



Mémoire présenté le :

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

Par : Anne-Sophie Musset

Titre Risque de longévité : la référence à la tendance de la population nationale est-elle justifiée ?

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

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

Membre présents du jury de l'Institut des Actuaire signature

Entreprise :

Nom : Périclès Actuarial

Signature :

Directeur de mémoire en entreprise :

Nom :

Signature :

Invité :

Nom : Denis Bourgeois

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

[Empty signature box for responsible company]

Signature du candidat

[Empty signature box for candidate]

Secrétariat

Bibliothèque :

Remerciements

Je souhaite remercier en tout premier lieu Frédéric Planchet, tant pour m'avoir donné l'opportunité de ce sujet de mémoire que pour sa disponibilité et ses conseils avisés tout au long de l'étude, ainsi que Julien Tomas pour son temps et ses avis particulièrement éclairants.

Je remercie particulièrement l'ISFA pour la mise en place d'un programme de VAE offrant l'opportunité aux professionnels de pouvoir accéder au Diplôme d'actuaire « a posteriori », et de bénéficier d'un enseignement de qualité.

Je voudrais également dire merci à Denis Bourgeois, pour son écoute et sa clairvoyance, ainsi qu'à toute l'équipe Périclès Actuarial, avec qui j'ai tant plaisir à travailler, pour son soutien et sa bienveillance !

Sans oublier -M- pour tes corrections avisées, -S- pour ta potion magique pour introduction et conclusion, -S- pour tes multiples relectures, et bien évidemment -3F- pour les escapades en mer que j'espère éternelles...

Comme le cumul d'un mémoire avec (tout !) le reste n'est jamais évident, j'ai une pensée toute exclusive pour mes proches. Merci d'avoir compris qu'il est de ces périodes où on est moins présent et disponible mais qui ne changent en rien la valeur essentielle des « gros cailloux ». « Et crever le silence... », je dédie d'ailleurs cette longévité à Manon, Camille et au futur petit bout des M&M's(c).

Je conclurai par un dernier clin d'œil à Jung sans qui la saveur de ce mémoire, à bien des égards, n'aurait pas été la même !

Résumé

Le risque de longévité se décompose sous plusieurs formes : du risque d'échantillonnage, au risque opérationnel en passant par les risques systématiques. Parmi ces derniers, celui d'estimation de la tendance pèse un poids non négligeable dans l'évaluation du choc de longévité sous Solvabilité 2. Cette même Directive impose par ailleurs une estimation au plus juste des engagements par le biais de la provision *best estimate*.

Ainsi, certaines critiques pourraient être émises dans la mesure où la tendance des tables de mortalité de place est usuellement extrapolée, par défaut, sur celle de la population nationale. Le manque d'informations centralisées sur la mortalité des assurés en est assurément l'une des principales raisons. La mise à disposition de données anonymisées d'assurés, issues de plusieurs organismes, offre l'opportunité d'évaluer la tendance de mortalité propre d'un échantillon significatif de place. Ces données sont en outre celles qui ont été utilisées par le groupe de travail « mortalité » de l'Institut des Actuaire pour notamment construire les tables de référence *best estimate* en 2014.

La première partie de ce mémoire fixe les points de repères démographiques et mathématiques essentiels pour une bonne compréhension de la démarche entreprise. Afin de construire le portefeuille agrégé, la régression de Cox est exploitée pour en gommer l'hétérogénéité tout en préservant la profondeur d'historique essentielle à l'objectif poursuivi. La troisième partie analyse la tendance propre obtenue à l'appui de deux modèles endogènes (Log-Poisson et GLM Poisson avec dépendance en âge et en année) et met en évidence les biais liés à cette estimation. Enfin, un éclairage sera apporté quant à la pertinence de la tendance des tables de référence assurantielles dans une approche au plus juste du calcul des provisions.

Abstract

From sampling risk to operational risk through systematic risks, the longevity risk takes several forms. Among the systematic risk, trend estimation has a significant weight on the longevity stress valuation used under Solvency II. Moreover, this Directive requires insurers to fully understand their risks by assessing a *best estimate* provision.

Thus, it appears open to criticisms that the trend of the reference prospective tables is mostly extrapolated from the one built on national population data. The lack of a centralised data on insured population is most likely one of the main causes. The availability of unnamed portfolios of different insurance companies is a real opportunity to assess the real mortality trend of a significant sample of insured people. These data were used by the Institut des Actuaire's working group on « mortality » to build the reference *best estimate* tables in 2014.

After an overview of the demographic and mathematical tools needed to understand the argumentation process, the paper presents the methodology employed to build the aggregated portfolio. To this aim, the Cox regression is used to take into account heterogeneity while keeping the historical depth of data required for the study. The last part of the paper is dedicated to the analysis of the resulted trend through two endogenous models (Log-Poisson and GLM Poisson depending on age and calendar year) and the highlight of the bias induced by this estimation. Finally, an additional perspective will be given regarding the relevance of the trend of the reference mortality tables considering a best estimate approach.

