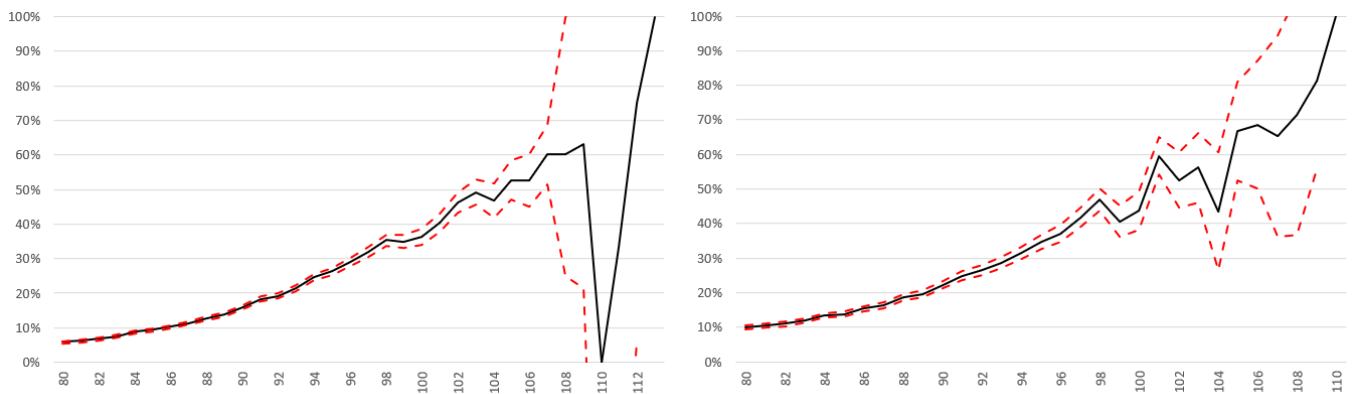


FIGURE 4.4 – *Intervalle de confiance à 95% du taux de mortalité obtenu via l'estimateur de Kaplan Meier pour les femmes (à gauche) et les hommes (à droite) pour la cohorte 1905*

La largeur de l'intervalle de confiance augmente rapidement à partir de 90 ans chez les hommes, et 95 ans chez les femmes. Les taux de mortalité après ces âges sont donc relativement peu fiables. On voit également une certaine volatilité à ces âges. À partir de 108 ans, l'intervalle de confiance ne fournit d'ailleurs plus aucune information étant que celui-ci n'est plus inclus dans l'intervalle $]0, 1[$.

Cet estimateur a donc des limites aux grands âges, où la volatilité des taux de mortalité obtenus grâce à l'estimateur de Kaplan-Meier reste très importante et les intervalles de confiance très larges.

Au vu de ces résultats, l'ajustement de la mortalité aux grands âges paraît alors nécessaire autour de 90 ans. Cet âge dépend, là encore, des méthodes utilisées. Avant d'étudier ces modèles, il reste encore à compléter la table avec la mortalité prospective.

4.4 Mortalité prospective

Plusieurs choix s'offrent à nous pour l'estimation de la mortalité prospective mais nous utilisons une méthode rapide à mettre en oeuvre. Il s'agit de la méthode décrite dans PLANCHET (2007b). Elle se fonde sur une approche exogène et l'utilisation d'une table externe.

Cela servira également au lissage des taux bruts de mortalité.

Les taux de mortalité prospectifs sont extrapolés en appliquant la formule suivante :

$$\begin{aligned}\log it(q_{x,t}) &= \log\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) \\ &= (\alpha + \beta b_x) \log it(q_{x,t}^{ref}) + b_x + \epsilon_{x,t}\end{aligned}$$

Avec $q_{x,t}^{ref}$ les taux de références provenant d'une table externe, et $\epsilon_{x,t}$ une erreur Gaussienne centrée de variance σ^2 .

La différence des $\log it$ entre la mortalité de référence et la mortalité des données d'observations est la suivante :

$$\begin{aligned}\Delta_x(t) &= \log it(q_{x,t}) - \log it(q_{x,t}^{ref}) \\ &= \log it(q_{x,t}^{ref})(\alpha + \beta b_x - 1) + b_x\end{aligned}$$

Pour estimer les paramètres, le but est de minimiser cette différence, en valeur absolue, en fonction de α , β et des b_x . En pratique cette minimisation a été faite génération par génération via la méthode du GRG³ non linéaire du solveur Excel.

Pour cette étude sur la population nationale française, les tables de référence qui sont utilisées sont les tables réglementaires TGH - TGF 05. Ces tables prévoient la mortalité jusqu'à la génération 2010 pour les âges allant de 0 jusqu'à 120 ans.

La mortalité est donc projetée jusqu'à la génération 2010, mais ne l'est pas pour les générations suivantes. Cela ne nous pose pas de problèmes, étant donné notre sujet d'étude.

La différence Δ_x obtenue est assez faible, ce qui montre une assez bonne cohérence des données par rapport aux tables réglementaires.

Nous obtenons alors la mortalité prospective suivante :

3. Generalized Reduced Gradient

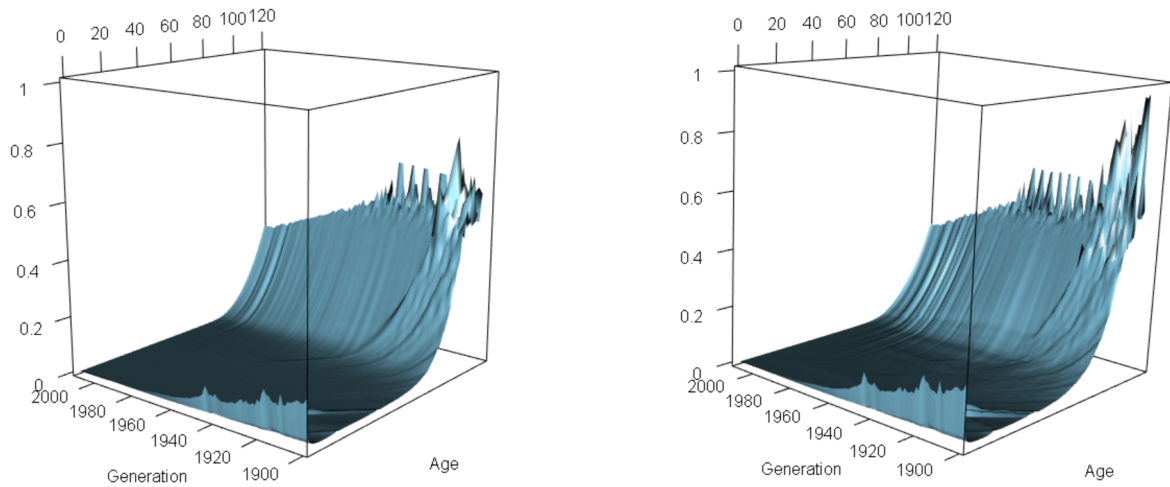


FIGURE 4.5 – *Mortalité prospective pour les femmes (à gauche) et pour les hommes (à droite)*

Nous pouvons voir que les taux de mortalité ainsi obtenus sont lissés, et cela entraîne quelques différences avec ceux présentés précédemment. Il reste une forte volatilité pour personnes âgées, notamment pour la population masculine.

En résumé nous avons pu construire des estimateurs des taux de mortalité, en particulier l'estimateur binomial et l'estimateur de Kaplan-Meier. Nous avons également pu projeter la mortalité future en nous fondant sur une table externe. Cela n'est cependant pas suffisant, il faut désormais se concentrer sur les grands âges, là où la taille des échantillons est très réduite, la volatilité élevée et où les intervalles de confiance obtenus sont très larges.

Chapitre 5

Ajustement aux âges avancés

Nous avons vu au Chapitre précédent que les taux de mortalité obtenus aux grands âges restent très volatils et assez peu fiables. Cela est vrai pour les taux obtenus via les deux estimateurs utilisés. Cela nous amène à nous concentrer spécifiquement sur la fermeture de la table. Pour cela, nous allons tester différents modèles, paramétriques ou non, et les ajuster aux données. Nous rappelons que les paramètres des modèles sont estimés génération par génération, et que les modèles prennent donc en compte la dérive de longévité bien que cela ne soit pas toujours précisé dans les notations. Prendre en compte des paramètres globaux qui n'évoluent pas avec la génération introduirait certainement un biais.

Il faut également garder en tête que d'autres modèles non testés ici existent. C'est par exemple le cas des modèles de Coale & Kisker ou Coale & Guo. Ces deux méthodes ont notamment été testées dans QUASHIE et DENUIT (2005), où ils montrent que les résultats obtenus ne sont pas très concluants sur leurs données. Un modèle bayésien ainsi que les modèles à noyau sont également développés en annexe (cf. Annexe B).

Les résultats seront présentés en 3D en prenant en compte à la fois la génération et l'âge. Cependant, ce type de représentation ne permet pas de se rendre compte efficacement de l'ajustement du modèle aux taux bruts. Il est préférable de se focaliser sur une génération spécifique. Les graphiques se concentreront sur l'année 1905, étant donné que ce sont des données historiques et non projetées. Bien entendu, les autres générations auraient également pu être considérées. Les résultats, commentaires et analyses présentés pour cette génération peuvent, s'étendre à toutes les cohortes.

Dans ce Chapitre, nous nous concentrerons principalement sur l'implémentation des modèles sans réellement les évaluer. Cela fera l'objet du Chapitre suivant, dans lequel nous développerons différentes méthodes pour nous rendre compte, de manière rigoureuse, de la qualité de l'ajustement.

Sauf mention contraire les taux bruts de mortalité utilisés sont les taux lissés obtenus grâce à l'estimateur de Kaplan-Meier.

5.1 Modèles paramétriques

Les modèles paramétriques sont intéressants, mais reposent sur des hypothèses généralement assez fortes quant à la forme des taux de mortalité. C'est le cas des modèles que nous allons tester et ajuster dans cette partie. Il faut donc, pour mettre en place ce type de modèle, avoir une idée de la forme des taux de mortalité. Par exemple, le modèle des P-splines fait l'hypothèse d'une évolution polynomiale par morceaux, tandis que celui de Kannistö, cas particulier du modèle logistique, propose d'ajuster une sigmoïde aux données.

5.1.1 Modèle des P-splines

- **Considérations théoriques**

L'idée de cette méthode est de découper l'intervalle à ajuster en plusieurs sous-intervalles, sur lesquels ajuster des fonctions assez simples appelées splines, qui sont, en général, des fonctions polynomiales. Les bornes des sous-intervalles sont appelées noeuds. Il faut faire attention au niveau de ces noeuds, âges de jonction entre les différentes fonctions afin que la courbe des taux reste lisse. En pratique, les polynômes considérés sont de degré 3, voire 4, mais rarement plus.

Mathématiquement, cela s'écrit assez simplement en une dimension mais l'écriture se complique avec l'augmentation du nombre de dimensions. Dans notre cas, nous étudions cette méthode avec deux dimensions : âge et génération.

Pour simplifier l'écriture de ce modèle, nous allons adopter une approche matricielle des polynômes. De manière formelle, il est possible d'écrire un polynôme de degré n , un entier strictement positif, de la manière suivante :

$$P(x, t) = {}^tTAX$$

Où tT est la transposée de T , $T = \begin{pmatrix} 1 \\ t \\ \vdots \\ t^n \end{pmatrix}$, $X = \begin{pmatrix} 1 \\ x \\ \vdots \\ x^n \end{pmatrix}$, et $A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{pmatrix}$.

On note $X^{(d)}$ le vecteur X dans lequel chaque terme a été dérivé d fois par rapport à x . De même, $T^{(d)}$ est la dérivée d -ième, terme à terme, du vecteur T par rapport à t .

$$\text{Par exemple } X^{(1)} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ nx^{n-1} \end{pmatrix}.$$

On peut également montrer, par une simple dérivation polynomiale, que les dérivées partielles s'écrivent de la manière suivante :

$$\begin{aligned} \frac{\partial P}{\partial x}(x, t) &= {}^tTAX^{(1)} & \frac{\partial^2 P}{\partial x^2}(x, t) &= {}^tTAX^{(2)} \\ \frac{\partial P}{\partial t}(x, t) &= {}^tT^{(1)}AX & \frac{\partial^2 P}{\partial t^2}(x, t) &= {}^tT^{(2)}AX \\ \frac{\partial^2 P}{\partial t \partial x}(x, t) &= {}^tT^{(1)}AX^{(1)} \end{aligned}$$

On note e_1, \dots, e_n les bornes des intervalles à l'intérieur desquels un unique polynôme est utilisé. Cela signifie que la table sera ajustée par intervalle $[e_i; e_{i+1}]$. La mortalité au sein de cet intervalle est définie par le polynôme P_i . Il est alors utile de définir les polynômes les uns en fonction des autres, afin que la courbe des taux de mortalité reste lisse.

Il est primordial de bien choisir les noeuds e_i , essentiellement leur nombre et leurs niveaux, pour avoir une représentation la plus juste possible de la mortalité.

Afin de comprendre les contraintes que cela entraîne, on introduit $P_1 = {}^tTAX$ et $P_2 = {}^tTBX$ deux polynômes qui ont le noeud commun e_2 . P_1 est donc défini sur $[e_1; e_2]$ et P_2 sur $[e_2; e_3]$.

Il est montré dans PLANCHET (2007a) que les raccordements au niveau du noeud e_2 impliquent les deux contraintes suivantes :

$$\begin{aligned} {}^tT^{(1)}(A - B) &= 0^1 \\ (A - B)X^{(1)} &= 0^1 \end{aligned}$$

Il y a, de plus, un critère de régularité et un critère de fidélité à prendre en compte. On peut combiner ces deux critères, ce qui revient à minimiser la quantité suivante :

$$\alpha \sum_x \sum_t \omega_{x,t} (P(x, t) - q_{x,t})^2 + (1 - \alpha) \iint \lambda(s, y) |H_P(s, y)|^2 ds dy$$

Avec P la spline à ajustée, $\alpha \in [0, 1]$ un paramètre permettant de privilégier soit la

1. La dimension étant donnée par le terme de gauche, nous notons indifféremment 0 de $(0, \dots, 0)$ ou ${}^t(0, \dots, 0)$.

fidélité (si $\alpha > 0,5$) soit la régularité (si $\alpha < 0,5$), H_P la matrice hessienne du polynôme P , $\omega_{x,t}$ et λ les poids associés aux points $q_{x,t}$.

Dans cette méthode, le choix des paramètres consiste à choisir les bons noeuds, ainsi que le degré des splines le plus adéquat.

- **Considérations pratiques**

En pratique, nous utilisons le package « splines » du langage R. Il permet l'ajustement des taux via cette méthode. L'ajustement a été fait, dans notre cas, à partir de 85 ans pour les hommes et de 93 ans pour les femmes. Les splines utilisées sont des splines de degré 4, étant donné que des polynômes de degré 2 et 3 ajustent les splines de manière trop approximative.

Les résultats obtenus sont les suivants :

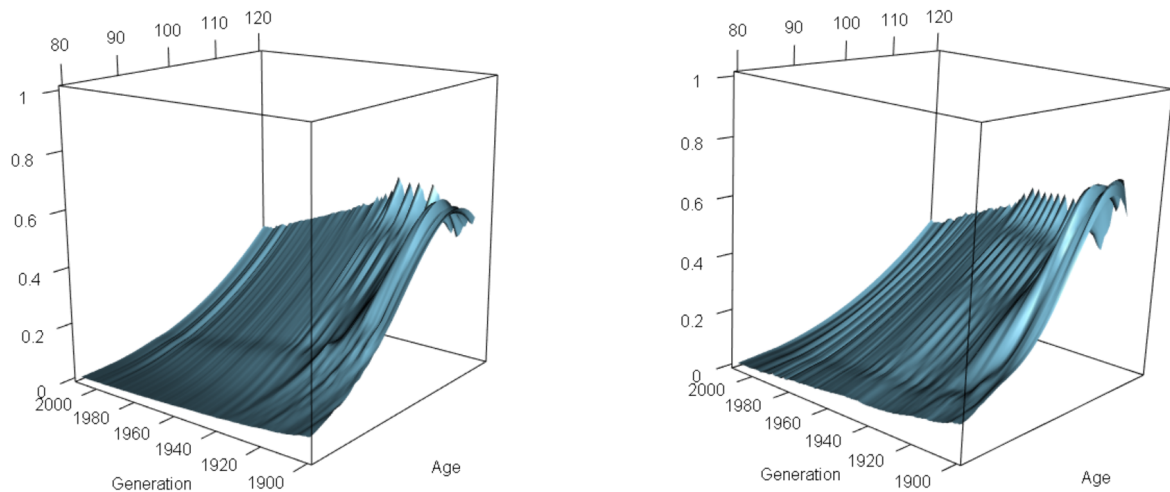


FIGURE 5.1 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle des P-splines*

Il reste un peu de volatilité à âge fixé en fonction des générations, en particulier pour les âges supérieurs à 95 ans. Cela reste cependant assez faible et ne pose donc pas réellement problème.

On constate, par ailleurs, que certains taux de mortalité ont tendance à diminuer à partir d'un certain âge, notamment pour les générations les plus anciennes.

Il est plus aisé de visualiser le niveau d'adéquation aux taux bruts sur un graphe en deux dimensions. Prenons, pour cela, l'exemple de la cohorte 1905 :

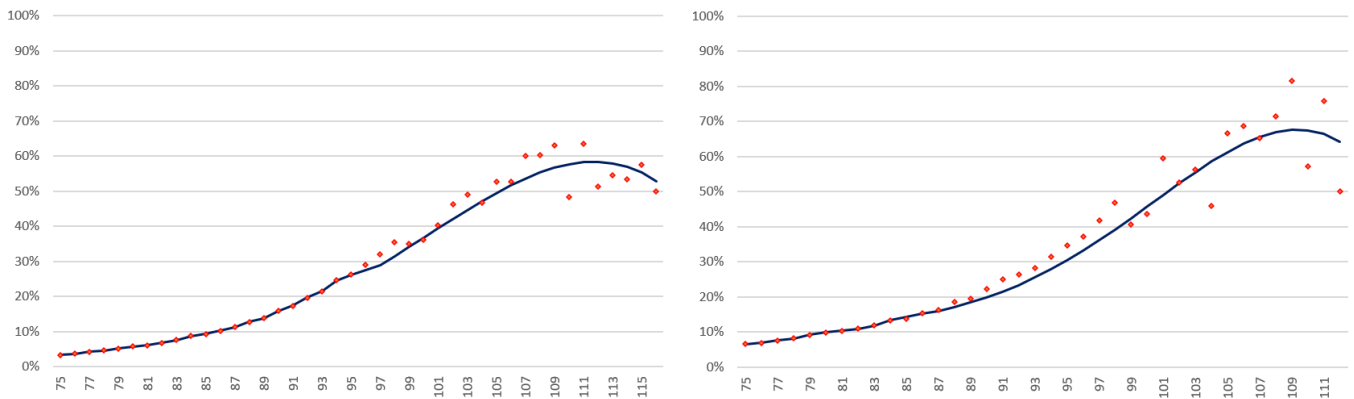


FIGURE 5.2 – Ajustement des taux de mortalité via la méthode des *P-splines* pour les femmes (à gauche) et les hommes (à droite) de la génération 1905

Nous constatons un bon ajustement, que ce soit pour les hommes ou pour les femmes, malgré ce qui semble être une légère sous-estimation des taux entre 90 et 100 ans. Cela semble provenir des oscillations aux grands âges, qui poussent les taux ajustés vers le bas. On observe également une légère baisse de la mortalité après 109 ans. Cependant, cela semble aller dans le sens des taux bruts obtenus.

Pour les autres générations, l'ajustement est aussi souvent sous-estimé pour les premiers âges suivants le début de l'ajustement. Cela conduira, entre autres, à une plus grande prudence dans le provisionnement. Cependant, cette sous-estimation reste assez légère et n'aura donc pas une très grande influence.

5.1.2 Modèle de Lindbergson

- **Considérations théoriques**

En travaillant sur des données suédoises au début du XXI^{ème} siècle, Lindbergson a proposé une approche permettant de lisser les taux tout en ajustant la mortalité aux grands âges, en se basant sur le modèle de Gompertz-Makeham qui fournit, en général, des taux trop élevés à partir un certain âge, d'où la nécessité d'un ajustement. Pour les âges élevés, Lindbergson a proposé de remplacer la croissance exponentielle de ce modèle par une croissance linéaire de la force de mortalité. Les taux de mortalité aux autres âges correspondent aux mêmes que ceux du modèles de Gompertz-Makeham. Cela est assez cohérent étant donné que plusieurs études démographiques montrent que cette croissance est concave en échelle logarithmique à partir d'un certain âge. C'est d'ailleurs l'une des constatations que l'on a pu faire sur les données considérées².

2. Cf. Chapitre 3.

Formellement, le modèle s'écrit de la manière suivante :

$$\mu_x(a, b, c, k, \omega) = \begin{cases} a + b \exp(cx) & \text{si } x \leq \omega \\ a + b \exp(c\omega) + k(x - \omega) & \text{si } x \geq \omega \end{cases}$$

Où il faut estimer les paramètres a, b, c et k ainsi que ω , l'âge charnière. Ils dépendent tous de la génération bien que, par souci de clarté on omettra l'indice.

Les paramètres a, b, c, k et ω peuvent être estimés en minimisant l'écart, sous une norme choisie, entre les taux bruts et les taux ajustés. Comme donnée par QUASHIE et DENUIT (2005), la quantité à minimiser est Q tel que :

$$\begin{aligned} (\hat{a}, \hat{b}, \hat{c}, \hat{k}) &= \arg \min_{(a,b,c,k,\omega) \in \mathbb{R}^4} Q(a, b, c, k, \omega) \\ &= \arg \min_{(a,b,c,k,\omega) \in \mathbb{R}^4} \sum_{x \geq x_0} \frac{(\hat{\mu}_x - \mu_x(a, b, c, k, \omega))^2}{\frac{\hat{\mu}_x}{L_x}} \\ &= \arg \min_{(a,b,c,k,\omega) \in \mathbb{R}^4} \sum_{x=x_0}^{\omega-1} \frac{(\hat{\mu}_x - a - b \exp(cx))^2}{\frac{\hat{\mu}_x}{L_x}} + \sum_{x \geq \omega} \frac{(\hat{\mu}_x - a - b \exp(c\omega) - k(x - \omega))^2}{\frac{\hat{\mu}_x}{L_x}} \end{aligned}$$

Où x_0 est un âge à partir duquel les taux de mortalité sont croissants. Pour bien nous assurer de cette propriété, nous nous positionnons après la bosse des accidents et nous choisissons de prendre x_0 valant 30 ans.

$\frac{L_x}{\hat{\mu}_x}$ sert ici de poids aux écarts afin de ne pas donner trop d'importance aux âges où l'effectif est très faible, et donc plus d'importance là où l'effectif est plus conséquent.

Notons que cette quantité admet bien un minimum, car en tant que somme de fonctions convexes, il s'agit d'une fonction convexe.

• Considérations pratiques

On minimise cette quantité de manière numérique. Ici, la minimisation a été faite via la fonction « constrOptim » de R. Cette fonction permet de faire de l'optimisation sous contraintes, et repose sur la méthode de minimisation de Nelder-Mead.

L'optimisation sous contraintes est intéressante dans notre cas, car la signification des paramètres les contraint notamment à être positifs. Les contraintes utilisées sont les contraintes naturelles, à savoir que les paramètres a, b, c et k doivent être compris entre 0 et 1. Aucune contrainte n'est mise sur l'âge charnière ω .

Après cet ajustement, on constate une forte volatilité en fonction des générations. Cette volatilité apparaît à travers l'oscillation du paramètre k et, également, l'oscillation de l'âge charnière. Un lissage peut ici s'avérer utile sur ces deux paramètres. Le lissage effectué est un lissage polynomial. On obtient ainsi :

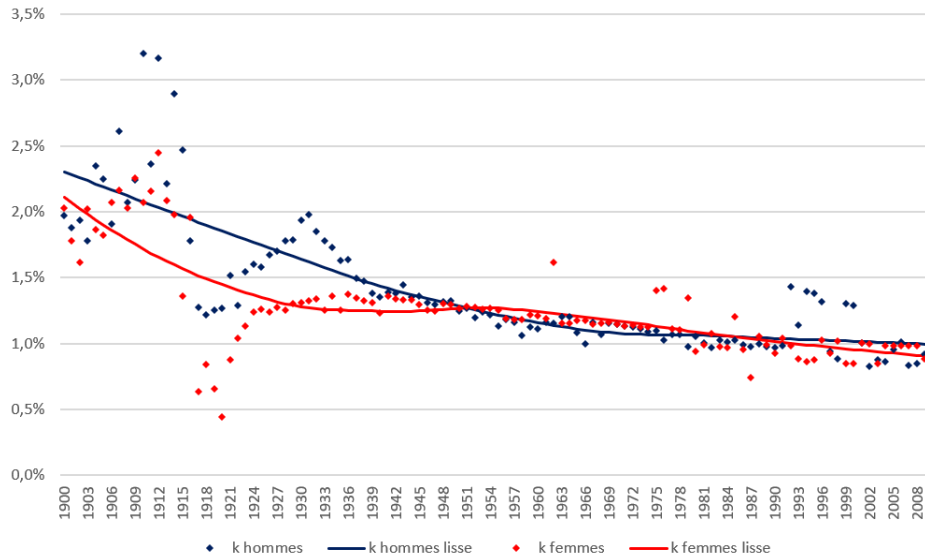


FIGURE 5.3 – *Évolution et lissage du paramètre k pour le modèle de Lindbergson*

Nous observons une volatilité importante, en particulier pour les générations les plus anciennes pour les paramètres non lissés. La volatilité pour les générations où les taux de mortalité ont été projetés est plus faible. Le lissage permet de réduire cette volatilité. Cette Figure montre également une baisse moyenne du paramètre k qui illustre la baisse de la mortalité aux grands âges.

La même méthode de lissage a été utilisée pour l'âge charnière, qui oscille beaucoup autour de 90 ans d'une génération à l'autre. Le lissage est également important sur ce paramètre, pour ne pas avoir de gros écarts d'une génération à l'autre.

Après le lissage les paramètres k et ω et le passage de la force de mortalité aux taux de mortalité, les résultats obtenus sont les suivants :

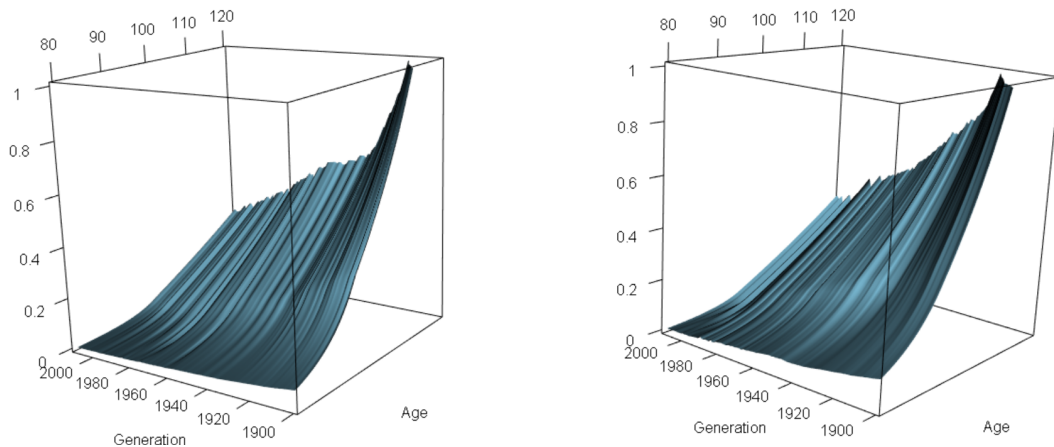


FIGURE 5.4 – Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Lindbergson

Les graphes montrent des taux lissés par rapport aux deux dimensions, et le changement de tendance apparaît naturel au niveau de l'âge charnière. La dérive de longévité, représentée par la baisse des taux de mortalité, est bien observable pour ce modèle.

Considérons à présent l'adéquation des taux obtenus avec ce modèle par rapport aux taux bruts de mortalité.

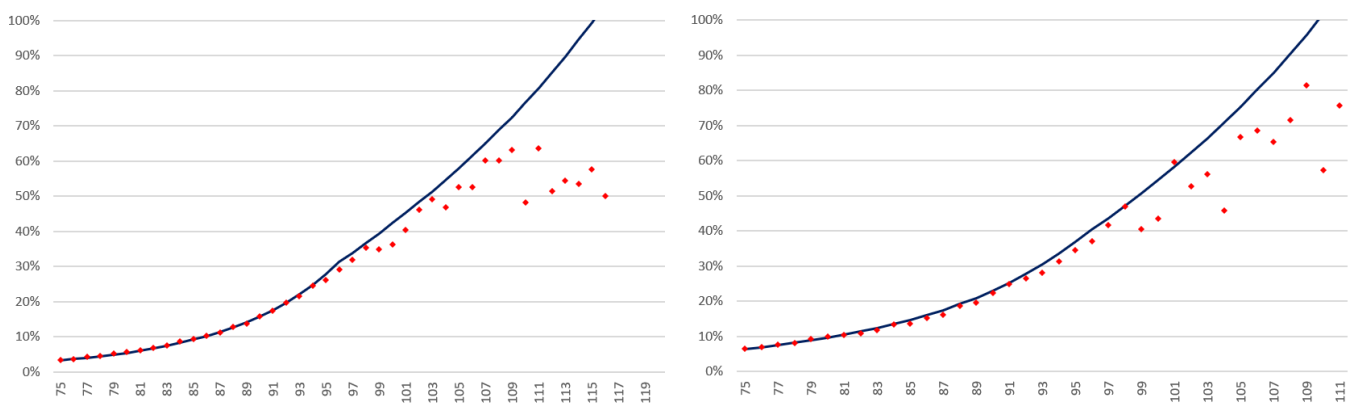


FIGURE 5.5 – Ajustement des taux de mortalité via la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) de la génération 1905

On remarque que l'ajustement semble assez éloigné des taux bruts de mortalité pour cette génération. Les taux de mortalité semblent surestimés, et atteignent même 100% avant que le dernier survivant ne disparaisse. Cela est dû au lissage horizontal, étant donné que la génération 1905 a, dans les faits, une mortalité plus forte que la moyenne. Cela peut notamment s'observer sur la Figure 5.3 dans laquelle le paramètre k lissé est plus faible que le paramètre k non lissé. Pour une génération où le lissage

des paramètres à un impact plus réduit, comme pour la génération 1910 par exemple, l'ajustement est le suivant :

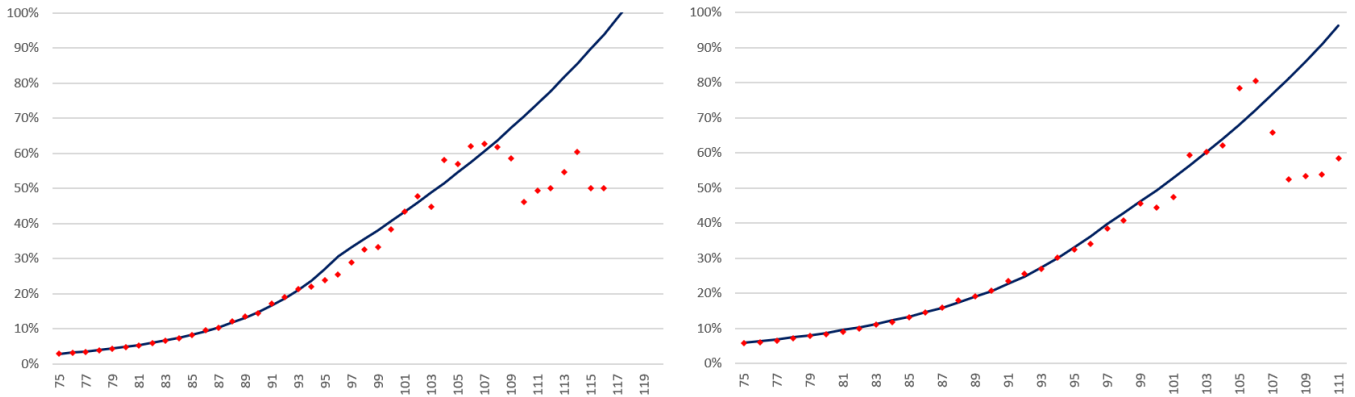


FIGURE 5.6 – *Ajustement des taux de mortalité via la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) de la génération 1910*

Celui-ci est nettement meilleur – notamment pour les hommes – mais, encore une fois, cet ajustement surestime les taux bruts de mortalité à partir de 110 ans pour les femmes et 106 ans pour les hommes. Il n’y a rien d’aberrant à ce résultat étant donné le faible échantillon.

Nous ne constatons pas de baisse du taux de mortalité à partir d’un certain âge comme c’est le cas avec le modèle des p splines.

Nous retiendrons que le point essentiel de cette méthode est le choix de l’âge charnière ω . Le choisir sans précaution conduirait irrémédiablement à une erreur plus importante. Le paramètre k joue également un rôle crucial.

5.1.3 Modèle de Denuit & Goderniaux

- **Considérations théoriques**

Le modèle de Denuit & Goderniaux, datant de 2005, postule que l’évolution du logarithme du taux de mortalité est un polynôme de degré 2. Cette méthode introduit également une contrainte sur l’âge le plus élevé de la table. Celle-ci est que le taux de mortalité pour l’âge le plus élevé est forcément égal à 100%. Cela est, en soi, assez logique car cela signifie que personne ne survit au delà de cet âge. Cependant, les autres modèles n’introduisent pas ce type de contrainte, ce qui conduit d’ailleurs à des taux de mortalité supérieurs à 100% avec le modèle de Lindbergson, bien qu’il reste des survivants. Le modèle des P-splines conduit, au contraire, à des taux de mortalité relativement bas.

Ce modèle s'appuie sur cette dynamique du taux de mortalité :

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

La contrainte se traduit par :

$$\begin{cases} q_{x_{max}} = 1 \\ q'_{x_{max}} = 0 \end{cases}$$

Avec q'_x la dérivée du taux de mortalité q_x par rapport à l'âge. La deuxième contrainte permet d'éviter une décroissance éventuelle au dernier âge.

Ces contraintes conduisent ainsi au système d'équations suivant :

$$\begin{cases} 0 = a + x_{max}b + x_{max}^2c \\ 0 = b + 2x_{max}c \end{cases}$$

D'où,

$$\begin{cases} a = cx_{max}^2 \\ b = -2cx_{max} \end{cases}$$

Afin d'ajuster le modèle, on utilise la méthode des moindres carrés ordinaires.

Un des avantages non-négligeables de ce modèle est qu'il n'y a qu'un seul paramètre à estimer, ce qui en fait in fine un modèle assez simple à mettre en pratique.

On pose également la contrainte que le paramètre c doit être négatif. Cela permet de garder les taux ajustés entre 0 et 1.

Dans la littérature actuarielle on retrouve en général x_{max} valant 130 ans. Cela paraît assez cohérent étant donné que la doyenne de l'humanité est décédée dans sa 123^{ème} année. Il nous semble également raisonnable et cohérent de considérer un x_{max} autour de 125 ans. Les deux éventualités ont ainsi été implémentées. Un autre âge plus petit que 125 ans ou plus grand que 130 ans n'aurait pas vraiment de sens en regard des observations actuelles.

- **Considérations pratiques**

Pour le cas où x_{max} vaut 125 ans, les résultats sont les suivants :

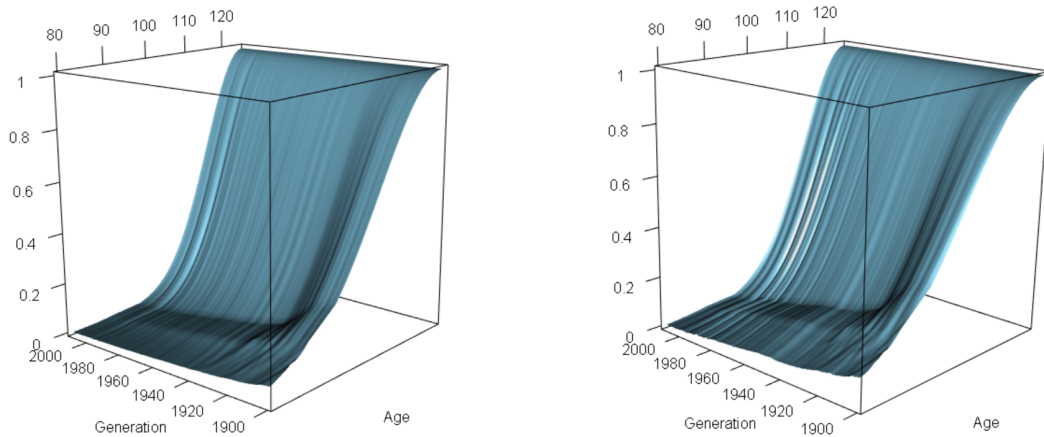


FIGURE 5.7 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Denuit & Goderniaux avec x_{max} valant 125 ans*

Nous remarquons trois choses sur les résultats obtenus :

- Bien que l'on ait lissé les taux de mortalité autour de l'âge charnière au moyen de moyennes pondérées, ce lissage n'est pas très bon étant donné que le changement de tendance est encore bien visible et assez brusque pour les générations après 1920.
- Il n'est pas nécessaire de réaliser un lissage horizontal comme cela a pu être le cas pour la méthode précédente.
- La forme imposée par ce modèle pousse la courbe des taux de mortalité à être concave aux grands âges.

À première vue, l'âge maximal x_{max} semble donc être raisonnable pour les générations avant 1920, mais légèrement trop jeune pour les suivantes.