TABLE DES MATIERES

Introduction	8
Partie 1 : Points de repères démographiques et mathématiques	11
A. Notions sur la mortalité et son évolution	12
1- Taux de mortalité moyen ou central.....	12
2- Deux autres terminologies majeures : quotient de mortalité et force de mortalité .	14
3- Diagramme de Lexis et mise en perspective générationnelle	16
B. Espérance de vie	18
1- Définitions	18
2- Observations.....	18
3- Perspectives d'évolution	20
4- Inégalités face à la longevite.....	21
C. Mise en perspective des modèles utilisés.....	24
1- Modèles paramétriques.....	24
2- Modèles semi-paramétriques : exemples des modèles relationnels.....	26
3- Modèles non paramétriques	28
Conclusion partie 1.....	28
Partie 2 : Tables prospectives assurantielles et omnipresence de la tendance de la population nationale.....	29
A. Tables prospectives et références	30
1- Les tables de référence de la population nationale.....	30
2- Les tables de références assurantielles	33
3- Comparaison des tables assurantielles et nationales.....	37
4- Enjeu des assureurs en regard du risque de longevite.....	40
B. Construction de la table des moments du « portefeuille des assurés »	42
1- Les données	42
2- Prise en compte de l'hétérogénéité par application du modèle de Cox.....	45

C.	Application du Modèle de Brass – Justification d’une tendance « assurés » propre	52
1-	Théorie	52
2-	Table des moments ajustée	53
3-	Table prospective	57
4-	Pertinence de la tendance issue des taux ajustés	58
	Conclusion partie 2	59
	Partie 3 : Evaluation de la tendance du portefeuille d’assurés agrégé	60
A.	Table des moments ajustée avec un modèle endogène de type GLM Poisson	60
1-	Théorie	60
2-	Résultats	61
3-	Etude d’impact des portefeuilles utilisés pour l’agrégation	64
4-	Conclusion	68
B.	Table des moments ajustée avec un modèle endogène de type Log Poisson	70
1-	Théorie	70
2-	Résultats pour l’ensemble du portefeuille	71
3-	Résultats ajustés de l’impact du portefeuille 6	74
4-	Comparaison et conclusion	76
C.	Comparaison des tables prospectives obtenues par rapport aux références	79
1-	Comparaison des tendances	79
2-	Comparaison des tendances sur une période d’ajustement restreinte	81
	Conclusion partie 3	86
	Conclusion	87
	Bibliographie	90
	ANNEXES	92
	A – Récapitulatif des retraitements réalisés sur la base de données	92
	B – Résultats détaillés de la régression de Cox Femmes / Hommes	93

INTRODUCTION

« N'attends pas que les évènements arrivent comme tu le souhaites ; décide de vouloir ce qui t'arrive et tu seras heureux ». Epictète aurait-il imaginé que ses préceptes de stoïcien sur cet état de plénitude qu'est le Bonheur puissent, encore plus de 2000 ans après sa mort, influencer la pensée humaine mais surtout son espérance de vie ?

En effet, être heureux aiderait l'homme à vivre plus longtemps. C'est en tout cas la conclusion de plusieurs études, comme la plus célèbre d'entre elles, la Harvard Grant Study qui recueille méthodiquement depuis 1938 autant d'informations que possible, basées sur l'expérience réelle d'une cohorte de 200 hommes, pour notamment identifier les secrets de longévité¹. Le bonheur serait ainsi une variable explicative de l'augmentation de la durée de vie constatée depuis toujours et dans le monde entier. Toutefois, le secret de la longévité ne se résume pas (hélas) au seul « état de bonheur » et de nombreux autres paramètres, plus significatifs, comme l'âge ou l'année calendaire, entrent en jeu dans cette modélisation complexe qui est celle de la mortalité.

Un assureur vie se doit de tarifer et provisionner le risque sous-jacent des garanties proposées à ses assurés, tant en cas de vie qu'en cas de décès. Ces évaluations passent par l'exploitation de tables de mortalité prospectives qui sont usuellement élaborées à l'appui de modèles mathématiques et d'hypothèses permettant d'expliquer la mortalité - **son niveau** - et d'en caractériser son évolution - **sa tendance**. Plus récemment, l'entrée en vigueur de la Directive Solvabilité 2 impose la quantification du risque de longévité, à l'appui d'une valorisation au plus juste de ses engagements, la provision *best estimate*, et de l'application d'un choc de longévité².

Si les tables de mortalité prospectives d'assurance, réglementaires ou d'expériences, déterminent le niveau de mortalité des assurés, la tendance est en revanche estimée à partir de celle observée pour la population nationale. Plusieurs arguments peuvent être avancés pour justifier d'une telle démarche. Tout d'abord, la probabilité est généralement élevée qu'une sous-population partage les mêmes caractéristiques que sa population globale d'appartenance. D'autre part, il n'existe pas de base de données centralisée des décès des assurés, à l'image de celle construite sur les fonds de pensions au Royaume-Uni (maintenue

¹ VAILLANT G. [2012] « *Triumphs of Experience* », citation « *Happiness is love. Full Stop.* ».