L'adéquation de l'ajustement aux taux bruts donne les résultats suivants :

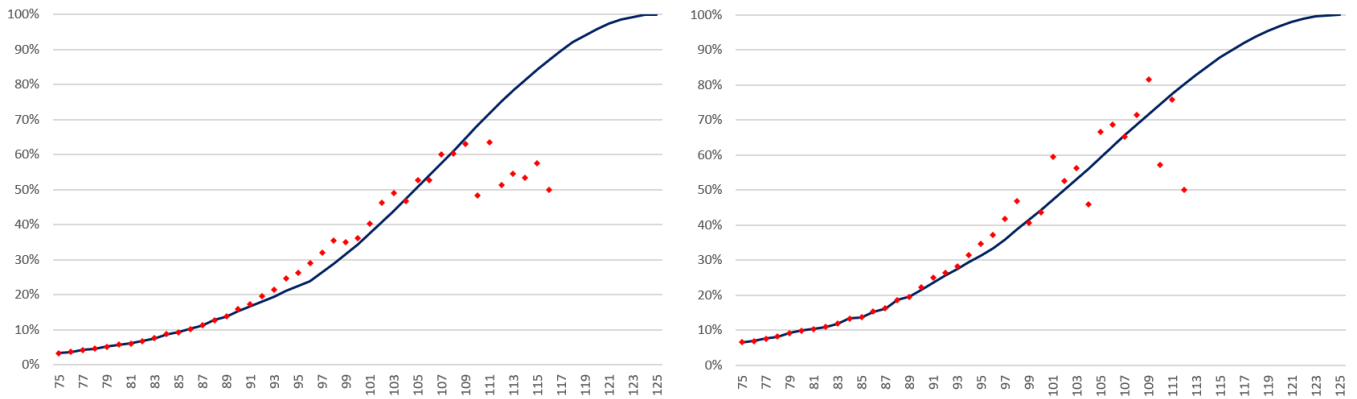


FIGURE 5.8 – *Ajustement des taux de mortalité via la méthode de Denuit & Goderniaux pour les femmes (à gauche) et les hommes (à droite) de la génération 1905*

Nous remarquons que ce modèle semble bien ajuster les taux bruts de mortalité, notamment chez les hommes. Pour les femmes, la mortalité semble surestimée à partir de 108 ans. Par ailleurs, après lissage via une moyenne pondérée au niveau de l'âge charnière, il est possible de remarquer un léger changement de tendance, notamment pour la population féminine.

Dans le cas où x_{max} vaut 130 ans et non plus 125 ans, les résultats sont les suivants :

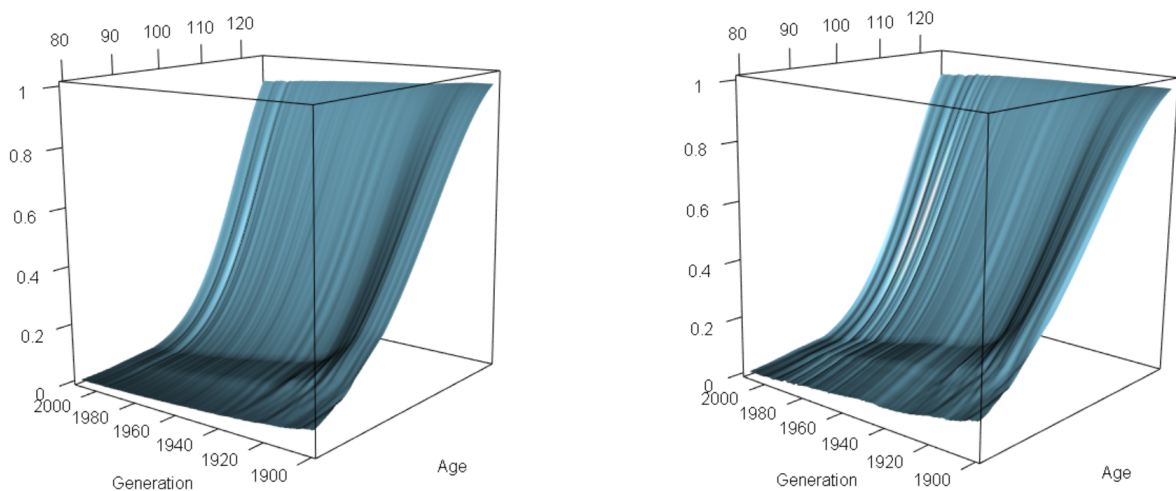


FIGURE 5.9 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Denuit & Goderniaux avec x_{max} valant 130 ans*

Nous constatons que, malgré un âge charnière moins marqué que dans le cas précédent, il reste quand même bien visible malgré le lissage par moyennes pondérées. Cependant, repousser encore l'âge maximal n'est pas forcément réaliste. Pour cette rai-

son nous ne testerons l'évolution de cette limite au fil du temps. Faire cette hypothèse – afin d'obtenir un prolongement naturel – conduirait probablement à des âges limites autour de 150 ou 160 ans pour les cohortes les plus jeunes.

5.1.4 Modèle logistique

- **Considérations théoriques**

On cherche, pour ce modèle paramétrique, à ce que la courbe de la force de mortalité ait la forme d'une sigmoïde. Les propriétés de ce type de fonction sont très intéressantes pour l'estimation de la mortalité. En effet, nous pouvons retrouver une forte croissance pour les âges dits intermédiaires, et une croissance plus faible près des âges les plus bas et les plus élevés.

La force de mortalité se définit comme suit :

$$\mu_x = c + \frac{a \exp(bx)}{1 + \sigma^2 \frac{a}{b} (\exp(bx) - 1)}$$

Le paramètre c joue un rôle « d'offset » qui peut être utile pour certaines populations, ce qui n'est a priori pas notre cas, puisque la mortalité est proche de 0 pour certains âges. Afin que la force de mortalité reste entre 0 et 1, il nous faut également introduire les contraintes suivantes : $\sigma^2 a \leq b$ et $a \leq 1$.

Il existe plusieurs variantes de ce modèle, dont le modèle de Bongaarts, que nous ne développerons pas ici, mais qui est développé dans KAMEGA et PLANCHET (2011). La variante que nous allons utiliser est le modèle de Kannistö, une autre forme particulière de ce modèle. Le nombre de paramètres à estimer est donc réduit à deux. Pour cela, on considère les valeurs $c = 0$ et $\sigma^2 = b$. De plus, $a \leq 1$ et b doit être positif. Cela signifie, qu'il n'y a plus d'offset, et que la variance de la distribution sous-jacente considérée est constante.

Le modèle finalement utilisé sera donc le suivant :

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

L'estimation des paramètres se fait en minimisant l'écart pondéré entre le modèle et la mortalité observée, ce qui signifie que :

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b) \in \mathbb{R}} \sum_x \frac{(\hat{\mu}_x - \mu_x(a,b))^2}{L_x}$$

- **Considérations pratiques**

Après le passage de la force de mortalité aux taux de mortalité, les résultats obtenus avec ce modèle sont les suivants :

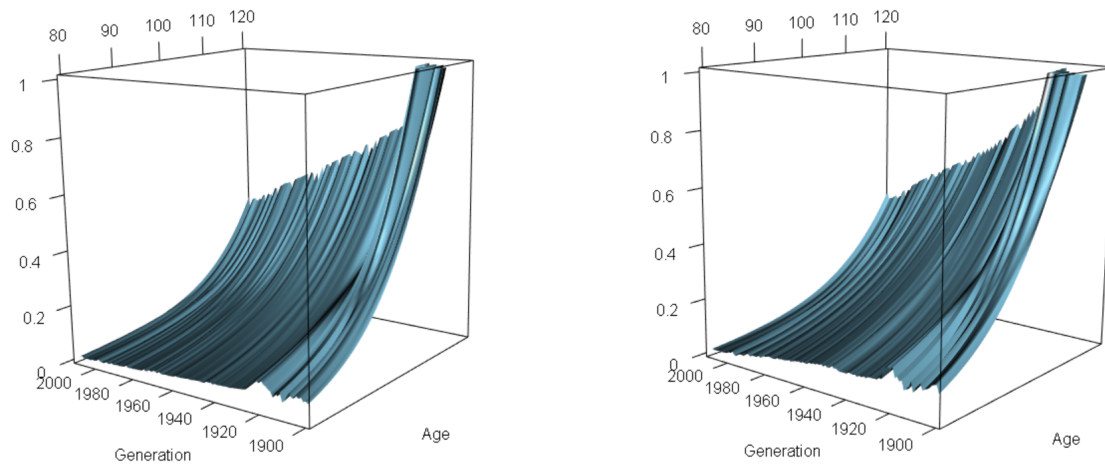


FIGURE 5.10 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Kannistö*

Nous constatons, d'une part, une légère volatilité horizontale, c'est-à-dire par rapport aux générations. En effet, les taux de mortalité varient d'une génération à la suivante, notamment autour des générations 1900-1910 avant de devenir plus lisses. Il aurait été possible de la lisser comme pour le modèle de Lindbergson. Cependant, nous considérons que cela n'est pas nécessaire étant donné que celle-ci reste finalement assez limitée. On remarque également que les taux de mortalité sont plus élevés que pour la méthode des P-splines par exemple, en particulier pour les générations antérieures à 1950. Une forte dérive de longévité est prévu par ce modèle.

L'ajustement aux taux bruts de mortalité reste correct :

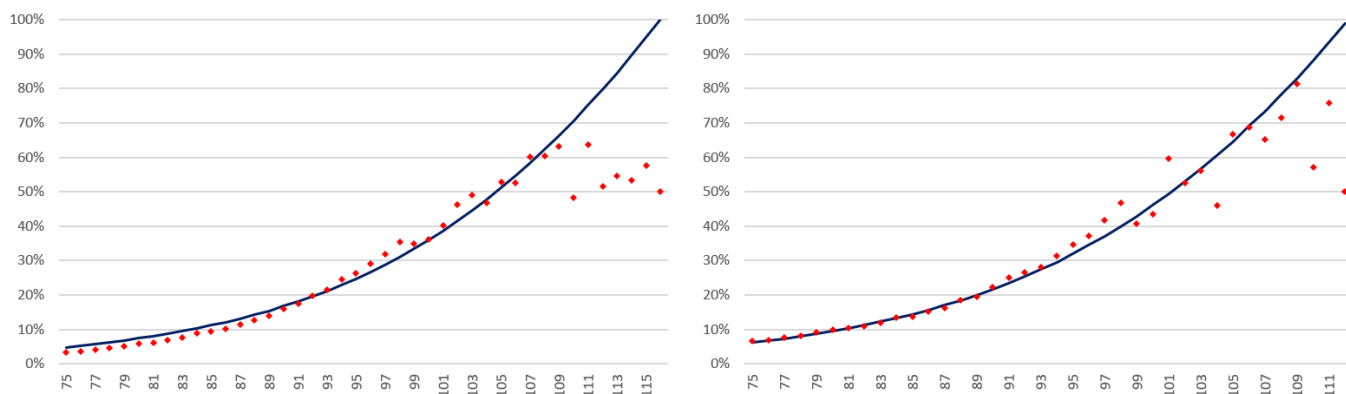


FIGURE 5.11 – *Ajustement des taux de mortalité via le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) de la génération 1905*

L’ajustement des taux bruts est satisfaisant et naturel. La mortalité est, encore une fois, surestimée pour les taux correspondant à des âges supérieurs à 109 ans. De ce fait, certains taux de mortalité prévus par le modèle sont supérieurs à 100% malgré la présence de survivants.

En résumé, nous retenons que les modèles paramétriques sont de bons outils de modélisation qui permettent d’extrapoler la mortalité aux grands âges. Le principal inconvénient de ce type de méthode est la nécessité de connaître la forme générale des taux à ajuster, ou, du moins, d’en faire l’hypothèse. Un ajustement par le modèle des P-splines présuppose, une forme polynomiale par morceaux, le modèle de Lindbergson une forme linéaire, Denuit & Goderniaux pensent, quant à eux, que le logarithme du taux de mortalité est quadratique, tandis que Kannistö opte pour une sigmoïde. Les paramètres ont été estimés via la méthode des moindres carrés ordinaires pondérés par l’exposition. D’après nos calculs, tous ces modèles ont l’air d’ajuster assez bien les taux bruts avant 109 ans, mais s’écartent de ceux-ci à partir de cet âge, excepté le modèle des P-splines.

Afin de ne plus poser de contrainte sur la forme des taux de mortalité et de se libérer de cette nécessité, il faut se tourner vers des modèles non-paramétriques qui ne font pas ce type d’hypothèse, et ajustent la mortalité via des méthodes alternatives.

5.2 Modèles non-paramétriques

Les modèles non-paramétriques ont l’avantage de ne pas faire d’hypothèse sur la forme de la courbe des taux de mortalité. Ce genre de modèle est donc utile lorsque l’on n’a pas d’idée sur ce point. Ces modèles laissent ainsi une plus grande liberté

ainsi qu'une plus grande adaptabilité que les modèles paramétriques. Cependant, ils dépendent plus de la qualité des données brutes et d'éventuelles valeurs aberrantes.

Les méthodes non-paramétriques ont également le désavantage de fournir une estimation de la fonction de survie non dérivable. Cela pose un problème, notamment lors de l'estimation du taux de risque instantané, pour lequel nous avons besoin de cette quantité.

5.2.1 Modèle de Whittaker-Henderson

- **Considérations théoriques**

Le principe de cette méthode est d'ajuster aux données une courbe alliant à la fois fidélité et régularité.

Bien que nous nous intéressions à un problème en deux dimensions, il est utile de développer la méthode en dimension 1 avant de l'étendre.

En dimension 1, le modèle s'écrit comme suit :

Le critère de fidélité s'écrit de la manière suivante :

$$F = \sum_{x=x_0}^{x_0+n} \omega_x (\hat{q}_x - q_x)^2$$

Avec x_0 le premier âge où l'on ajuste les taux, et $n + 1$ le nombre d'âges concernés par l'ajustement.

Il s'agit simplement d'une somme pondérée des erreurs quadratiques.

Le critère de régularité s'énonce comme suit :

$$R = \sum_{x=x_0}^{x_0+n} (\Delta^z q_x)^2$$

Avec z un paramètre du modèle et $\Delta q_x = q_{x+1} - q_x$. Δ^z est la composition de z opérateur Δ .

On cherche donc à minimiser la quantité :

$$Q = hF + (1 - h)R$$

Avec $h \in [0, 1]$, le poids permettant de donner plus d'importance à l'un des deux critères par rapport à l'autre.

Écrit sous la forme matricielle, cela revient à l'égalité :

$$Q = h \times {}^t(q - \hat{q})\omega(q - \hat{q}) + (1 - h) \times {}^tq^t K_z K_z q$$

Avec q (respectivement \hat{q}) la matrice colonne des q_x (respectivement \hat{q}_x), ω une matrice diagonale avec les poids ω_x comme coefficients diagonaux et K_z la matrice de taille $(n+1-z, n+1)$ tel que $\Delta^z q = K_z q$.

Par définition de Δ , nous pouvons déduire que les termes de la matrice K_z sont les coefficients binomiaux d'ordre z , dont le signe alterne et commence positivement pour z pair et négativement pour z impair.

La quantité Q étant convexe, comme somme de fonctions convexes, et dérivable par rapport aux taux de mortalité, on a :

$$\frac{\partial Q}{\partial q} = 0, \forall x \in [x_0, x_0 + n]$$

Donc :

$$\begin{aligned} \frac{\partial Q}{\partial q} &= 2h\omega q - 2h\omega\hat{q} + 2(1 - h) \times {}^t K_z K_z q = 0 \\ q (h\omega + (1 - h) \times {}^t K_z K_z) - h\omega\hat{q} &= 0 \end{aligned}$$

La matrice $(h\omega + (1 - h) \times {}^t K_z K_z)$ est inversible. Ce résultat n'est pas trivial et nécessite que l'on s'y attarde. En effet, on peut montrer, comme précisé dans PLANCHET (2021a), que la matrice $(1 - h) \times {}^t K_z K_z$ n'est généralement pas inversible. Cependant, en ajoutant le terme de poids $h\omega$, le déterminant de la matrice devient non-nul et, par conséquent elle devient donc inversible. Cela est dû au fait que la matrice $h\omega$ contient des éléments non nuls sur sa diagonale. Le taux de mortalité q vaut alors :

$$q = (h\omega + (1 - h) \times {}^t K_z K_z)^{-1} \omega \hat{q}$$

Il est possible de généraliser ce raisonnement en dimension 2 voire en dimension supérieure :

Le critère de fidélité est équivalent à celui en dimension 1, en sommant également sur la génération. Cela signifie que les taux q_x (respectivement \hat{q}_x) se transforment en $q_{x,t}$ (respectivement $\hat{q}_{x,t}$), et la matrice colonne q (respectivement \hat{q}) se transforme en matrice de taille $n \times g$ avec n le nombre d'âges considérés et le g le nombre de générations considérées.

Le critère de régularité se divise quant à lui en deux sous-critères, l'un correspondant au critère de régularité par rapport à l'âge et l'autre par rapport aux cohortes.

D'où :

$$\begin{aligned} R &= \alpha R_a + \beta R_g \\ &= \alpha \sum_{x=x_0}^{x_0+n} \sum_{t=t_0}^{t_0+g-z} (\Delta^z q_{x,t})^2 + \beta \sum_{x=x_0}^{x_0+n-y} \sum_{t=t_0}^{t_0+g} (\Delta^y q_{x,t})^2 \end{aligned}$$

Avec t_0 la première cohorte considérée, z l'indice de régularité par âge, y l'indice de régularité par génération et, α et β des coefficients permettant de privilégier la régularité selon l'âge plutôt que selon la cohorte ou inversement.

Q vaut donc :

$$\begin{aligned} Q &= hF + (1-h)R \\ &= hF + (1-h)\alpha R_a + (1-h)\beta R_g \end{aligned}$$

Par analogie avec le cas en dimension 1, il est possible de définir également K_z^{age} et K_y^{gen} . On peut aussi déterminer u un vecteur de taille $n \times g$ tel que $u_{g(i-1)+j} = q_{x_i, t_j}$, et la matrice de poids Ω .

En dérivant et en annulant la dérivée de Q par rapport à q_x (de la même manière qu'en dimension 1), on obtient les taux ajustés suivants :

$$q = (h\Omega + (1-h)\alpha^t K_z^{age} K_z^{age} + (1-h)\beta^t K_y^{gen} K_y^{gen})^{-1} \Omega u$$

PLANCHET (2021a) précise que ce modèle peut également être vu comme un modèle bayésien.

5.2.2 Modèle de Whittaker-Henderson vu comme un ajustement bayésien

• Considérations théoriques

Avant de développer ce modèle en utilisant cette approche bayésienne, nous allons commencer par un rappel rapide de quelques notions de la statistique bayésienne. Le principe de ce type de modèle est de proposer une loi a priori qui nous permettra d'ajuster la loi a posteriori. On rappelle la formule de Bayes :

$$f_{q|\hat{q}}(q|\hat{q}) = \frac{f_{\hat{q}|q}(\hat{q}|q)f_q(q)}{f_{\hat{q}}(\hat{q})}$$

Afin de comprendre le modèle précédent à l'aide d'une approche bayésienne, nous utilisons la mesure de régularité R pour définir la loi a priori f_q .