² EIOPA [2014] « *Delegated Acts* », Choc de 20% appliqués uniformément sur les taux de mortalité.

par le CMI, Continuous Mortality Investigation). Enfin, et même si une tendance propre de la mortalité des assurés pouvait être statistiquement évaluée, aucun élément factuel ne permettrait d'en attester de sa justesse prospective. Ainsi, la tendance des tables de place est extrapolée, par défaut, à partir de celle de la population nationale mesurée par l'INSEE dans la mesure où cette dernière rassemble un nombre de données et une profondeur temporelle largement plus conséquente.

La tendance réelle et spécifique d'une population d'assurés n'a donc que très rarement été étudiée à ce jour. Pourtant, dans un contexte Solvabilité 2, le poids du risque d'estimation de la tendance est non négligeable dans l'évaluation du choc de longévité : il représente environ 40% à 55 ans et décroît progressivement pour atteindre environ 10% à 85 ans³. Ainsi, la possibilité, dans le cadre de ce mémoire, d'utiliser le regroupement d'un ensemble de données issues de plusieurs organismes assurantiels offre l'opportunité d'évaluer cette tendance sur un échantillon significatif de la place.

L'exploitation de ce type de données, comme un tout, nécessite l'atténuation des disparités entre organismes, issues aussi bien des périodes d'observation distinctes, que des inégalités face à la mortalité des assurés (sociales, médicales, géographiques...) ou que de l'anti-sélection liée aux contrats (asymétrie d'information, garantie, caractère optionnel...). Les discontinuités ainsi induites sont usuellement corrigées par la prise en compte d'une période d'observation restreinte et commune à l'ensemble des organismes, ce qui est ici incompatible avec l'objectif poursuivi. La régression de Cox procure une solution alternative intéressante pour préserver l'historique de données. L'application de ce modèle ainsi que la justification de l'existence d'une tendance propre du portefeuille d'assurés étudié feront l'objet de la **deuxième partie**.

Les conclusions de l'analyse de la tendance de ce portefeuille agrégé d'assurés seront enfin élaborées en **troisième partie**, à l'appui de deux modèles : un modèle semi-paramétrique construit dans le cadre d'un modèle linéaire généralisé de Poisson (GLM Poisson) intégrant une dépendance à l'âge et à l'année calendaire, et un modèle paramétrique de type Log-Poisson. Ces deux références ont précisément été sélectionnées dans le cadre de cette étude,

³ DUCROS JC., LEROY G., PLANCHET F. [2015] « Risques systématiques associés aux tables d'expérience », Groupe de Travail Mortalité de l'Institut des Actuaire, IA. Résultat « tendance 6 ans ».

pour leur caractère endogène, pour leur robustesse⁴, ainsi que pour leur approche à la fois différente et complémentaire qui permettra de justifier les biais d'estimation rencontrés.

Avant d'aborder la **première partie** de ce mémoire qui fixe les points de repères démographiques et mathématiques essentiels pour la bonne compréhension de la démarche entreprise, il paraît nécessaire de préciser que les travaux présentés ci-après n'ont ni pour vocation d'élaborer une nouvelle table de mortalité de place de référence, ni d'aborder la longévité avec un regard sur l'état de dépendance prenant en compte un critère de bonne ou de mauvaise santé au décès, ni d'approfondir les méthodes de couverture du risque de longévité.

L'objectif de ce mémoire est en définitive de pouvoir challenger la pertinence de la tendance des tables de références assurantielles quant à une approche au plus juste ou *best estimate* du calcul des provisions, tout en mettant en évidence les biais liés à cette estimation.

⁴ Robustesse liée au cadre d'un GLM pour le modèle GLM Poisson, et justifiée par une approche statistique plus juste que le modèle Lee-Carter, détaillée par BROUNHS et Al. [2002] pour le modèle Log-Poisson.

PARTIE 1 : POINTS DE REPERES DEMOGRAPHIQUES ET MATHÉMATIQUES

L'étude publiée en 2012 par Oskar Burger, Annette Baudisch et James W.Vaupel⁵ a démontré que la diminution de la mortalité de l'homme a été plus conséquente durant les 100 dernières années que lors de son évolution à Homo Sapiens : l'espérance de vie d'un homme de 72 ans de nos jours correspondrait à celle d'un chasseur-cueilleur de 30 ans il y a 10 000 ans. Aujourd'hui encore, la durée de vie continue d'augmenter progressivement : chaque année, les nouveau-nés français gagnent en moyenne 2 à 3 mois de durée de vie. Mais qu'en sera-t-il demain ?

Si d'une manière générale, le passé ne peut pas nécessairement préjuger de l'avenir, les modèles mathématiques prédictifs s'appuient néanmoins sur un historique de données.

Dans cette perspective, cette première partie s'attachera à présenter les modèles envisageables pour évaluer la tendance de la mortalité de l'échantillon significatif de place étudié et la comparer à celle de la population nationale.

Au préalable, il est important de préciser les notions essentielles à la compréhension de l'évolution de la mortalité ainsi qu'à la formalisation des modèles mathématiques sous-jacents. Pour compléter cette approche, un point de vue démographique et médical viendra en outre enrichir cette partie.

⁵ BURGER O., BAUDISCH A., VAUPEL JW. [2012] « Human mortality improvement in evolutionary context »

A. NOTIONS SUR LA MORTALITE ET SON EVOLUTION

1- TAUX DE MORTALITE MOYEN OU CENTRAL

i. Evolution selon l'année d'observation

En démographie, un taux de mortalité correspond au rapport entre le nombre annuel de décès et la population totale moyenne sur une période observée. Usuellement, cette définition correspond plus spécifiquement au *taux de mortalité moyen ou central*, et il est distingué par âge x et par période d'observation. Ce taux sera noté m_x pour une période d'observation annuelle communément employée.

L'observation de l'évolution du taux de mortalité moyen annuel en France par âge (dont un extrait est présenté ci-dessous) confirme bien la décroissance reconnue de la mortalité. Depuis 1950, progrès techniques et scientifiques, amélioration de l'hygiène de vie, évolutions en termes de médecine, avancées dans le traitement de maladies sont les raisons de cette baisse observée. Cette tendance est bien évidemment antérieure à 1950, mais la période considérée permet une analyse relativement neutre de tout événement majeur sur le territoire français, de type catastrophe naturelle, épidémie, guerre...

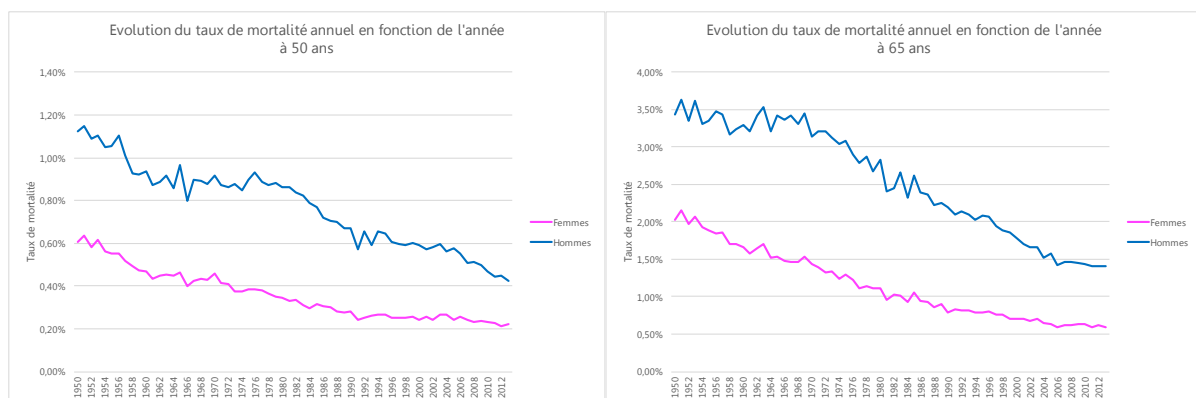


Figure 1 – Evolution du taux de mortalité à 50 et 65 ans des françaises et des français
- Source : Human Mortality DataBase

La Figure 1 illustre, à 50 et 65 ans, les constats, également vérifiés à tout âge, selon lesquels la mortalité des femmes est plus faible par rapport à celle des hommes, et que cet écart a tendance à se réduire. Il semblerait que cette convergence s'explique par la baisse de la mortalité liée au tabac et au cancer chez les hommes.

Par ailleurs, si l'évolution globale se montre assez linéaire, ou linéaire par morceaux, les taux de mortalité, ici bruts de tout retraitement, sont relativement volatiles. Cette volatilité est

qualifiée d'intrinsèque lorsqu'elle est la conséquence de facteurs externes annuels ponctuels (épidémie, catastrophe naturelle...). Elle peut également être la conséquence d'une fluctuation d'échantillonnage liée notamment à l'hétérogénéité du portefeuille.

ii. Positionnement, tendance et volatilité

Positionnement, tendance et volatilité sont les trois composantes permettant d'expliquer l'évolution de la mortalité :

- Le positionnement ou le niveau de mortalité correspondant à la valeur des taux de mortalités actuels ;
- La tendance qui se matérialise par la pente de l'évolution des taux de mortalité ;
- La volatilité identifiée par le « bruit ».

La comparaison de l'évolution de la mortalité de deux populations se matérialise ainsi par un écart de positionnement et un écart de tendance, comme l'illustre la figure ci-dessous entre les femmes et les hommes.