On pose alors la densité a priori comme étant une loi exponentielle de paramètre λ . Cette loi est cohérente avec l'idée qu'avec la croissance de l'âge, la régularité décroît et que la densité de probabilité augmente, ce qui signifie que le taux de mortalité croît. Il s'agit donc d'une loi qui semble convenir à la durée de vie d'un individu, et qui, de plus, est beaucoup utilisée en modèle de durée. La loi a priori s'écrit donc comme suit :

$$f_q(q) \propto \exp(-\lambda R)$$

De plus, en faisant l'hypothèse que les résidus $\hat{q}_x - q_x$ suivent une loi normale centrée et dont la variance est notée σ_x^2 , sous une hypothèse d'indépendance entre les différents âges, et en considérant $n + 1$ âges, on obtient :

$$f_{\hat{q}|q}(\hat{q}|q) \propto \exp\left(-\frac{1}{2} \sum_{k=x_0}^{x_0+n} \left(\frac{\hat{q}_k - q_k}{\sigma_k}\right)^2\right)$$

Ces deux quantités suffisent à définir la distribution a posteriori étant donné que $f_{\hat{q}}(\hat{q})$, ne dépendant pas de q , agit comme une constante de normalisation. Il est possible de la retrouver si besoin puisque l'on sait que $\int f_{q|\hat{q}}(q|\hat{q}) = 1$ par définition d'une densité de probabilité.

On obtient donc la distribution a posteriori suivante :

$$f_{q|\hat{q}}(q|\hat{q}) \propto \exp\left(-\lambda R - \frac{1}{2} \sum_{k=x_0}^{x_0+n} \left(\frac{\hat{q}_k - q_k}{\sigma_k}\right)^2\right)$$

Il s'agit également de la vraisemblance du modèle, que l'on cherche à maximiser. Cela revient à minimiser la quantité :

$$\lambda R + \frac{1}{2} \sum_{k=x_0}^{x_0+n} \left(\frac{\hat{q}_k - q_k}{\sigma_k}\right)^2$$

On retrouve alors les mêmes résultats que dans la section précédente étant donné que la quantité à minimiser correspond à Q .

- **Considérations pratiques**

Les résultats de l'approche de Whittaker-Henderson sont les suivants :

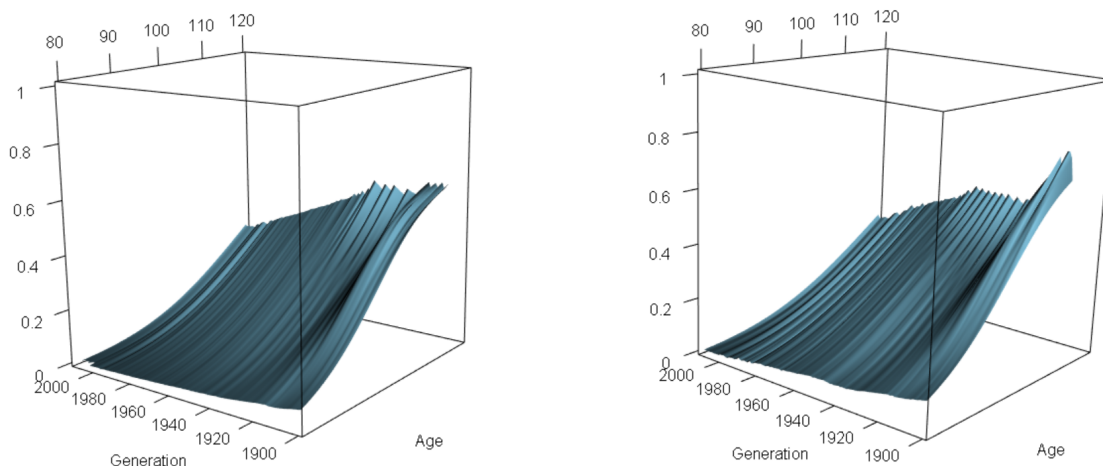


FIGURE 5.12 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Whittaker-Henderson*

On note que les taux sont assez lisses par rapport aux deux composantes. Avec cette méthode et le paramètre de régularité ajusté, les taux de mortalité montrent une évolution assez intuitive, que ce soit en fonction de l'âge ou de la génération.

Observons maintenant l'adéquation aux taux bruts :

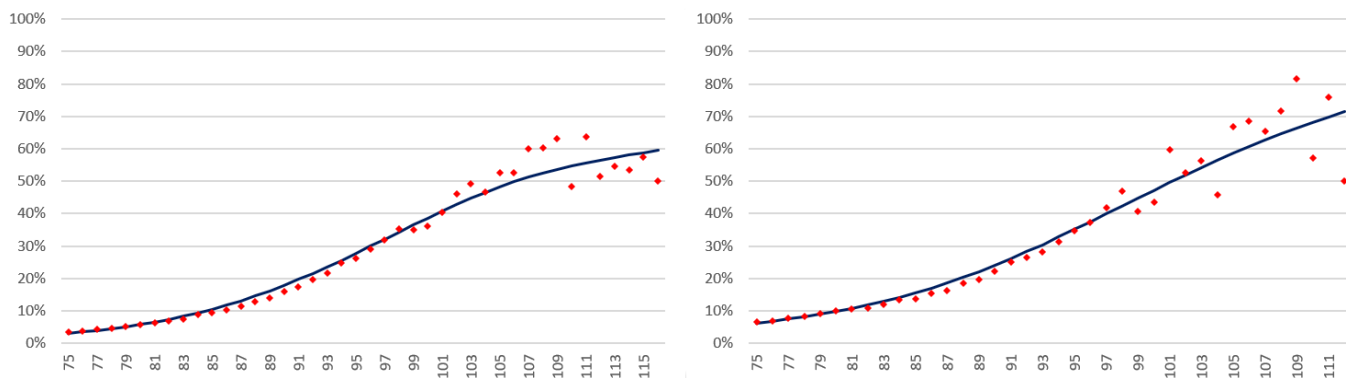


FIGURE 5.13 – *Ajustement des taux de mortalité via le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite) de la génération 1905*

Nous remarquons une bonne adéquation des taux ajustés aux taux bruts. Il s'agit du premier modèle, après celui des P-splines, à ne pas surestimer les taux bruts aux très grands âges (après 109 ans). L'ajustement se fait sans discontinuité, contrairement à d'autres modèles comme celui de Denuit & Goderniaux ou de Lindebergson, qui a nécessité un ajustement supplémentaire.

Le fait de ne pas présupposer de forme pour le taux de mortalité est un avantage certain. Cela permet, notamment, d'avoir une allure légèrement concave pour les centenaires. Cela n'est pas le cas avec certains modèles paramétriques, qui imposent une évolution donnée, quitte à s'éloigner des taux bruts. Cette méthode présente néanmoins le désavantage d'être plus sensible à des points potentiellement aberrants. Ce phénomène est, en particulier, observable en cas de petit échantillon, où chaque décès peut avoir un fort impact sur les taux bruts de mortalité.

5.2.3 Modèles relationnels

Les modèles relationnels représentent un autre type de modèle non-paramétrique. Une telle approche se fonde sur une référence externe, et ces modèles font donc partie des méthodes d'ajustement exogènes contrairement à ceux que nous avons vu jusqu'à présent. Ces modèles consistent à évaluer l'écart entre la table que l'on cherche à construire et la table de référence, et ainsi la positionner par rapport à cette dernière.

Ce sont, en général, des modèles assez simples mais non pas efficaces si la population assurée est assez proche de la population de référence. Dans notre cas, la table à ajuster se fonde sur la population nationale, ce qui nous donne de bons espoirs quant à la fiabilité de ce type de méthode. En effet, cela nous permettra d'utiliser les tables réglementaires, également basée sur la population nationale, comme références.

Modèle de Brass

- **Considérations théoriques**

Le modèle de Brass, modèle relationnel, consiste à modéliser le $\log it$ de la mortalité comme étant une transformation affine du $\log it$ d'une mortalité de référence. Le modèle s'écrit donc :

$$\log it(q_x) = a \times \log it(q_x^{ref}) + b$$

Nous cherchons à optimiser les paramètres par la méthode des moindres carrés ordinaires. On minimise donc la quantité :

$$Q = \sum_x L_x (q_x - q_x^{ref})^2$$

Cette méthode est ainsi assez proche de la méthode utilisée pour obtenir la mortalité prospective. La différence réside dans le fait que les paramètres a et b ne dépendent, ici, pas de l'âge. Par conséquent, la quantité à minimiser diffère également.

Les taux de référence que nous utiliserons sont les taux issus des tables réglementaires TGH et TGF 05.

- **Considérations pratiques**

Les résultats obtenus avec le modèle de Brass sont les suivants :

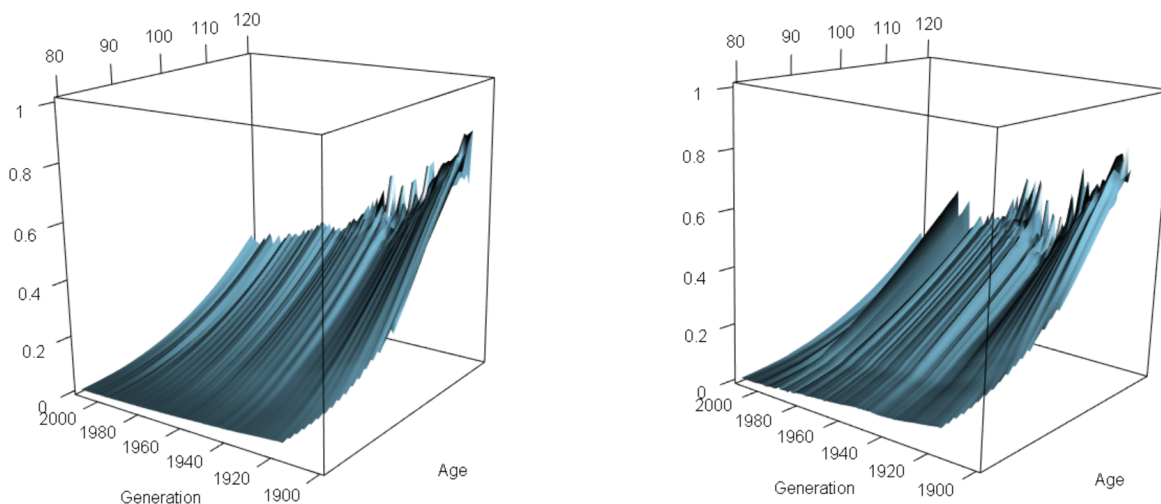


FIGURE 5.14 – *Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Brass*

Tout d'abord, nous remarquons que les taux de mortalité à certains âges et pour certaines générations, ne sont pas donnés. Cela est le cas pour les âges inférieurs à 91 ans pour la génération 1905. Cela est dû au fait que la table de référence ne donne pas non plus ces taux. Cela n'est pas un problème en soi, étant donné que l'objectif est d'obtenir la mortalité des plus âgés, pour lesquels nous avons bien les taux de mortalité.

Les taux obtenus sont assez lisses, mais semblent sous-estimés pour les cohortes les plus jeunes, avec des taux de mortalité autour de 35% à plus de 110 ans. Cela signifie que les chances de survie sont supérieures à 65% ce qui reste élevé, mais cela n'est pas en contradiction avec les taux bruts obtenus à la section 4.5. Cela nous laisse donc penser que le modèle ajuste correctement les taux bruts. De plus, les table de référence TGH et TGF 05 ont des taux de mortalité assez similaires, voire plus faibles, pour ces projections.

Regardons maintenant l'ajustement du modèle aux taux bruts :

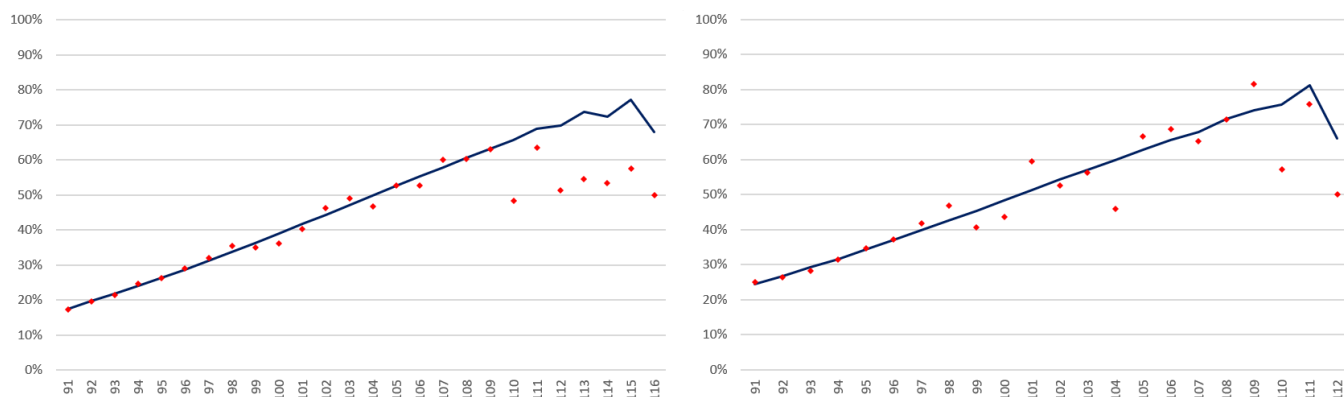


FIGURE 5.15 – *Ajustement des taux de mortalité via le modèle de Brass pour les femmes (à gauche) et les hommes (à droite) de la génération 1905*

Pour la génération 1905, la table nous donne seulement les taux de mortalité pour les personnes âgées de plus de 91 ans, d'où l'abscisse plus restreinte sur cette Figure que sur celles des modèles précédents. La méthode d'ajustement se révèle être de bonne qualité. Il reste une légère volatilité, en particulier chez les femmes après 113 ans.

Nous constatons la baisse du taux de mortalité à partir d'un certain âge. Cela rejoint les conclusions tirées du modèle des P-splines.

En conclusion de ce Chapitre, nous pouvons retenir qu'il existe de nombreuses méthodes de fermetures de tables de mortalité. Certaines, paramétriques, présupposent la forme des taux de mortalité ce qui constitue une limite à ce type de modèles. Ils sont cependant beaucoup utilisés, et semblent convenir dans notre cas en fournissant, généralement, une bonne adéquation aux taux bruts. Les modèles non paramétriques permettent, quant à eux, de s'adapter aux taux bruts tout en produisant un lissage permettant d'éviter la volatilité aux grands âges. Ils sont cependant très dépendant de la qualité des données en entrée, en comparaison des modèles paramétriques. Enfin, les modèles relationnels, qui sont non-paramétriques, constituent une troisième alternative ; ils consistent à évaluer l'écart entre la population étudiée et une population de référence. Cela permet de produire un lissage, en général, à moindre coût et de bonne qualité. Dans notre cas, les modèles non-paramétriques fournissent de bons résultats du fait de la bonne qualité des données en entrée.

D'autres type d'approches exogènes peuvent être mises en place, avec l'application d'un coefficient d'abattement ou des méthodes de décalage par âge par exemple.

Certains modèles comme le modèle de Brass, de Whittaker-Henderson et des P-splines font l'hypothèse d'une possible baisse de la mortalité à partir d'un certain âge tandis que les trois autres modèles développés ne permettent pas d'avoir cette baisse du

fait de la forme de la courbe de taux supposée. Bien que les ajustements paraissent, à première vue, tous convenir et être en bonne adéquation aux taux bruts sur la majorité des âges, il est important d'évaluer les modèles développés via différentes méthodes statistiques.

Chapitre 6

Erreur de modèle

Maintenant que nous avons ajusté les différents modèles aux grands âges, nous avons besoin d'outils et de méthodes pour évaluer leurs fiabilités et les comparer entre eux. Ce Chapitre a pour objectif de développer ces outils.

Pour tester la fiabilité des modèles, les techniques utilisées sont assez classiques en statistiques, et permettent de mesurer le degré de confiance que l'on peut leur donner.

6.1 Tests de cohérence

Un simple test de cohérence peut dégager des premières impressions sur les modèles utilisés. Pour réaliser ce test nous vérifions la cohérence des taux de mortalité à certains âges donnés.

La ligne « Taux de référence » correspond aux taux de mortalité obtenus avec les tables de référence TGH-TGF 05.

	Génération 1905			Génération 1955			Génération 2005		
	<i>100 ans</i>	<i>110 ans</i>	<i>120 ans</i>	<i>100 ans</i>	<i>110 ans</i>	<i>120 ans</i>	<i>100 ans</i>	<i>110 ans</i>	<i>120 ans</i>
Taux de référence	26,1%	47,8%		16,3%	31,5%		8,9%	20,2%	
Taux bruts	36,3%	48,3%		17,6%	32,5%		6,5%	17,1%	
P-splines	36,8%	57,7%		16,3%	30,9%		5,8%	15,8%	
Lindbergson	42,3%	76,6%		19,6%	36,3%	56,5%	7,4%	17,9%	29,6%
Denuit&Goderniaux	34,6%	68,2%	95,8%	14,9%	50,4%	92,7%	5,4%	35,0%	89,0%
Kannistö	36,0%	70,6%		15,9%	32,9%	63,4%	5,6%	15,5%	39,4%
Whittaker-Henderson	38,6%	54,8%		17,7%	32,6%		7,0%	17,5%	
Brass	39,0%	65,7%		16,4%	29,9%		5,4%	13,7%	

FIGURE 6.1 – Cohérence des taux bruts femmes à 100, 110 et 120 ans pour les générations 1905, 1955, et 2005

	Génération 1905			Génération 1955			Génération 2005		
	100 ans	110 ans	120 ans	100 ans	110 ans	120 ans	100 ans	110 ans	120 ans
Taux de référence	34,7%	60,0%		21,1%	36,8%		11,6%	22,9%	
Taux bruts	43,6%	57,2%		25,1%	38,9%		9,0%	20,0%	
P-splines	45,8%	67,5%		23,7%	37,5%		8,2%	18,7%	
Lindbergson	54,4%			26,6%	43,9%	64,9%	9,5%	21,3%	34,6%
Denuit & Goderniaux	44,3%	74,6%	96,8%	21,9%	57,9%	94,1%	7,8%	39,9%	90,3%
Kannistö	46,0%	88,1%		23,4%	40,0%	65,0%	7,9%	18,8%	41,9%
Whittaker - Henderson	47,2%	68,1%		24,8%	38,9%		9,4%	20,4%	
Brass	48,4%	75,7%		24,3%	37,6%		7,9%	17,1%	

FIGURE 6.2 – *Cohérence des taux bruts hommes à 100, 110 et 120 ans pour les générations 1905, 1955, et 2005*

Les cases grisées correspondent à des couples âge/génération donnant un taux de mortalité supérieurs à 100% ou non défini. Cela signifie que le modèle ne projette déjà plus aucun survivant à un âge inférieur, et donc, a fortiori, pour cet âge.

Les taux de mortalité obtenus à ces grands âges nous semblent assez optimistes pour les générations les plus jeunes avec des taux de mortalité autour de 20% à 110 ans. Cependant, le fait que les taux de mortalité soient aussi bas à ces âges est dû au fait que les taux prospectifs sont également bas, du fait de la méthode de projection choisie. Nous remarquons d'ailleurs que les taux de référence sont du même ordre.

À ce stade, une autre manière de faire est de comparer l'écart des ajustements aux taux bruts. Toutes les méthodes donnent des résultats assez satisfaisants en moyenne. Néanmoins les modèles de Whittaker-Henderson et de Brass semblent les plus proches. Le modèle de Denuit & Goderniaux surestime les taux bruts en raison de la contrainte à l'âge maximal.

6.2 Standardized Mortality Ratio (SMR)

Cette méthode d'évaluation des modèles de mortalité est définie par l'INED comme étant « le ratio entre le nombre de décès observés dans une population sur une période donnée et le nombre de décès attendu dans la population étudiée sur la même période. Si le ratio est supérieur à 1 cela est interprété comme une surmortalité de la population observée ». Dans le cas où le SMR est inférieur à 1, on l'interprète comme une sous-mortalité dans la population observée.

6.2. STANDARDIZED MORTALITY RATIO (SMR)

Ce ratio se calcule donc comme :

$$SMR = \frac{\sum_x D_x^{observé}}{\sum_x D_x^{modèle}}$$

Il est intéressant de faire le calcul de SMR par génération et/ou par classe d'âge.

De plus, ce ratio peut sert également à la mise en place d'un modèle relationnel. Cette méthode consiste à appliquer le ratio SMR comme coefficient d'abattement à une mortalité de référence. Cette approche a été développée, entre autres, dans CAPPE (2018) mais ne sera pas développée ici.

Les résultats des ratios SMR sont les suivants :

	Tranche d'âge	1900	1925	1950	1975	2000
P-splines	80 - 100	1,001	1,002	1,013	1,027	1,050
	101 - 120	0,898	0,971	0,938	0,943	0,960
Lindbergson	80 - 100	0,994	0,994	0,999	1,405	0,936
	101 - 120	2,335	1,182	1,017	0,611	1,065
Denuit & Goderniaux	80 - 100	1,004	1,044	1,103	1,193	1,340
	101 - 120	0,678	0,600	0,674	0,733	0,817
Kannistö	80 - 100	1,002	1,023	0,999	0,980	0,773
	101 - 120	0,596	1,270	1,120	1,092	1,413
Whittaker-Henderson	80 - 100	1,000	1,000	0,988	0,972	0,949
	101 - 120	1,390	1,059	1,067	1,066	1,045
Brass	80 - 100		1,003	1,013	1,019	1,026
	101 - 120	1,015	0,974	0,942	0,961	0,978

FIGURE 6.3 – *Standard Mortality Ratio à différents âges et générations pour les modèles testés chez les femmes*

	Tranche d'âge	1900	1925	1950	1975	2000
P-splines	80 - 100	1,001	1,005	1,011	1,023	1,064
	101 - 120	0,524	0,777	0,803	0,847	0,905
Lindbergson	80 - 100	1,001	1,010	1,005	0,994	1,744
	101 - 120	3,863	0,829	0,995	1,086	0,575
Denuit & Goderniaux	80 - 100	1,001	1,012	1,037	1,084	1,235
	101 - 120	0,644	0,569	0,563	0,614	0,752
Kannistö	80 - 100	0,989	1,026	1,025	1,010	0,974
	101 - 120	0,518	0,965	1,131	1,197	1,086
Whittaker-Henderson	80 - 100	1,005	1,005	1,003	0,998	0,976
	101 - 120	1,298	1,073	1,059	1,053	1,044
Brass	80 - 100		1,001	1,007	1,003	1,039
	101 - 120	1,001	0,949	0,877	0,976	0,938

FIGURE 6.4 – *Standard Mortality Ratio à différents âges et générations pour les modèles testés chez les hommes*

Le modèle de Brass ne fournit pas de résultats pour la génération 1900 sur la première tranche d'âge, étant donné que le nombre de décès obtenu est biaisé par le fait que la table ne commence à compter le nombre de décès qu'à partir de 96 ans (du fait de la table de référence utilisée), et donc les expositions ne sont pas sur le même périmètre.

Nous remarquons que le modèle de Denuit & Goderniaux fournit les nombres de décès, en moyenne, les plus éloignés des nombres de décès observés et les surestime. Selon ce modèle il y aurait donc une forte sous-mortalité, en moyenne, dans la population étudiée. Cela est vrai notamment sur les âges supérieurs à 101 ans étant donné les ratios autour de 0,6. C'est une conséquence de la contrainte mise à l'âge maximal fixé à 125 ans. Les trois modèles qui présentent des résultats les plus proches des nombres de décès observés sont ceux des P-splines, de Whittaker-Henderson ainsi que celui de Brass. Le modèle de Lindbergson fournit des résultats mitigés, dans le sens où l'adéquation aux nombres de décès bruts est soit très bonne, soit très mauvaise (génération 1900 sur la tranche d'âge 101-120 ans par exemple). Le modèle de Kannistö fournit des résultats convenables malgré quelques taux de mortalité qui sont fortement surestimés par rapport aux taux bruts (notamment sur la génération 1900).

On remarque également, de manière générale, une meilleure adéquation des nombres de décès ajustés aux décès bruts sur la tranche d'âge 80-100 ans que sur les âges suivants, notamment pour les modèles paramétriques. Nous pouvons nous y attendre, étant donné la très forte volatilité présente pour cette deuxième tranche d'âge, ce qui rend les taux compliqués à ajuster, et crée forcément des écarts.

On peut aussi constater que les modèles ajustent les taux de mortalité, en moyenne, aussi bien d'une génération à une autre.

Le SMR est une quantité intéressante assez simple à appréhender et à calculer. On peut également s'intéresser à une mesure d'écart sur les taux de mortalité et non plus sur les nombres de décès. Il est pertinent de se tourner vers la quantité MAPE.

6.3 Mean Absolute Percentage Error (MAPE)

Il s'agit principalement d'une méthode utilisée pour mesurer une erreur de prédiction en statistique. Cette quantité se calcule comme :

$$MAPE = \frac{100}{x_{max} - x_0 + 1} \sum_{x=x_0}^{x_{max}} \left| \frac{q_x^{obs} - q_x^{modèle}}{q_x^{obs}} \right|$$

Elle permet de mesurer l'écart relatif des taux ajustés aux taux bruts observés.

6.3. MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

Une meilleure adéquation des taux se traduit par une MAPE faible. Au contraire, si cette quantité est élevée cela signifie que les taux bruts sont éloignés des taux prévus par le modèle. Cette statistique ne permet cependant pas de savoir si le modèle prévoit une surmortalité ou une sous mortalité moyenne.

Comme précédemment, pour le SMR, il est intéressant de le calculer par tranche d'âge et par génération. Les résultats obtenus sont les suivants :

	Tranche d'âge	1900	1925	1950	1975	2000
P-splines	80 - 100	0,970	0,652	1,660	2,233	2,699
	101 - 120	6,296	5,111	4,598	5,071	6,202
Lindbergson	80 - 100	11,285	9,107	6,668	35,243	10,545
	101 - 120	18,755	12,991	6,709	18,453	8,003
Denuit & Goderniaux	80 - 100	4,267	11,424	13,121	15,591	18,162
	101 - 120	17,560	25,264	43,215	60,273	79,382
Kannistö	85 - 100	10,587	30,004	30,642	45,429	190,866
	101 - 120	17,916	5,635	6,814	8,609	20,665
Whittaker-Henderson	80 - 100	6,756	8,304	6,946	9,967	15,520
	101 - 120	6,750	2,024	0,992	1,056	2,234
Brass	80 - 100		4,863	2,791	2,236	2,229
	101 - 120	15,998	0,620	5,933	4,463	4,993

FIGURE 6.5 – Mean Percentage Absolute Error à différents âges et générations pour les modèles testés chez les femmes

	Tranche d'âge	1900	1925	1950	1975	2000
P-splines	80 - 100	6,453	4,786	5,327	5,940	7,977
	101 - 120	8,713	3,833	2,737	3,169	4,514
Lindbergson	80 - 100	11,156	6,564	4,640	4,474	51,736
	101 - 120	14,423	10,531	12,443	8,932	17,124
Denuit & Goderniaux	80 - 100	3,294	8,401	11,538	13,980	17,280
	101 - 120	13,334	11,074	35,243	53,684	74,337
Kannistö	80 - 100	9,739	13,251	18,514	27,656	44,627
	101 - 120	14,309	2,820	7,139	6,050	7,108
Whittaker-Henderson	80 - 100	4,665	5,309	5,144	5,972	6,824
	101 - 120	9,869	1,419	1,819	0,449	1,537
Brass	80 - 100		3,260	2,940	0,827	4,422
	101 - 120	10,999	1,052	5,288	1,237	8,288

FIGURE 6.6 – Mean Percentage Absolute Error à différents âges et générations pour les modèles testés chez les hommes

On ne donne pas les résultats obtenus via le modèle de Brass pour la génération 1900 car nous ne pouvons calculer l'écart aux taux bruts que pour les âges postérieurs à 96 ans. La tranche d'âge ne serait donc pas respectée, ce qui aurait pour effet de biaiser le résultat obtenu.

L'adéquation aux taux bruts semble légèrement moins bonne d'une génération à la suivante, notamment pour les modèles de Denuit & Goderniaux et de Kannistö. Les résultats vont dans le même sens que ceux obtenus avec le SMR. En effet, nous constatons que les modèles des P-splines, de Whittaker-Henderson ainsi que celui de Brass fournissent, là encore, les résultats les plus proches de la table des taux bruts. Le modèle de Lindbergson fournit lui aussi des résultats assez proches des taux bruts de manière générale.

Contrairement à l'évaluation des écarts avec les SMR, la première tranche d'âge n'est plus forcément celle sur laquelle les taux sont les plus proches des taux bruts. Cela dépend des modèles et des générations. Nous constatons d'ailleurs que le modèle de Kannistö prévoit des taux plus proches des taux bruts sur la tranche d'âges la plus âgée.

6.4 Tests de sensibilité

Cette section a pour objectif de mettre en évidence l'importance respective des paramètres, et les conséquences de leurs variations sur le taux de mortalité. Pour les modèles relationnels, ce type de tests peut également porter sur la table de référence choisie. Afin de pouvoir tirer des conclusions par paramètre, il est important de tester chaque paramètre indépendamment les uns des autres.

- **P-splines**

En ce qui concerne la méthode des P-splines, ce test porte sur le degré des polynômes utilisés ainsi que sur les noeuds choisis. Nous testerons les cas suivants :

- Le degré des splines utilisées passe de 4 à 3 dans un premier temps puis à 5.
- L'utilisation de noeuds à 90 ans et à 100 ans, puis l'utilisation de nouveaux noeuds à 95 ans et 105 ans.

Pour la génération 1905, les résultats sont les suivants :

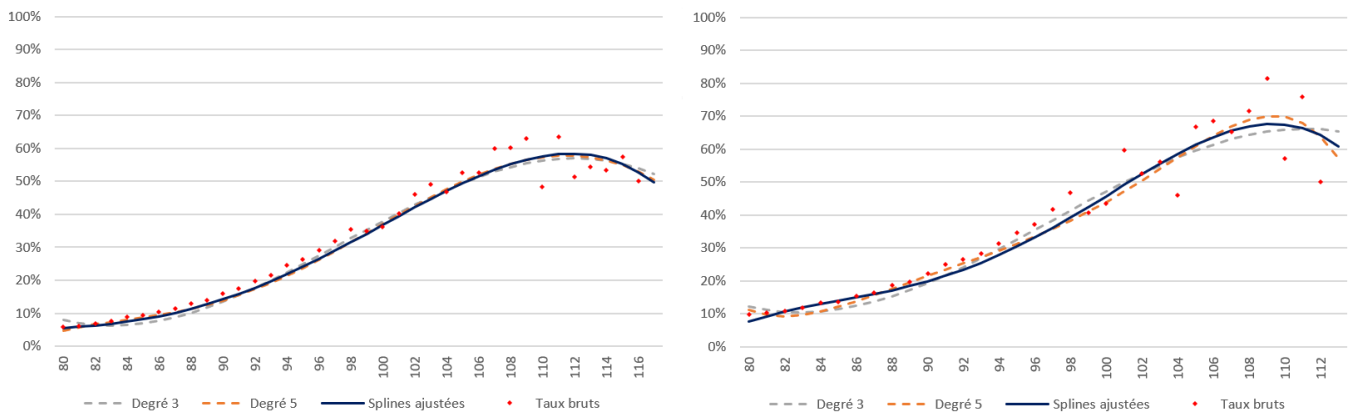


FIGURE 6.7 – *Test de sensibilité sur le degré des splines pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905*

La sensibilité au degré des splines est assez faible. La courbe reste quasiment inchangée pour les femmes en faisant varier ce paramètre. La sensibilité est un peu plus forte pour les hommes, pour qui les taux sont plus volatils, ce qui explique cet écart de sensibilité. Nous pouvons remarquer que dans ce cas, le degré 4 est un bon compromis entre le degré 3 et le degré 5 pour lesquels l'adéquation aux taux bruts est légèrement moins bonne. L'ajustement au niveau des âges entre 80 et 87 ans se dégrade faiblement avec le changement de degré des splines.

Quant au positionnement des noeuds les résultats sont les suivants :

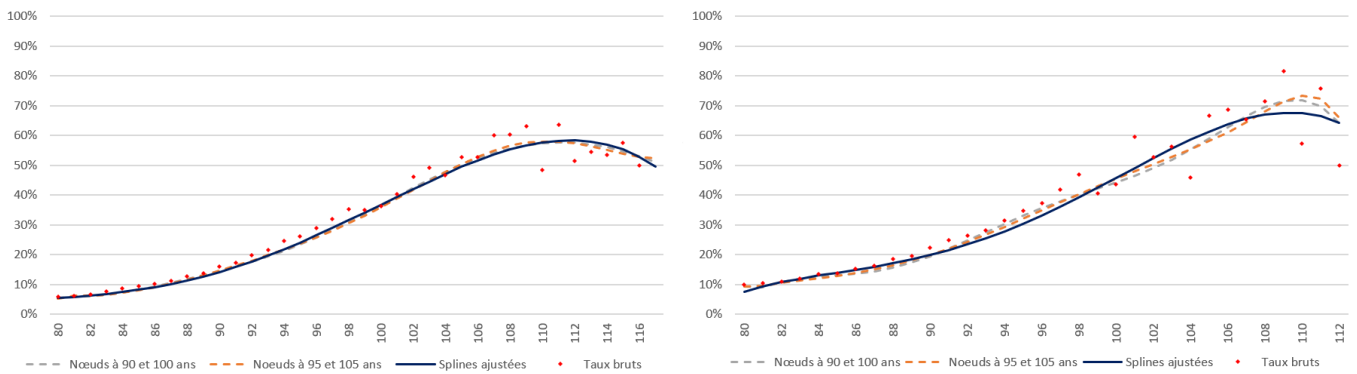


FIGURE 6.8 – *Test de sensibilité sur les noeuds des splines pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905*

Une nouvelle fois, l'impact reste faible sur la population féminine. Il est légèrement plus important sur la population masculine, notamment après 100 ans, où la forte volatilité et le noeud à 105 ans ont pour effet de faire augmenter les taux ajustés de presque 10% à 110 ans. L'inflexion, à l'âge de 111 ans est très accentué par le changement des noeuds. Lors de la mise en place de cet ajustement il est donc essentiel

de porter une attention particulière à ce paramètre.

Ce modèle est par conséquent assez robuste pour les populations avec la volatilité des taux de mortalité contrôlée, mais peut tout de même s'avérer sensible, notamment aux niveaux des noeuds, dans le cas contraire.

- **Lindbergson**

Dans la deuxième approche, celle adoptée par Lindbergson, la sensibilité porte seulement sur le paramètre k . La sensibilité aux paramètres a , b et c ne sera pas testée, étant donné que l'on cherche les paramètres ajustant au mieux les grands âges. Un test sur l'influence de l'âge charnière a également été réalisé. Les variations étant très faibles ces derniers résultats ne seront pas illustrés mais seulement évoqués.

Les résultats du test de sensibilité sur le paramètre k sont les suivants :

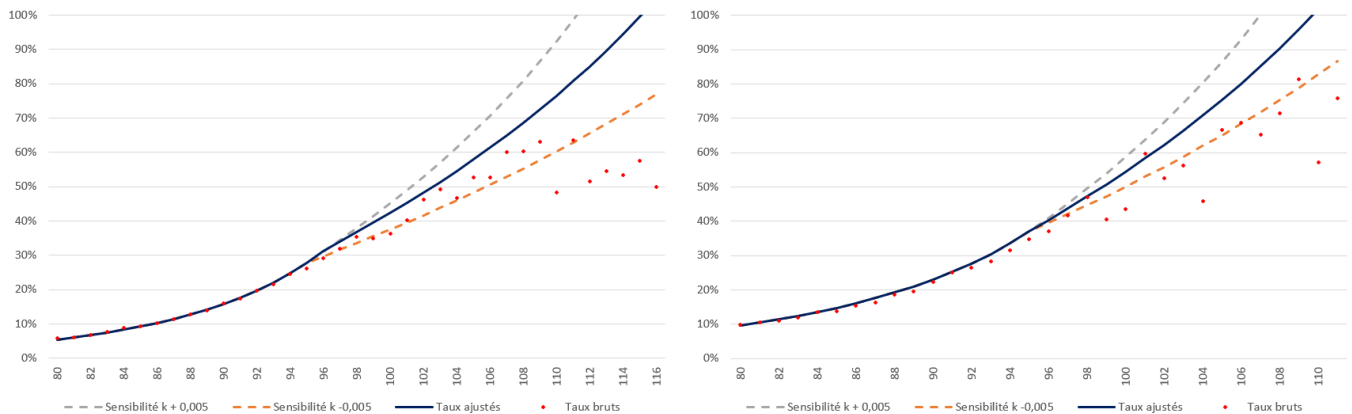


FIGURE 6.9 – Test de sensibilité sur le paramètre k de la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905

D'une part, on rappelle que l'ajustement des taux ajustés aux taux bruts n'est pas optimal, pour cette génération, à cause du lissage horizontal, qui a pour effet de surestimer la mortalité de cette cohorte. D'autre part, nous remarquons que l'ajustement de Lindbergson est très sensible au paramètre k , étant donné qu'une variation de 0,5% de sa valeur aboutit à une différence de plus de 10% au niveau des âges les plus avancés. Cela est cohérent étant donné qu'il s'agit de la pente de la courbe après l'âge charnière.

Au contraire, une variation de 1 ou 2 ans du paramètre ω n'a pas de grande influence sur l'ajustement. Cela ne fait que décaler le point à partir duquel la courbe devient une droite.

Comme nous pouvions nous y attendre, cette méthode est, de fait, sensible au paramètre k . Il est donc important de l'estimer avec une précision suffisante.

- **Denuit & Goderniaux**

Pour ce modèle le test de sensibilité porte uniquement sur le paramètre c . Nous ne testerons pas ici la sensibilité au paramètre x_{max} . Cela a été fait lors de la mise en place du modèle au cours de laquelle deux cas ont été implémentés. Ces deux cas ont également permis à tester la sensibilité à ce paramètre, qui est finalement assez réduite pour une variation de 5 ans.