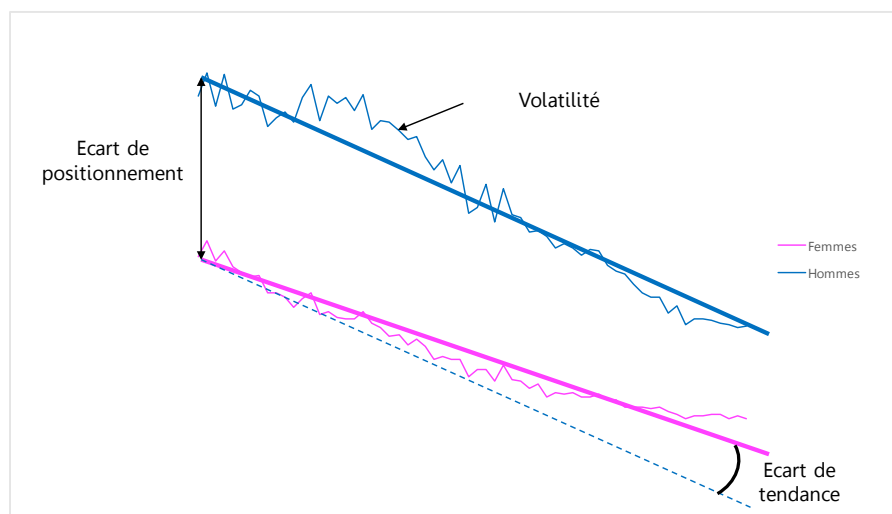


Figure 2 – Caractérisation de l'évolution de la mortalité

iii. Evolution en fonction de l'âge

L'évolution des taux de mortalité en fonction de l'âge fait apparaître quant à elle une forme exponentielle aux âges adultes qui conduit usuellement à les représenter par une échelle logarithmique.

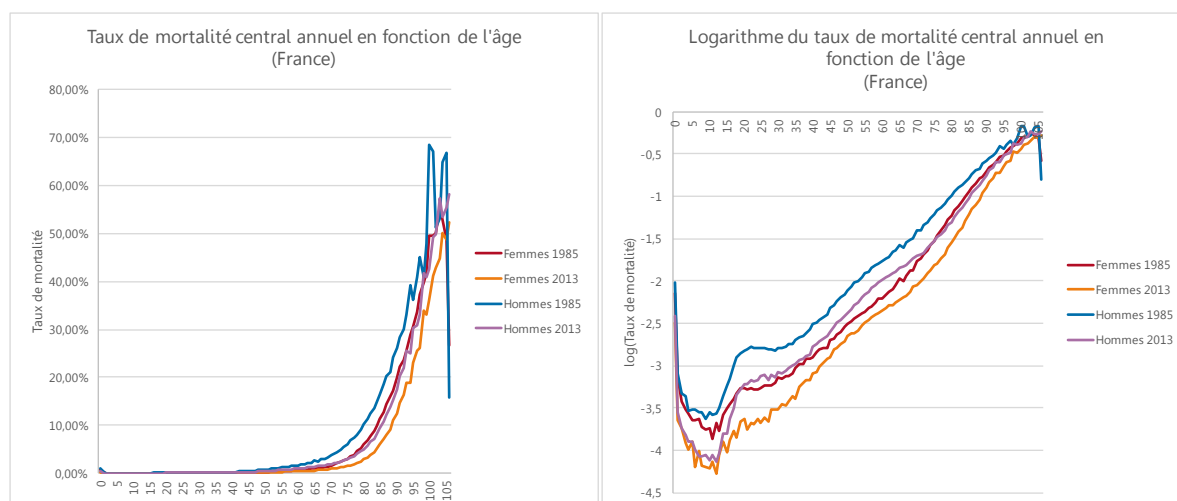


Figure 3 – Taux de mortalité en fonction de l'âge en 1985 et 2013 des françaises et des français
- Source : Human Mortality DataBase

Cette échelle permet ici également d'apprécier l'évolution de la structure de la mortalité sur les 30 dernières années. Ainsi, si certains modèles mathématiques, détaillés par la suite, peuvent reproduire la mortalité réelle d'une population, cette adéquation ne reste généralement valide qu'un temps et nécessite alors un réajustement, de paramétrage ou de modèle, à horizon d'une dizaine ou vingtaine d'années.

Ce mémoire se concentrera sur l'analyse de la tendance et ne cherchera donc pas à expliquer les formes de la mortalité à tout âge, comme le maximum local avant 30 ans⁶.

2- DEUX AUTRES TERMINOLOGIES MAJEURES : QUOTIENT DE MORTALITE ET FORCE DE MORTALITE

La notion de mortalité ne se limite pas au taux de mortalité central. Elle fait également intervenir deux fonctions mathématiques essentielles, employées dans de nombreux modèles, à savoir, le quotient de mortalité et la force de mortalité.

⁶ DEBONNEUIL E. [2014] « Modèle paramétrique de mortalité en fonction de l'âge pour des applications sur un portefeuille de retraite », reproduction de la forme de la mortalité à l'aide de trois modèles selon tranche d'âge.

i. Quotient de mortalité

Le taux de mortalité central, tel qu'explicité, présente une dimension inverse d'une unité de temps, tandis que le quotient de mortalité ${}_tq_x$ est sans dimension et correspond à la probabilité de décéder entre les âges x et $x+t$. Il se définit de la manière suivante :

$${}_tq_x = 1 - {}_tp_x = 1 - \frac{L_{x+t}}{L_x}$$

avec :

${}_tp_x$: Probabilité de survie entre l'âge x et l'âge $x+t$

L_x : Fonction de survie de la durée de vie résiduelle d'un individu sachant que ce dernier est en vie à l'âge x – ou nombre de personnes encore en vie à l'âge x

ii. Quotient de mortalité versus taux de mortalité central

Il convient de ne pas confondre le taux de mortalité central et le quotient de mortalité, bien que les graphiques ci-dessous mettent en évidence que ce n'est qu'aux âges particulièrement avancés (après 90 ans) que l'écart est particulièrement notable.

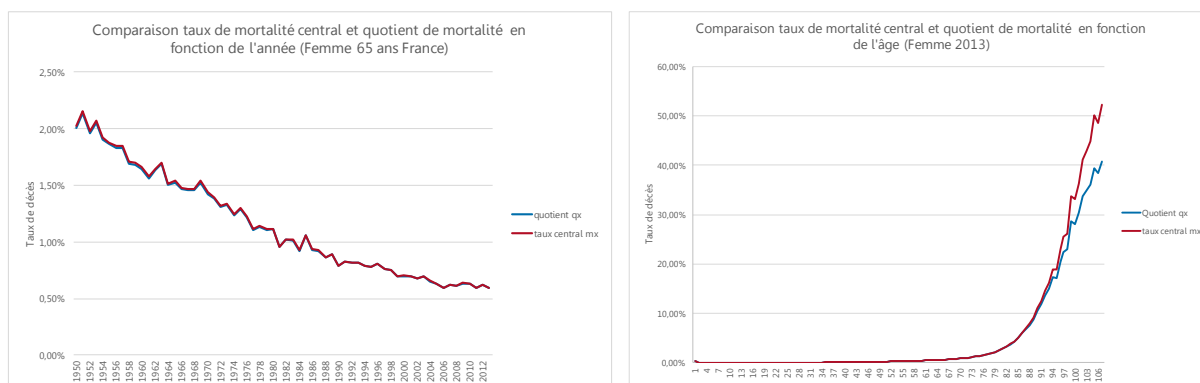


Figure 4 – Comparaison taux de mortalité central et quotient de mortalité

La littérature emploie couramment le terme « taux » en lieu et place du terme « quotient ». Dans la mesure où l'étude qui suit se concentrera sur la tranche d'âge 30 à 90 ans, le terme « taux » sera également employé.

iii. Force de mortalité ou taux de mortalité instantané

La force de mortalité à l'âge x de l'année calendaire t , $\mu_x(t)$, est définie par :

$$\mu_x(t) = \lim_{h \rightarrow 0} \frac{{}_{t+h}q_x - {}_tq_x}{h \cdot {}_tp_x} = \frac{\frac{d}{dt} {}_tq_x}{{}_tp_x}$$

Par intégration, le taux ou quotient de mortalité s'exprime en fonction de la force de mortalité de la manière suivante.

$${}_tq_x = 1 - {}_tp_x = 1 - e^{-\int_0^t \mu_{x+s} ds}$$

En considérant que la force de mortalité est constante par morceaux, à savoir que :

$$\mu_{x+k}(t+h) = \mu_x(t) \text{ pour } 0 \leq h < 1 \text{ et } 0 \leq k < 1 \text{ et } x, t \text{ entiers}$$

Le taux de mortalité s'écrit : ${}_tq_x = 1 - {}_tp_x = 1 - e^{-\mu_x(t)}$, avec $\mu_x(t)$ équivalent au taux de mortalité central $m_x(t)$.

Il est à noter que cette hypothèse sera considérée dans la suite de ce mémoire.

3- DIAGRAMME DE LEXIS ET MISE EN PERSPECTIVE GENERATIONNELLE

La perception des subtilités temporelles liées au taux de mortalité peut-être plus facile à appréhender graphiquement. Le diagramme de Lexis en schématise les trois paramètres essentiels :

- L'âge, correspondant à l'ordonnée ;
- L'année d'observation, représentée en abscisse ;
- La génération d'appartenance, caractérisée par la bissectrice.

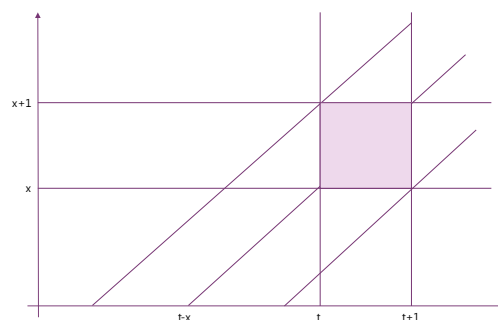


Figure 5- Principe du diagramme de Lexis

L'hypothèse de constance par morceaux de la force de mortalité consiste à considérer que cette dernière est constante dans les carrés unitaires tel que celui colorié dans le précédent diagramme.

Les taux de mortalité jusqu'ici mentionnés sont plus précisément appelés taux de mortalité transversaux, à distinguer des taux de mortalité longitudinaux qui prennent en considération la vision générationnelle des individus (bissectrice du diagramme). Ces deux taux de mortalité se déduisent l'un de l'autre par simple changement de variable⁷.