Les résultats des tests de sensibilité au paramètre c sont les suivants :

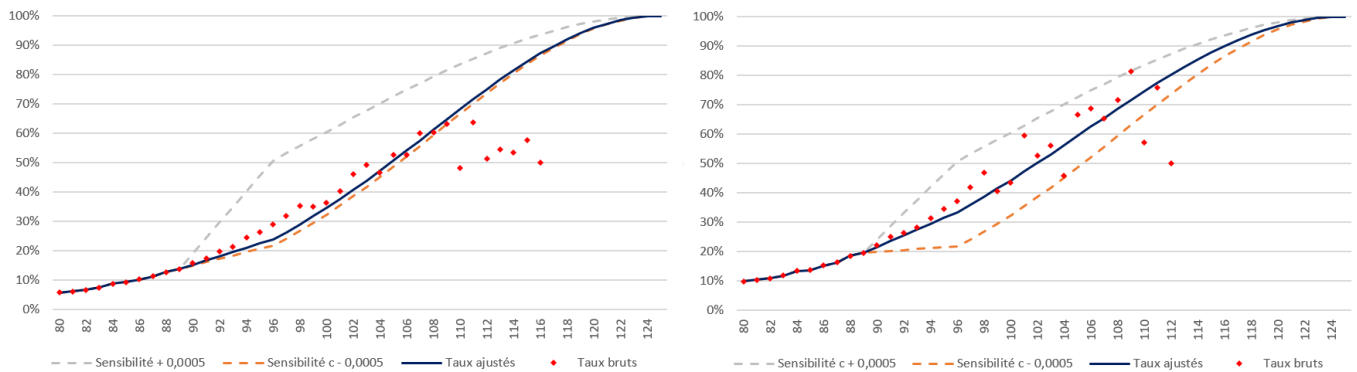


FIGURE 6.10 – *Test de sensibilité sur le paramètre c de la méthode de Denuit & Goderniaux pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905*

La sensibilité au paramètre c est très importante. Cela est aussi une conséquence du fait que le modèle est entièrement déterminé par cet unique paramètre. En le modifiant de l'ordre de 10^{-4} , les taux de mortalité peuvent varier de plus de 20%. Pour les femmes, le paramètre optimal obtenu via la méthode des moindres carrés étant assez proche du point d'inflexion de la courbe de la quantité à minimiser, une variation de cet ordre n'a pas vraiment d'impact. En raison de cette variation, l'ajustement par moyennes pondérées au niveau de l'âge charnière marque une forte discontinuité.

Il est alors primordial, pour cette méthode, d'être le plus précis possible dans l'ajustement du paramètre c , au risque de faire une grande erreur d'estimation des taux de mortalité.

- **Kannistö**

Pour le modèle de Kannistö, le test porte sur la sensibilité à chacun des deux paramètres a et b du modèle.

Pour les tests de sensibilité au paramètre a , avec une variation de l'ordre de 10^{-4} , les résultats sont les suivants :

6.4. TESTS DE SENSIBILITÉ

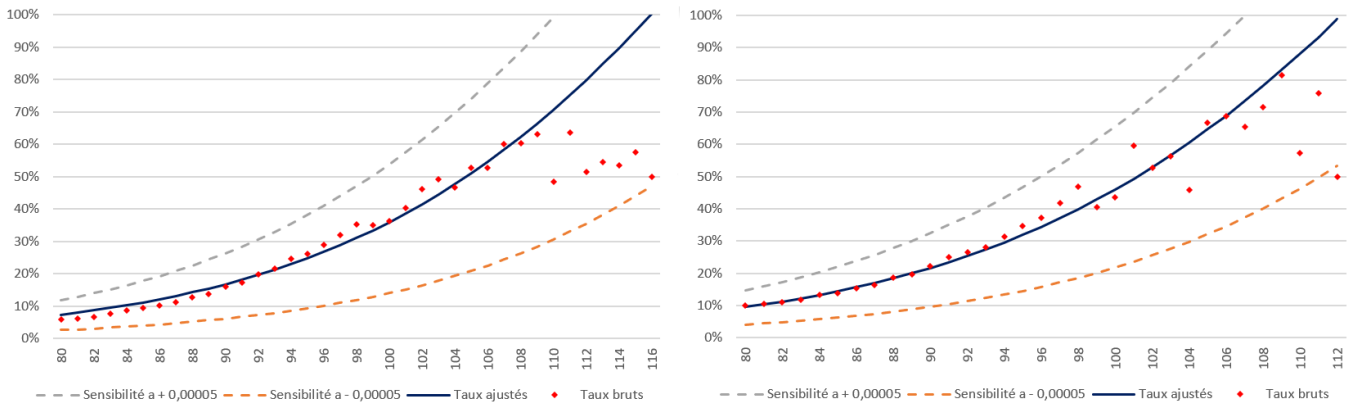


FIGURE 6.11 – *Sensibilité au paramètre a pour le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905*

Nous constatons une très forte sensibilité au paramètre a , avec des écarts de plus de 30% pour certains âges, et toujours supérieurs à 8%. Cette sensibilité est très réduite lorsque l'ordre de la variation est de 10^{-5} et non plus 10^{-4} . Afin d'ajuster ce modèle, il faut alors au moins une précision de ce niveau là pour ce paramètre.

La sensibilité au paramètre b est moindre que pour le paramètre a mais reste forte :

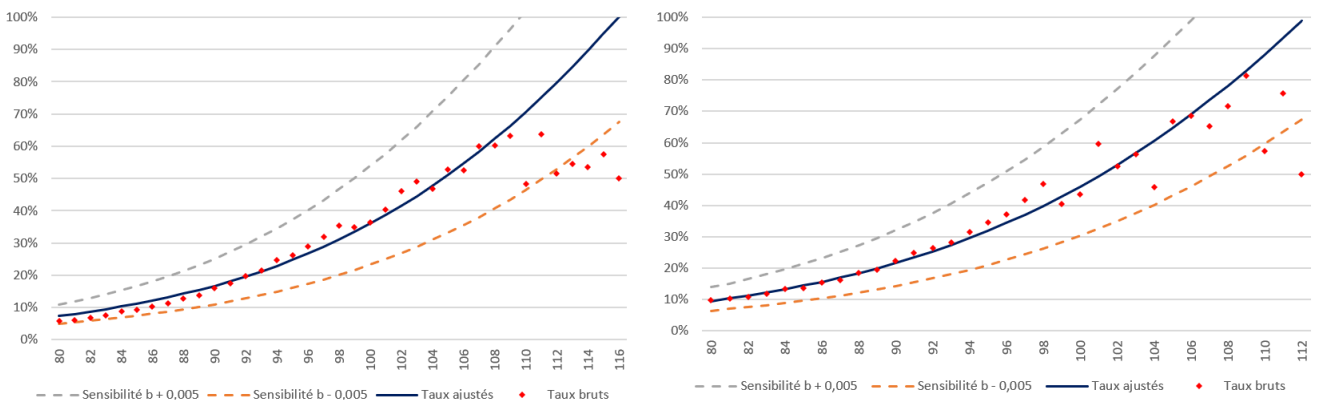


FIGURE 6.12 – *Sensibilité au paramètre b pour le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905*

Avec une précision de l'ordre de 10^{-3} pour ce paramètre, les écarts sont marqués mais de réduisent pour une variation de 10^{-4} .

Le modèle de Kannistö fournit des résultats très sensibles. Là encore les écarts sont assez importants avec des variations parfois supérieures à 20%. Il est donc essentiel d'estimer les paramètres avec une précision suffisante au risque de faire une grande erreur d'estimation.

- **Whittaker-Henderson**

Les tests de sensibilité portent, pour ce modèle, sur le paramètre de régularité ainsi que sur celui de fidélité.

Les résultats obtenus sur les tests portant sur le paramètre de régularité sont les suivants :

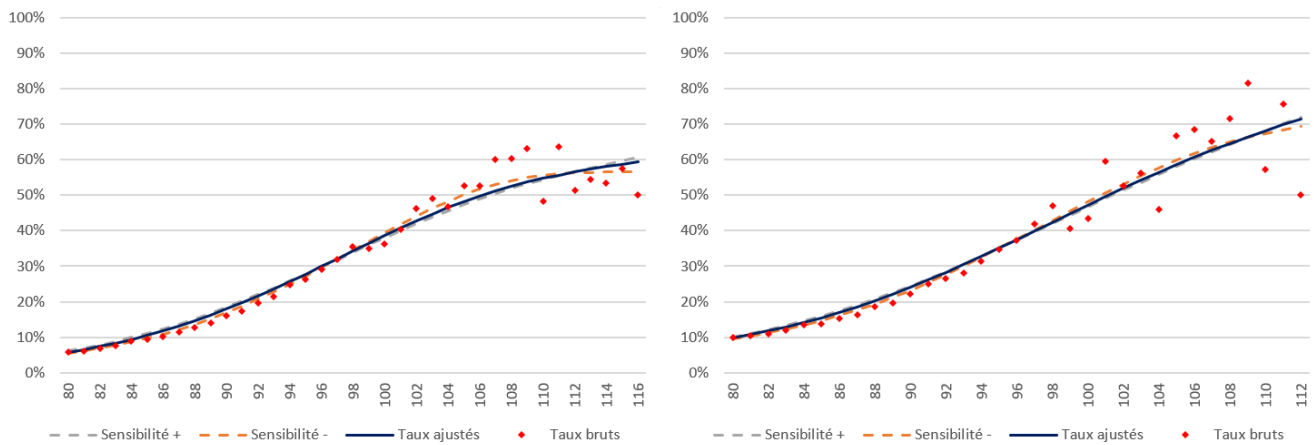


FIGURE 6.13 – *Sensibilité au paramètre de régularité pour le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite)*

Les tests effectués montrent que le paramètre de régularité n'a pas une très grande influence. En effet, mis à part les âges après 100 ans, pour lesquels les taux sont modifiés de manière très légère, la sensibilité est très réduite. Le paramètre de fidélité « écrase » celui de régularité ici.

Le modèle est très sensible au paramètre de fidélité :

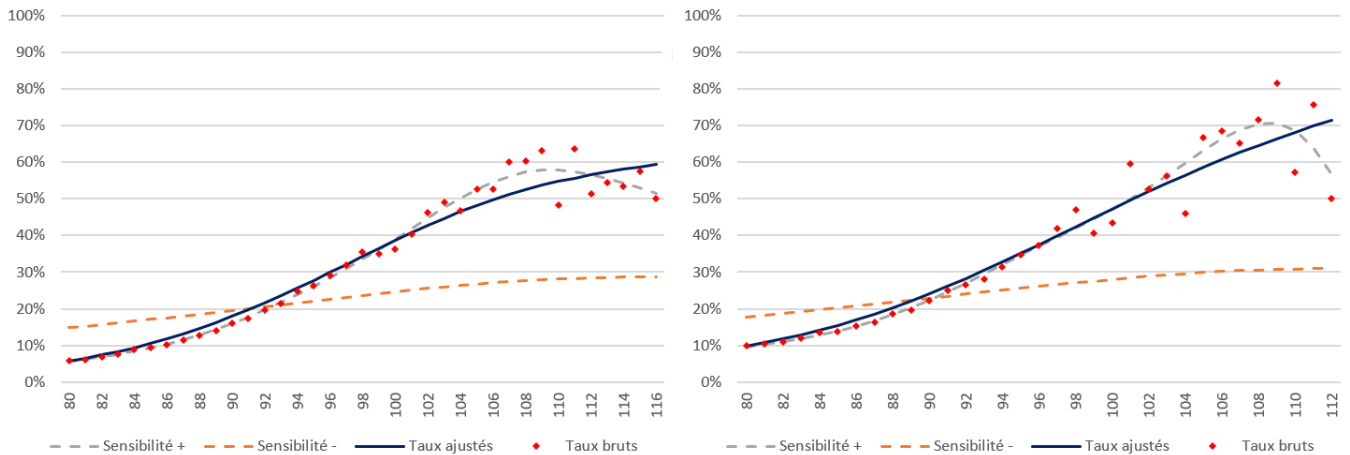


FIGURE 6.14 – *Sensibilité au paramètre de fidélité pour le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite)*

Nous constatons qu'en abaissant le paramètre de fidélité, on obtient des résultats incohérents. Cependant, en l'augmentant plus qu'illustrer sur la Figure ci-dessus, on se rend compte que l'on peut rapidement faire de l'overfitting¹. Cela peut notamment être le cas si le paramètre de fidélité est encore augmenté ou, si l'on diminue le paramètre de régularité et que l'on augmente le paramètre de fidélité en même temps.

En conclusion le paramètre le plus important de ce modèle est le paramètre de fidélité. Le paramètre de régularité n'est pas à négliger pour autant. En effet, l'abaisser de manière trop forte aurait abouti à de l'overfitting. Nous aurions donc obtenu un modèle donnant, à très peu de chose près, les taux bruts, il n'y aurait donc pas eu d'ajustement et la volatilité serait restée élevée aux grands âges.

- **Brass**

Le test de sensibilité porte sur les paramètres dont la table de référence utilisée. Les tables de référence prises afin de pouvoir réaliser le test de sensibilité sont les tables TH/TF 00-02, qui ne sont pas des tables prospectives. Cela ne pose cependant pas de problème, étant donné que les paramètres du modèle doivent être capables de capter la dérive de longévité en faisant un ajustement par génération.

1. Il s'agit d'une problématique de machine learning désignant un sur-ajustement aux données aboutissant à un modèle peu robuste.

Les résultats sur les tests de sensibilité aux deux paramètres sont les suivants :

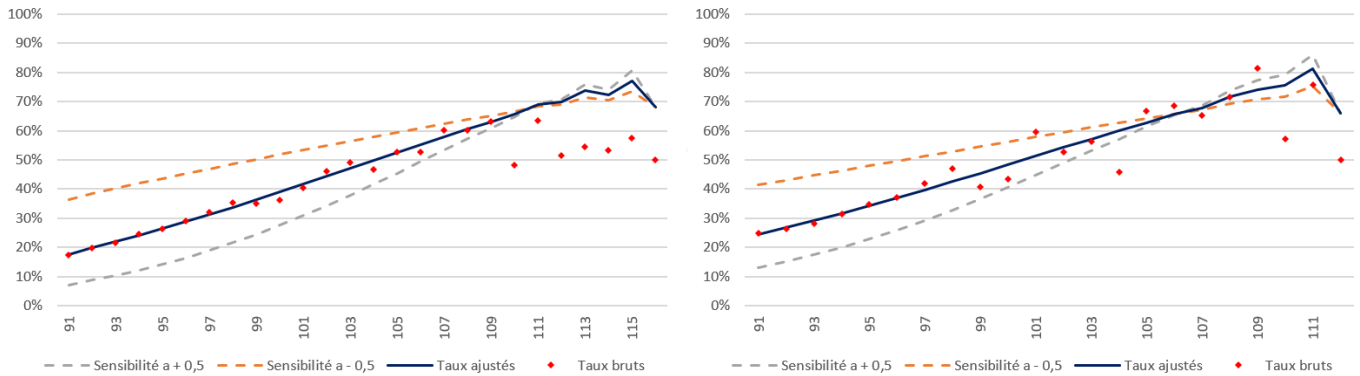


FIGURE 6.15 – Sensibilité au paramètre a pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)

Le paramètre a , représentant la pente du taux de mortalité, est un paramètre auquel le modèle est sensible. Une légère variation de celui-ci conduit à une pente mal ajustée, et donc des taux qui ne sont pas en phase avec les taux bruts.

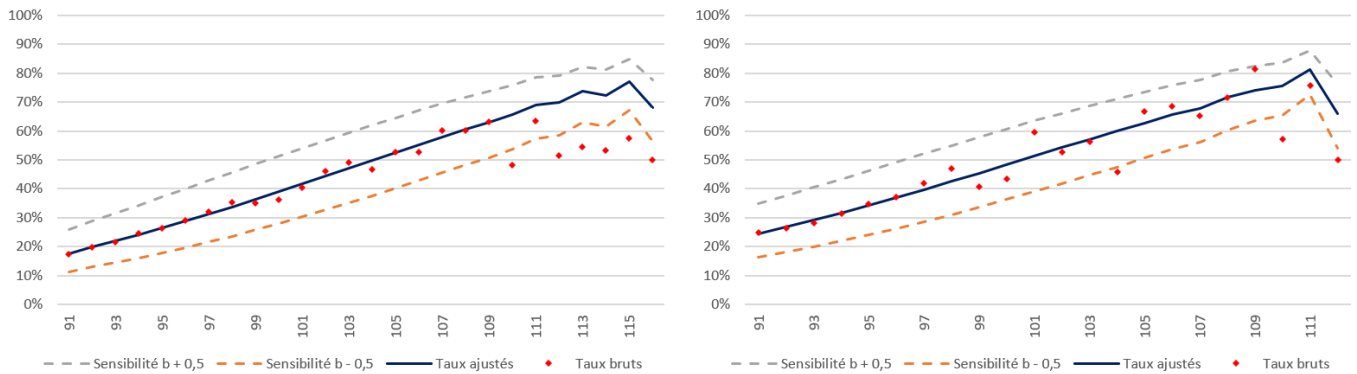


FIGURE 6.16 – Sensibilité au paramètre b pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)

Le paramètre b , jouant un rôle d'« offset », est également un paramètre auquel le modèle est sensible. Une légère variation de celui-ci peut avoir un fort impact sur les résultats.



FIGURE 6.17 – *Sensibilité à la table de référence utilisée pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)*

Nous constatons que la table de référence utilisée joue également un rôle important dans la détermination des taux de mortalité par la méthode de Brass. En effet, en changeant cette référence, nous trouvons des résultats sensiblement différents, avec des écarts de plus de 10% pour certains âges.

En conclusion, nous retiendrons que, si l'on préfère se fier au test de cohérence des taux ajustés, les modèles de Whittaker-Henderson, des P-splines et de Brass fournissent les meilleurs résultats pour capter la mortalité. Sur la base des méthodes SMR ou MAPE ce sont les mêmes modèles qui présentent les résultats qui ajustent le mieux les taux bruts. La contrainte introduite par le modèle de Denuit & Goderniaux ne permet finalement pas une très bonne adéquation aux taux bruts. Le modèle de Kannistö tend à surestimer les taux de mortalité. Le modèle de Lindbergson fournit, quant à lui, des résultats qui peuvent être soit très proches des taux bruts, soit assez éloignés selon les générations et âges.

Les tests ont également porté sur la sensibilité des modèles aux paramètres et hypothèses. Ils ont ainsi révélé les faiblesses de certains modèles très sensibles aux paramètres. C'est notamment le cas des modèles de Whittaker-Henderson sensible au paramètre de fidélité, ou de celui de Kannistö, très sensible à ses deux paramètres. Les autres modèles sont également sensibles à leurs paramètres respectifs, notamment ceux de Lindbergson et Denuit & Goderniaux, dans une moindre mesure. Le modèle des P-splines est celui qui est le moins sensible.

Chapitre 7

Application à un portefeuille de rentes

Introduisons désormais la base de données assurantielle sur laquelle nous allons étudier l'impact des ajustements proposés sur le provisionnement du portefeuille de rentes. Ce dernier pourra notamment être assimilé à un portefeuille retraite. Il est donc utile de commencer par un rappel du principe de fonctionnement et des chiffres clés de la retraite en France.