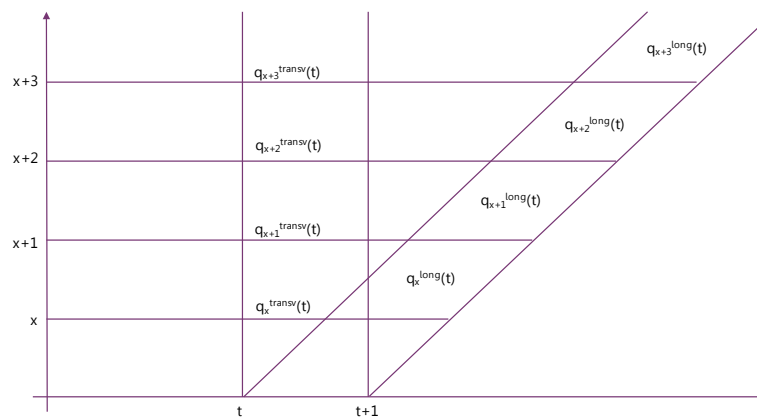


Figure 6- Illustration du principe des taux de mortalité longitudinaux et transversaux

Jusqu'à présent l'existence d'un effet générationnel significatif dans l'étude des taux de mortalité n'a pas été démontrée en France. Il n'a pas été constaté de progrès médical qui n'aurait concerné qu'une cohorte spécifique. Ni épidémie qui aurait conduit à constater des décès pour une unique cohorte à une date donnée et ultérieurement. Une étude récente⁸ sur la qualité et l'exploitation commune des données de la HMD (Human Mortality Database) et de la HFD (Human Fertility Database) a d'ailleurs permis de démontrer qu'en retraitant le calcul de la mortalité à la naissance et en le répercutant aux âges adultes, cet effet cohorte légèrement observable en était considérablement atténué.

Nous nous focaliserons dans la suite sur l'évaluation et l'analyse des taux de mortalité transverses, tant pour le portefeuille étudié que pour les références de mortalité employées.

⁷ Par le changement de variable $t' = t-x$, à savoir $q_x^{\text{transv}}(t) = q_x^{\text{long}}(t-x)$

⁸ Etude présentée dans le cadre de la chaire « Actuariat durable et stabilité du secteur de l'assurance à long terme » (Milliman R&D, Laboratoire ISFA, Institut Louis Bachelier) – Décembre 2015

B. ESPERANCE DE VIE

1- DEFINITIONS

D'une manière générale, les taux de mortalité sont assez peu « parlants » et la terminologie la plus employée pour décrire l'évolution de la longévité reste l'espérance de vie. Tout comme pour les taux de mortalité, l'espérance de vie générationnelle, qui ne sera pas employée dans le cadre de ce mémoire, permet d'appréhender le nombre d'années restant à vivre à un individu d'un âge donné et d'une génération donnée.

Notre attention se porte sur l'espérance de vie (ou espérance de vie instantanée) résiduelle à l'âge x . Elle se définit comme le nombre moyen d'années restant à vivre au-delà de cet âge x , dans les conditions de mortalité par âge de l'année considérée. Cela revient donc à considérer une génération fictive représentée par tous les âges d'une année calendaire donnée. Du fait des perspectives de longévité, l'espérance de vie (instantanée) est aujourd'hui inférieure à l'espérance de vie générationnelle.

L'espérance de vie, équivalente à la durée de survie moyenne à un âge x , est intrinsèquement liée aux taux de mortalité et peut s'exprimer de la manière suivante, avec les notations précédemment explicitées :

$$e_x = \sum_{h>0} \frac{L_{x+h}}{L_x} = \sum_{h>0} (1 - {}_h q_x)$$

2- OBSERVATIONS

En un siècle, l'espérance de vie à la naissance (pour un âge « nul » donc) a considérablement augmenté : hommes et femmes ont gagné quasiment 15 ans de vie en moyenne en France. Il existe encore une disparité non négligeable entre hommes et femmes puisque les femmes ont plus de 6 années d'avance sur l'espérance de vie des hommes (85 ans contre 78,8 ans). Comme observé sur les taux de mortalité, cet écart semble progressivement diminuer. Mais, selon les perspectives actuelles, 60 années seraient tout de même encore nécessaires pour atteindre une égalité hommes/femmes.

L'amélioration de l'espérance de vie à la naissance est aujourd'hui principalement expliquée par la lutte contre la mortalité adulte, en particulier aux âges élevés où se concentrent de plus en plus les décès, compte tenu d'une mortalité infantile particulièrement basse.

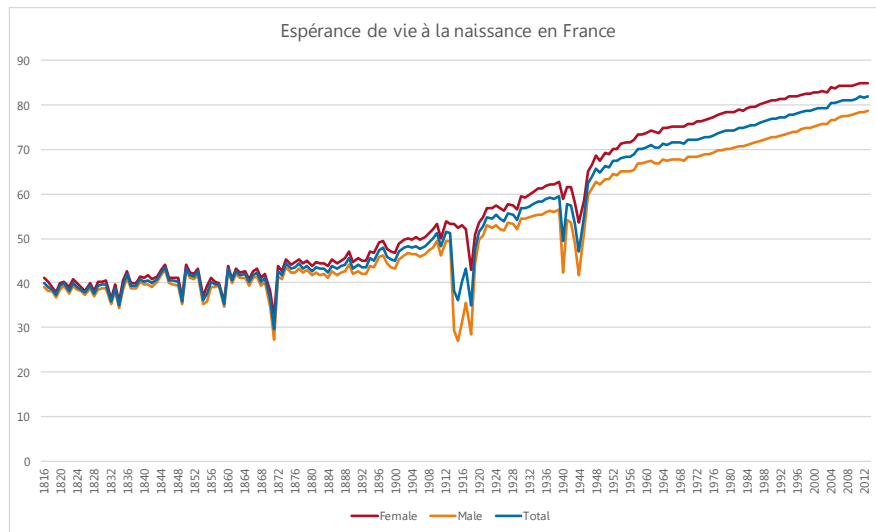


Figure 7 – Espérance de vie à la naissance en France - Source : Human Mortality DataBase