7.1 Généralités

La retraite est une prestation de prévoyance sociale essentielle, elle représente une grande partie de ce type de prestations. Voici, pour exemple, les chiffres de 2017 :

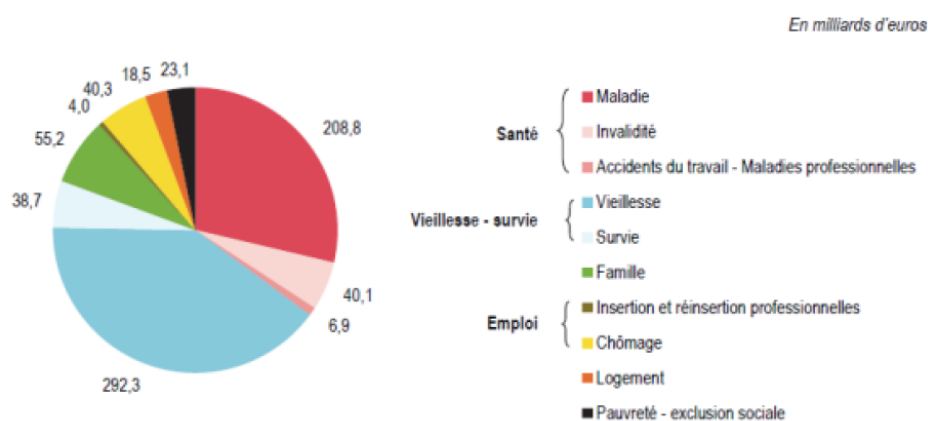


FIGURE 7.1 – *Composition des prestations de protection sociale en 2017 par risque*

En 2017, le risque vieillesse et donc, en partie la retraite, représente plus du tiers des prestations globales de protection sociale en France. La part des contrats d'épargne

retraite reste cependant relativement faible : selon la DREES¹ cette part s'élève à 2,1% des prestations vieillesse, soit 6,1 milliards d'euros.

Un régime de retraite peut se définir comme étant un système de transfert d'une fraction non immédiatement consommée des revenus générés par la population des actifs vers la population des retraités. En France, il existe 42 régimes de retraite différents. Ils se divisent en deux types de régimes :

- Les régimes à prestations définies : les prestations qui seront servies par le régime de retraite sont définies contractuellement, dans le règlement intérieur du régime, les cotisations du régime sont ajustées régulièrement. La prestation peut être définie de manière absolue (régime additionnel) ou relative (régime différentiel). Les régimes à prestations définies sont générateurs de passif social.
- Les régimes à cotisations définies : le niveau des cotisations est défini contractuellement, le niveau des prestations fluctue en fonction du contexte économique, démographique et financier.

Plusieurs types de financement sont possibles :

- Le financement par capitalisation : les prestations servies sont prélevées sur les réserves financières constituées au fil des années par les cotisations de l'assuré.
- Le financement par répartition : la totalité des prestations de l'année est assurée par les cotisations de l'année. L'avantage principal de ce type de financement est qu'il permet l'attribution immédiate de droits.

Les caractéristiques de chaque type de régime de retraite sont résumées dans la Figure suivante, qui fournit également des exemples par type de régime et de financement.

	Répartition	Capitalisation
Cotisations définies	Obligation de moyen. Réserves nulles ou faibles. Le rapport démographique et la conjoncture économique influent principalement sur l'équilibre du régime. <i>Exemple : ARRCO, AGIRC.</i>	Obligation de moyen. Engagements gagés par des actifs financiers. Le niveau des taux réels de placement (taux bruts diminués de l'inflation) influe principalement sur les performances du régime. Droits individualisés. <i>Exemple : Contrat "Article 83", PERE, PERP.</i>
Prestations définies (additionnels et différentiels)	Obligation de résultat. Réserves nulles ou faibles. Le rapport démographique et la conjoncture économique influent principalement sur l'équilibre du régime. <i>Exemple : Régime des fonctionnaires.</i>	Obligation de résultat. Engagements gagés par des actifs financiers. Le niveau des taux réels de placement (taux bruts diminués de l'inflation) influe principalement sur les performances du régime. Droits non individualisés (et non individualisables ?) <i>Exemple : Contrat "Article 39"/"Article 115".</i>

FIGURE 7.2 – Régimes de retraite par droits et financement

1. Direction de la Recherche, des Études, de l'Évaluation et des Statistiques.

La situation démographique est un des principaux facteurs pouvant influencer le bon fonctionnement de régimes par répartition. Pour les régimes par capitalisation, l'inflation peut avoir un fort impact.

Il existe 4 niveaux différents de couverture vieillesse :

- La retraite de base (obligatoire employeurs et salariés).
- La retraite complémentaire (obligatoire employeurs et salariés).
- La retraite supplémentaire (facultatif employeur).
- L'assurance vie (facultatif individus).

La masse de dépenses pour les systèmes de retraite s'élève en 2019, selon les chiffres indiqués sur le site du gouvernement, à 331 milliards d'euros. On se rend compte que ce chiffre est en hausse par rapport à 2017 (292,3 milliards en 2017, cf. Figure 7.1). Cela est dû, principalement, au vieillissement de la population, ce qui signifie plus de prestations vieillesse à verser. En effet, le site indique aussi qu'il y a environ 700 000 nouveaux retraités chaque année. Aujourd'hui, on estime le nombre de retraités en France à près de 15,9 millions, soit plus de 20% de la population nationale.

7.2 Présentation des données

Les données utilisées sont celles d'un portefeuille constitué de personnes de plus de 80 ans. Initialement, ce portefeuille ne couvre pas une garantie qui offre des prestations sous forme de rentes viagères, mais les données fournies sont intéressantes pour notre étude, étant donné l'âge moyen du portefeuille. Pour chaque individu, on observe les variables suivantes :

- Numéro identifiant
- Date de naissance
- Sexe

Pour les futurs calculs, on considère les hypothèses suivantes :

- Taux d'actualisation : 0,00%.
- Taux de revalorisation annuelle de la rente : 1,17% – évolution moyenne du SMIC² sur les 5 dernières années.
- Montant de la rente annuelle par individu : 1€.

Il s'agit d'une base de données contenant 1782 lignes, ce qui est assez réduit. Cependant, étant donné le faible échantillon à ces âges là, il est difficile de constituer des bases de données importantes.

2. Salaire Minimum Interprofessionnel de Croissance.

7.3 Statistiques descriptives

Le but de cette section est de donner une première approche sur les données du portefeuille. Cela nous permettra de constater, ou non, un rapprochement entre cette base de données et les données nationales sur lesquelles nous avons travaillées pour construire les tables. Cependant, cette approche est assez limitée étant donné que la profondeur d'historique dont nous disposons sur ce portefeuille est seulement d'un an. Cela fait qu'il ne nous est pas possible d'évaluer des taux de mortalité pour cette population par exemple, et d'ainsi pouvoir évaluer un écart avec la population nationale. Observons la répartition hommes/femmes du portefeuille :

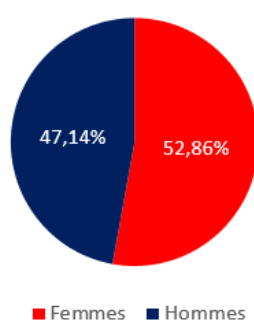


FIGURE 7.3 – Répartition hommes/femmes du portefeuille de rentes

Nous constatons que les hommes sont surreprésentés dans cette base par rapport à la population nationale. En effet, la proportion d'hommes passe de 32,82% au niveau national à 47,14% dans ce portefeuille. Cela a également pour effet une sous-représentation des femmes en proportion.

Il est également pertinent de s'intéresser à la répartition par âge des assurés :

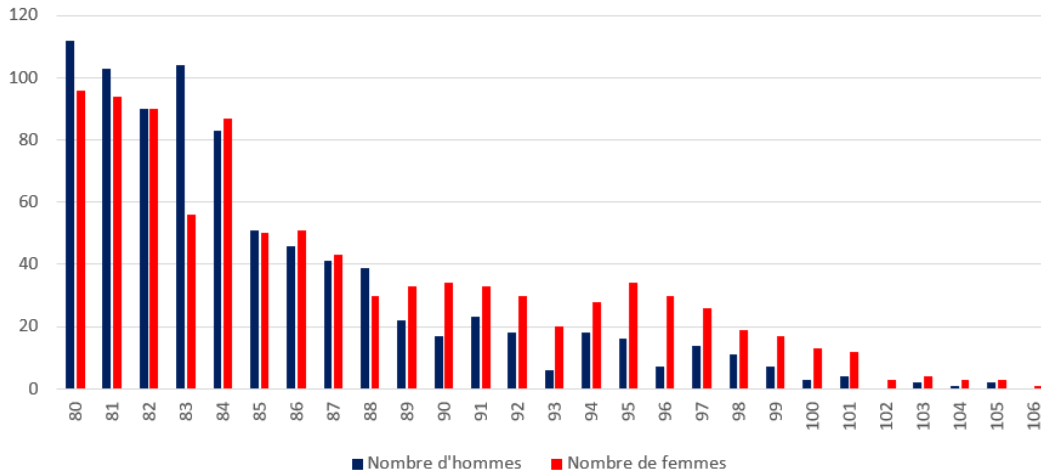


FIGURE 7.4 – Répartition par âge du portefeuille de rentes

Les âges s'étendent jusqu'à 106 ans (âge révolu au 31 décembre 2020). Les hommes sont, en moyenne, plus jeunes que les femmes et, comparativement à celles-ci, relativement peu d'entre eux dépassent les 90 ans. L'homme le plus âgé du portefeuille atteint 105 ans.

Ces deux Figures montrent ainsi que la population assurée est différente de la population globale, notamment au niveau de la répartition hommes/femmes. Cependant, la répartition par âge reste proche, étant donné la chute du nombre de personnes autour de 85 ans, puis une baisse moins forte sur les âges les plus élevés. L'exposition de la population masculine est moins forte que celle de la population féminine à partir de 89 ans.

7.4 Impacts financiers des ajustements

Calculons le montant des provisions de notre portefeuille de rentes viagères. Pour un individu donné, nous utilisons la formule suivante :

$$PM = \sum_{k=x_0}^{x_{max}} \frac{kP_{x_0}}{(1+i)^k} R(1+\tau)^k$$

Avec R le montant de la rente annuelle en x_0 , τ le taux de revalorisation annuel de la rente, i le taux d'actualisation, x_0 l'âge actuel de l'individu considéré.

Il est intéressant d'ajouter dans la comparaison des provisions mathématiques de ce portefeuille, celles que l'on obtient avec les tables réglementaires TGH/TGF 05.

	PM	Écart aux taux sans ajustement
Sans ajustement	15 682 €	
Tables réglementaires	17 215 €	9,78%
P-splines	16 079 €	2,53%
Lindbergson	15 722 €	0,26%
Denuit & Goderniaux	16 855 €	7,48%
Kannistö	14 782 €	-5,74%
Whittaker-Henderson	15 193 €	-3,12%
Brass	16 009 €	2,09%

FIGURE 7.5 – *Impacts financiers des différents ajustements*

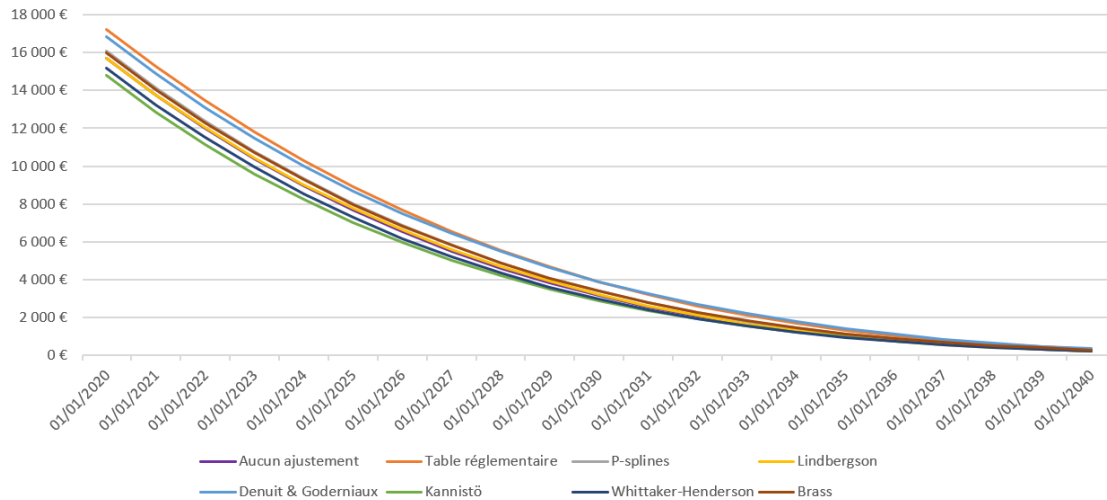
Malgré la taille réduite de notre base de données, nous visualisons des impacts financiers importants, avec des écarts s'élevant parfois à plus de 10%. Cela confirme l'importance du choix de la méthode d'ajustement.

Les tables réglementaires, sont, par nature, assez prudentes, – et semblent même légèrement trop prudentes sur ce portefeuille au vu des autres résultats. La méthode de Denuit & Goderniaux se révèle également prudente contrairement au modèle de Kannistö qui fait preuve du moins de prudence.

Le modèle permettant de s'approcher au mieux, en terme de provisionnement, des taux bruts est le modèle de Lindbergson, suivi du modèle de Brass et de celui des P-splines. Le modèle de Whittaker-Henderson est finalement assez peu prudent sur notre portefeuille.

Il est également intéressant de projeter dans le temps l'évolution du montant des provisions pour ce portefeuille de rentes. Nous n'avons pas assez d'historique sur les données pour estimer une loi d'entrée dans ce portefeuille, c'est pourquoi nous fondons notre raisonnement sur le fait que le portefeuille est fermé. Cela signifie que nous ne supposons aucune entrée dans le portefeuille pendant toute la durée de projection.

Les résultats sont les suivants :

FIGURE 7.6 – *Projection du montant des provisions selon le modèle d'ajustement*

Du fait du grand âge des assurés, pour chaque modèle d'ajustement, le montant des provisions baisse fortement dans les premières années de projection. La baisse est beaucoup plus réduite sur la fin de vie du portefeuille étant donné le faible nombre de sorties à ces dates. L'écart obtenu pour l'exercice 2020 selon la méthode d'ajustement utilisée, se poursuit et reste, à peu de choses près, similaire en proportion à celui de cette première année.

7.5 Tests de sensibilité

Des tests de sensibilités sur les résultats obtenus ont également été réalisés. Les résultats obtenus sont les suivants :

	PM	
	Sensibilité : + 0,5%	Sensibilité : - 0,5%
Sans ajustement	15 226 €	16 164 €
Table réglementaire	16 680 €	17 782 €
P-splines	15 599 €	16 588 €
Lindbergson	15 263 €	16 208 €
Denuit & Goderniaux	16 327 €	17 416 €
Kannistö	14 360 €	15 229 €
Whittaker-Henderson	14 760 €	15 652 €
Brass	15 530 €	16 518 €

FIGURE 7.7 – *Sensibilité des résultats obtenus au taux d'actualisation*

Le fait d'augmenter le taux d'actualisation permet de limiter l'impact de la composante de longévité, ce qui a pour conséquence de baisser le montant des provisions,

mais aussi de réduire les écarts entre les différents modèles. Avec une augmentation plus forte du taux d'actualisation les écarts auraient été moins importants. Au contraire, une diminution de ce taux donne encore plus d'importance au risque de longévité. Le test de sensibilité avec une diminution du taux d'actualisation de 0,5 points amène à utiliser un taux d'actualisation négatif tandis qu'en pratique, en assurance vie, nous retiendrons un taux d'actualisation positif.

En conclusion, il est important de retenir que la sélection de la méthode de fermeture de tables est essentielle lors de leur construction. En effet, selon la méthode choisie, le montant des provisions peut varier de plus de 10%.

Malgré des résultats mitigés sur les erreurs de modèles, c'est la méthode de Lindbergson qui donne les résultats les plus proches de ceux obtenus avec les taux bruts. La méthode de Whittaker-Henderson, bien que fournissant des taux assez proches des taux bruts, reste finalement assez peu prudente pour nos données. Cela est dû à l'écart entre la population assurée et la population nationale sur laquelle est construite la table de mortalité. Le modèle développé qui fournit les taux de mortalité les plus prudents est celui de Denuit & Goderniaux, dont le montant des provisions obtenu est 7,5% plus élevé que celles obtenues avec les taux bruts.

Conclusion

Afin de conclure ce mémoire sur l'ajustement des taux de mortalité aux grands âges, il faut retenir qu'il ne s'agit pas d'une nouvelle problématique d'assurance vie, mais qu'elle reste encore ouverte à ce jour. Elle prend de plus en plus d'ampleur, en particulier dans le contexte actuel où les taux d'intérêts sont bas. Le risque de longévité n'est plus écrasé par le taux d'actualisation, ce qui a pour effet d'augmenter les provisions mathématiques.

Le contexte réglementaire a également un rôle à jouer dans cette problématique, étant donné le poids des tables de mortalité dans la directive Solvabilité 2. De plus, le passage de la table réglementaire TPRV93 aux tables TGH/TGF 05 et le défaut de provisionnement suffisant à ce moment là, donnent encore plus d'importance au sujet de la dérive de longévité.

Afin de réaliser cette étude, nous avons d'abord travaillé sur des données publiques issues de la Human Mortality Database, que nous avons complétées aux âges supérieurs à 110 ans sans trop de pertes de qualité des données. À partir de cela, il a été possible de construire une table de mortalité via l'estimateur binomial, ainsi que l'estimateur de Kaplan-Meier, mais la faible taille de l'échantillon de données disponibles aux grands âges conduit à une volatilité et à des intervalles de confiance très larges. Les résultats obtenus n'avaient alors plus vraiment de crédibilité et beaucoup d'incertitude. C'est pourquoi nous avons développé des modèles d'ajustement pour ces âges.

Certaines de ces méthodes font des hypothèses sur la forme des taux de mortalité : il s'agit des modèles paramétriques. Ces modèles fournissent des résultats relativement bons, avec des erreurs de modèle assez réduites pour certains comme pour le modèle des P-splines. D'autres, comme le modèle de Denuit & Goderniaux, fournissent des résultats plus éloignés des taux bruts, mais qui semblent parfois plus cohérents. En effet, la méthode de construction des taux prospectifs utilisée dans ce mémoire, conduit à des taux de mortalité finalement assez faibles. En effet, bien que cela réponde à des problématiques de prudence, des taux de mortalité inférieurs à 20 ou 25% après 110 ans nous semblent assez optimistes quant à l'évolution de la longévité.

D'autres modèles ne font aucune hypothèse sur la forme des taux de mortalité : il s'agit de modèles non-paramétriques. Ainsi la méthode de Whittaker-Henderson consiste à se rapprocher des taux bruts de mortalité tout en gardant une certaine régularité. Les modèles relationnels, qui se fondent sur une table de référence et évaluent un écart avec cette dernière, font également partie de cette famille de modèles. Il est cependant important de bien choisir la table de référence utilisée car les taux de mortalité obtenus peuvent y être assez sensibles. Les résultats obtenus avec ces méthodes semblent plus proches des taux bruts que ceux obtenus avec les modèles paramétriques, mais plus sensibles aux potentielles anomalies contenues dans les taux bruts. Dans notre cas, en raison d'une bonne qualité des données en entrée, ces approches fournissent de bons résultats.

Enfin, nous avons également analysé les impacts financiers des différents ajustements sur un portefeuille de rentes. Cette analyse a mis en évidence l'importance de la méthode sélectionnée sur le calcul du montant des provisions, étant donné les écarts importants entre les résultats obtenus avec certains modèles. En effet, certains donnent un montant de provisions qui s'écarte de plus de 10% d'autres modèles. Il y a par exemple 14% d'écart entre les PM obtenues avec le modèle de Denuit & Goderniaux et celles obtenues avec celui de Kannistö. Nous avons également remarqué que les tables obtenues sont toutes moins prudentes que les tables réglementaires. Le modèle de Lindbergson donne les résultats les plus proches de ceux obtenus avec les taux bruts sur ce portefeuille, bien qu'il y ait des écarts sur certaines générations. En effet, les écarts se concentrent notamment sur les générations les plus jeunes, qui ne rentrent pas dans le cadre de l'étude des portefeuilles retraites. Le modèle de Kannistö ainsi que celui de Whittaker-Henderson sous-estiment le montant des provisions du fait de la légère surmortalité trouvée, en particulier pour le premier, tandis que le modèle de Denuit & Goderniaux, plutôt prudent, conduit à des résultats contraires.

À l'inverse des tables réglementaires et du modèle de Denuit & Goderniaux, ceux de Whittaker-Henderson et de Kannistö fournissent donc des montants trop peu prudents, tandis que le modèle de Lindbergson, qui fournit le montant le plus proche de celui obtenu avec les taux bruts, ne permet pas beaucoup d'anticiper les éventuels imprévus. Cela nous poussent donc à opter pour un compromis entre l'adéquation aux taux bruts et la mesure de prudence, et donc, à utiliser le modèle de Brass, ou celui des P-splines, qui sont les deux modèles les plus adaptés dans ce cas.

Table des figures

1.1	<i>Historique des tables réglementaires</i>	13
1.2	<i>Évolution de la mortalité en France depuis les années 1970 selon le modèle de Lee Carter</i>	14
2.1	<i>Les différentes composantes du risque biométriques de longévité et leurs effets sur le logarithme du taux de mortalité en fonction du temps</i>	17
2.2	<i>Facteurs du risque de longévité pour les années à venir</i>	20
2.3	<i>Schéma simplifié d'un swap de longévité</i>	22
3.1	<i>Répartition hommes/femmes de l'exposition (à gauche) et des décès (à droite) pour les personnes de plus de 80 ans depuis les années 1990</i>	26
3.2	<i>Répartition de l'exposition par âge depuis les années 1990</i>	26
3.3	<i>Répartition des décès par âge depuis les années 1990</i>	27
3.4	<i>Taux bruts de mortalité par âge et par sexe toutes cohortes confondues</i>	27
3.5	<i>Taux bruts de mortalité par âge et par sexe, en échelle logarithmique, toutes cohortes confondues</i>	28
3.6	<i>Évolution de l'exposition à âge fixe en fonction de la cohorte pour les femmes (à gauche), et pour les hommes (à droite)</i>	29
3.7	<i>Représentation en 3 dimensions des taux bruts de mortalité des femmes (à gauche) et des hommes (à droite)</i>	29
3.8	<i>Représentation en 3 dimensions des taux bruts de mortalité des femmes (à gauche) et des hommes (à droite) en échelle logarithmique</i>	30
3.9	<i>Tests de sensibilité n°1 aux décès des supercentenaires</i>	32
3.10	<i>Test de sensibilité n°1 sur le nombre de décès pour les supercentenaires femmes (à droite) et hommes (à gauche)</i>	32
3.11	<i>Tests de sensibilité n°2 aux décès des supercentenaires</i>	33
3.12	<i>Test de sensibilité n°2 sur le nombre de décès pour les supercentenaires femmes</i>	33
4.1	<i>Diagramme de Lexis</i>	35
4.2	<i>Intervalle de confiance à 95 % pour l'estimateur binomiale à partir de 80 ans pour les femmes (à gauche) et les hommes (à droite) pour la cohorte 1905</i>	40
4.3	<i>Taux d'accroissement de la largeur de l'intervalle de confiance par sexe pour l'estimateur binomial pour la cohorte 1905</i>	41
4.4	<i>Intervalle de confiance à 95% du taux de mortalité obtenu via l'estimateur de Kaplan Meier pour les femmes (à gauche) et les hommes (à droite) pour la cohorte 1905</i>	43

4.5	<i>Mortalité prospective pour les femmes (à gauche) et pour les hommes (à droite)</i>	45
5.1	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle des P-splines</i>	49
5.2	<i>Ajustement des taux de mortalité via la méthode des P-splines pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	50
5.3	<i>Évolution et lissage du paramètre k pour le modèle de Lindbergson</i>	52
5.4	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Lindbergson</i>	53
5.5	<i>Ajustement des taux de mortalité via la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	53
5.6	<i>Ajustement des taux de mortalité via la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) de la génération 1910</i>	54
5.7	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Denuit & Goderniaux avec x_{max} valant 125 ans</i>	56
5.8	<i>Ajustement des taux de mortalité via la méthode de Denuit & Goderniaux pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	57
5.9	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Denuit & Goderniaux avec x_{max} valant 130 ans</i>	57
5.10	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Kannistö</i>	59
5.11	<i>Ajustement des taux de mortalité via le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	60
5.12	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Whittaker-Henderson</i>	65
5.13	<i>Ajustement des taux de mortalité via le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	65
5.14	<i>Évolution du taux de mortalité chez les femmes (à gauche) et chez les hommes (à droite) en fonction de l'âge et de la génération pour le modèle de Brass</i>	67
5.15	<i>Ajustement des taux de mortalité via le modèle de Brass pour les femmes (à gauche) et les hommes (à droite) de la génération 1905</i>	68
6.1	<i>Cohérence des taux bruts femmes à 100, 110 et 120 ans pour les générations 1905, 1955, et 2005</i>	70
6.2	<i>Cohérence des taux bruts hommes à 100, 110 et 120 ans pour les générations 1905, 1955, et 2005</i>	71
6.3	<i>Standard Mortality Ratio à différents âges et générations pour les modèles testés chez les femmes</i>	72
6.4	<i>Standard Mortality Ratio à différents âges et générations pour les modèles testés chez les hommes</i>	72
6.5	<i>Mean Percentage Absolute Error à différents âges et générations pour les modèles testés chez les femmes</i>	74

6.6	<i>Mean Percentage Absolute Error à différents âges et générations pour les modèles testés chez les hommes</i>	74
6.7	<i>Test de sensibilité sur le degré des splines pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	76
6.8	<i>Test de sensibilité sur les noeuds des splines pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	76
6.9	<i>Test de sensibilité sur le paramètre k de la méthode de Lindbergson pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	77
6.10	<i>Test de sensibilité sur le paramètre c de la méthode de Denuit & Goderniaux pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	78
6.11	<i>Sensibilité au paramètre a pour le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	79
6.12	<i>Sensibilité au paramètre b pour le modèle de Kannistö pour les femmes (à gauche) et les hommes (à droite) pour la génération 1905</i>	79
6.13	<i>Sensibilité au paramètre de régularité pour le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite)</i>	80
6.14	<i>Sensibilité au paramètre de fidélité pour le modèle de Whittaker-Henderson pour les femmes (à gauche) et les hommes (à droite)</i>	81
6.15	<i>Sensibilité au paramètre a pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)</i>	82
6.16	<i>Sensibilité au paramètre b pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)</i>	82
6.17	<i>Sensibilité à la table de référence utilisée pour le modèle de Brass pour les femmes (à gauche) et les hommes (à droite)</i>	83
7.1	<i>Composition des prestations de protection sociale en 2017 par risque</i>	84
7.2	<i>Régimes de retraite par droits et financement</i>	85
7.3	<i>Répartition hommes/femmes du portefeuille de rentes</i>	87
7.4	<i>Répartition par âge du portefeuille de rentes</i>	88
7.5	<i>Impacts financiers des différents ajustements</i>	89
7.6	<i>Projection du montant des provisions selon le modèle d'ajustement</i>	90
7.7	<i>Sensibilité des résultats obtenus au taux d'actualisation</i>	90

Références

- BLAKE, D., CAIRNS, A., COUGHLAN, G., DOWD, K. et MACMINN, R. (2016). « Le nouveau marché du risque de longévité ». *Revue d'Économie Financière*, 122:1–4 et 23–24.
- BOUMEZOUED, A. (2016). « Approches micro-macro des dynamiques de populations hétérogènes structurées par âge. application aux processus auto-excitants et à la démographie ».
- BOURREAU, C. (2014). « Prise en compte de l'hétérogénéité des individus dans le modèle de Lee-Carter ».
- BOUYEKHF, R. (2016). « Construction et validation de tables de mortalité prospectives en cas de petits échantillons ».
- CAPPE, C. (2018). « Établissement d'une table de mortalité spécifique pour un régime de retraite existant et évaluation des impacts financiers ».
- DEBONNEUIL, E., LOISEL, S. et PLANCHET, F. (2015). « Do actuaries believe in longevity deceleration ? ». *Les cahiers de recherches de l'ISFA*, pages 2–21.
- DEVINEAU, L., BOUMEZOUED, A., HABART, M., BALLAND, F. et POPA, T. (2017). « La qualité des données de mortalité dans le cadre d'un modèle interne ». pages 5–23.
- EL KAROUI, N. et LOISEL, S. (2017). « Le risque de longévité est-il assurable ? ». *Revue d'Économie Financière*, pages 1–8.
- EL KAROUI, N., LOISEL, S. et SALHI, Y. (2015). « Minimax optimality in robust detection of a disorder time in poisson rate ». pages 2–25.
- GUILLON VERNE, T. (2015). « Construction de tables de mortalité d'expérience sur de petits échantillons pour l'estimation de la sinistralité décès ».
- KAMEGA, A. et PLANCHET, F. (2011). « Mortalité prospective en cas de petits échantillons : modélisation à partir d'informations externes en utilisant l'approche de bongaarts ». *Bulletin Français d'Actuariat*, 13:14–19.
- KANDJI, M. (2016). « Le swap de longévité : modélisation et mesure d'impact d'une solution de réassurance innovante sous solvabilité 2 ».
- KIMELDORF, G. et JONES, D. (1967). « Bayesian graduation ». *TSA*, 19.
- LELIEUR, V. (2005). « Utilisation des méthodes de Lee-Carter et Log-Poisson pour l'ajustement de tables de mortalité dans le cas de petits échantillons ».

-
- LUZON, A. (2019). « Mise à jour de la loi d'incidence et de maintient d'un contrat dépendance et impact sur le provisionnement ».
 - MARRON, J. et PADGETT, W. (1987). « Asymptotically optimal bandwidth selection for kernel density estimators for randomly right censored samples ». *The Anal of Statistics*, 15(4):1520–1535.
 - PLANCHET, F. (2007a). « L'utilisation des splines bidimensionnels pour l'estimation de lois de maintien en arrêt de travail ». *Bulletin Français d'Actuariat*, pages 3–8.
 - PLANCHET, F. (2007b). « Tables TGH / TGF 05 : Construction ». pages 4–8.
 - PLANCHET, F. (2020-2021a). « Méthodes de lissage et d'ajustement ». pages 6–17.
 - PLANCHET, F. (2020-2021b). « Statistiques des modèles non paramétriques ». pages 17–19.
 - PLANCHET, F. et LELIEUR, V. (2007). « Utilisation des méthodes de Lee-Carter et Log-Poisson pour l'ajustement de tables de mortalité dans le cas de petits échantillons ». *Bulletin Français d'Actuariat*, pages 2–4 et 13–16.
 - QUASHIE, A. et DENUIT, M. (2005). « Modèle d'extrapolation de la mortalité aux grands âges ». pages 2–14.
 - ROGER, A. (2015). « Modélisation du risque de surmortalité et étude d'impact sur l'activité de réassurance vie en prévoyance ».
 - SALHI, Y. (2018). « Tables de mortalité best estimate : une approche de crédibilité pour les portefeuilles de petite taille ».
 - SILUÉ, Z. (2007). « Élaboration d'une table de mortalité spécifique au régime d'assurance vieillesse de la cnps ».
 - WENDJA, C. (2007). « Construction d'une table de mortalité prospective à partir d'un échantillon réduit ».
 - WERTEL, S. (2015). « Construction de tables de mortalité par ajustement dans le cadre d'un régime de retraite en point ».
 - WILMOTH, J., ANDREEV, K., GLEI, D. et RIFFE, T. (2021). « Methods protocol for the human mortality database ».

Annexe A

Modèle de Lee-Carter

Le modèle de Lee-Carter permet d'estimer la mortalité prospective, La dynamique de ce modèle est donnée par l'équation suivante :

$$\log \hat{\mu}_x(t) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

Avec :

- α_x qui correspond à la forme moyenne de la force de mortalité qui augmente avec l'âge.
- β_x la déviation de la mortalité par rapport au comportement moyen α_x .
- κ_t est une série temporelle décrivant l'évolution de la mortalité au cours du temps.
- $\epsilon_{x,t}$ correspond aux résidus du modèle. Il sont supposés gaussiens de moyenne nulle et de variance σ^2 .

Ce modèle a des contraintes sur ses paramètres¹ :

$$\sum \beta_x = 1 \quad \sum \kappa_t = 0$$

Ces contraintes permettent au modèle d'être identifiable, et donc de pouvoir avoir de bonnes propriétés statistiques.

Il faut ensuite ajuster ce modèle à nos données en utilisant la méthode des moindres carrés ordinaires et ainsi obtenir des estimateurs pour chacun des trois paramètres.

1. Une explication des contraintes est notamment développée dans BOURREAU (2014).

Annexe B

Autres modèles d'ajustement aux grands âges

B.1 Modèle de Kimeldorf-Jones

Cette méthode a été étudiée par Kimeldorf et Jones en 1967¹. Elle s'applique principalement pour les lois de maintien en incapacité/invalidité, mais est adaptable aux tables de mortalité. Elle repose sur une approche bayésienne.

La densité de la loi $f_{\hat{q}|q}$, étant un produit de lois binomiales, elle peut être approchée par une loi normale de moyenne q puisque les estimateurs \hat{q} sont non-biaisés. On peut donc écrire :

$$f_{\hat{q}|q}(\hat{q}|q) = \frac{1}{\sqrt{2\pi^p} \sqrt{\det(W)}} \exp\left(-\frac{1}{2}{}^t(\hat{q} - q)W^{-1}(\hat{q} - q)\right)$$

Avec W la matrice de covariance des \hat{q}_x . Elle est diagonale, dont les éléments sont les variances des lois binomiales associées, à savoir $L_x q_x (1 - q_x)$. Afin d'être en mesure de faire les calculs et, étant donné que les taux q_x ne sont pas connus nous les remplacerons par leurs estimateurs \hat{q}_x .

Il reste à faire l'hypothèse de la loi a priori. Pour cette loi, une loi Gamma de paramètres (k, θ) , qui généralise la loi exponentielle aurait été appropriée. Ces deux lois sont beaucoup utilisées en analyse de survie, étant donnée qu'elles sont seulement définies sur des valeurs positives et peuvent donc assez bien représenter le temps. De plus, l'évolution des taux de mortalité correspond assez bien à la densité de ce type de distributions. Cette loi Gamma pose cependant problème dans notre cas puisqu'il s'agit d'une loi univariée. On peut cependant l'approcher par une loi Gaussienne de

1. Cf. KIMELDORF et JONES (1967).

paramètres $\mu = (k - 1)\theta$ et $\sigma^2 = (k - 1)\theta^2$ dans le cas où k est assez grand. La loi a priori choisie sera donc également une loi normale de paramètres μ et V (supposés connus). On obtient donc :

$$f_q(q) = \frac{1}{\sqrt{2\pi^p} \sqrt{\det(V)}} \exp\left(-\frac{1}{2}{}^t(q - \mu)V^{-1}(q - \mu)\right)$$

La quantité $f_{\hat{q}}$, ne dépendant pas de q , n'interviendra donc pas dans le calcul. On peut alors écrire que :

$$\begin{aligned} f_{q|\hat{q}}(q|\hat{q}) &= \frac{f_{\hat{q}|q}(\hat{q}|q)f_q(q)}{f_{\hat{q}}(\hat{q})} \\ &\propto \exp\left(-\frac{1}{2}{}^t(\hat{q} - q)W^{-1}(\hat{q} - q)\right) \exp\left(-\frac{1}{2}{}^t(q - \mu)V^{-1}(q - \mu)\right) \\ &\propto \exp\left(-\frac{1}{2}({}^t(\hat{q} - q)W^{-1}(q - \mu) + {}^t(q - \mu)V^{-1}(q - \mu))\right) \end{aligned}$$

En factorisant à l'intérieur de l'exponentielle, on trouve :

$$f_{q|\hat{q}}(q|\hat{q}) \propto \exp\left(-\frac{1}{2}{}^t(q - z)A^{-1}(q - z)\right)$$

Avec : $A^{-1} = (V^{-1} + W^{-1})^{-1}$ et $z = (V^{-1} + W^{-1})^{-1}(W^{-1}\hat{q} + V^{-1}\mu) = A^{-1}(W^{-1}\hat{q} + V^{-1}\mu)$.

Ce résultat montre, d'une part, que $q|\hat{q}$ suit une loi normale de paramètres (z, A) et, d'autre part, permet de trouver les taux lissés. Afin de les obtenir, on cherche à maximiser la quantité ci-dessus et on trouve donc que : $q_{lisse} = z$.

La principale difficulté pratique de cette méthode est de connaître les paramètres μ et V de la loi a priori.

B.2 Modèle à noyau

Pour mettre en place ce type de modèle, il faut repartir de la définition de la fonction de répartition empirique. On cherche ici à l'estimer à l'aide d'une fonction régulière, qui permettra à l'estimation de la fonction de survie associée d'être dérivable. Il est possible d'estimer la fonction de répartition empirique comme étant :

$$\hat{F}_{n,h}(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)$$

La fonction K est appelée noyau. Cette méthode consiste à choisir un noyau à la

fois continue et dérivable, positif et ayant une intégrale égale à 1.

Les noyaux souvent utilisés en pratique sont la densité d'une loi normale centrée réduite – i.e $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ –, ainsi que le noyau d'Epanechnikov – i.e $K(x) = \frac{3}{4}(1 - x^2)1_{|x| \leq 1}$.

Cependant, ce type de modèles est très sensible à la censure, ce qui risque de biaiser les résultats obtenus. Il convient alors, comme proposé dans MARRON et PADGETT (1987), de réajuster l'estimation de la manière suivante :

$$\hat{f}_{n,h}(x) = \frac{1}{h} \sum_{j=1}^n \frac{d_j}{n_j} K\left(\frac{X_j - x}{h}\right)$$

Avec $\frac{d_j}{n_j}$ le saut de l'estimateur de Kaplan-Meier.

Un point important de cette méthode est également de déterminer le paramètre h (aussi appelé « bande passante ») de manière optimale.