Les maladies cardio-vasculaires et les cancers sont désormais les principales causes de décès à ces âges adultes. De premiers succès rencontrés dans la lutte contre ces maladies ont fait reculer la mortalité des adultes et des personnes âgées à partir des années 1970 (diagnostics plus précoces, recul des comportements à risques comme le tabagisme et l'alcoolisme...). Plus récemment, les progrès qui ont le plus accru l'espérance de vie ont surtout été réalisés aux âges avancés : autour de 80 ans pour les femmes et 70 ans pour les hommes, principalement à travers une amélioration du traitement des tumeurs et des maladies de l'appareil respiratoire.

Depuis la moitié du XXème siècle, la tendance constatée sur l'espérance de vie à la naissance en France est d'environ un trimestre par an.

Cette augmentation de longévité, mise en évidence sur la France, se retrouve en quelque sorte au niveau mondial. En effet, c'est à partir de cette tendance linéaire que le démographe et fondateur du Max Planck Institute, James W. Vaupel a introduit la notion de « Best Practice line ». Jusqu'en 1840, dans aucun pays du monde, les personnes ne vivaient plus de 45 ans en moyenne. Par la suite, cette espérance de vie maximale a augmenté en ligne droite, selon une pente de 0,243.

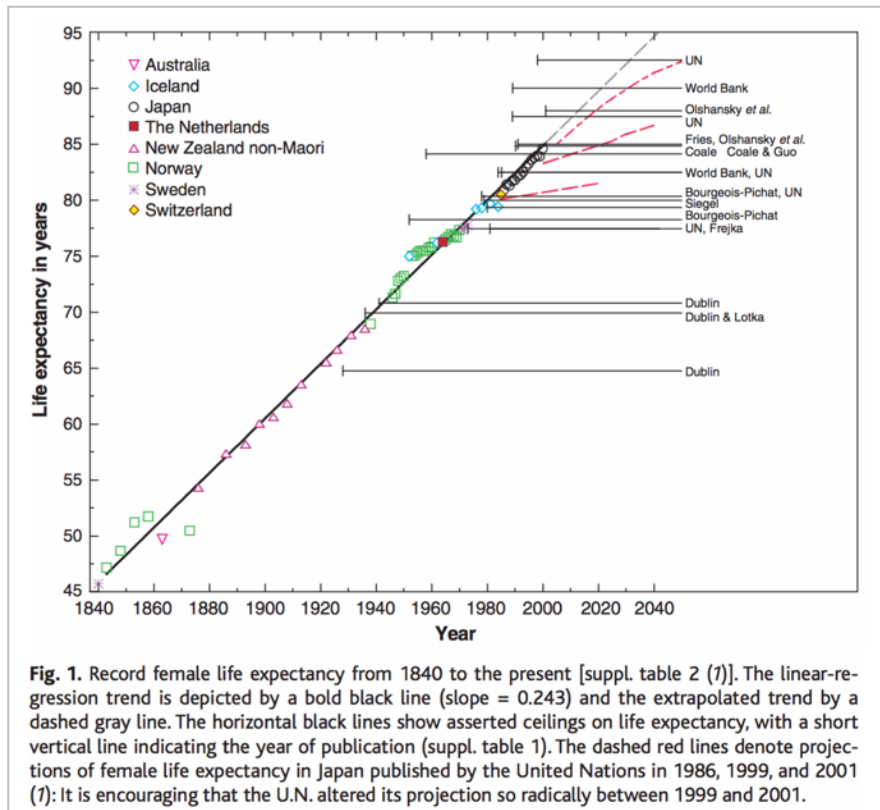


Figure 8 – Best Practice line - OEPPEN J., VAUPEL J.W. [2002] « Broken Limits to Life Expectancy », Science, vol. 296

3- PERSPECTIVES D'ÉVOLUTION

Même si l'évolution linéaire constatée depuis de nombreuses années sur l'espérance de vie pourrait présager d'une longue stabilité, toute perspective reste cependant envisageable à l'avenir :

- Evolutions médicales dans le traitement des maladies les plus pénalisantes : la médecine modélise notamment avec intensité le vieillissement afin d'identifier ces pathologies les plus pénalisantes et les traitements curatifs associés, avec une attention particulière à ce jour sur les risques cardio-vasculaires, l'obésité et le diabète ;
- Voire une évolution génétique de l'Homme, qui pourrait par exemple régénérer naturellement ses tissus à des âges élevés, le rendant moins sensible aux effets de la vieillesse, comme cela peut être observé chez les animaux aujourd'hui ;
- Développement de risques existants ou de nouveaux risques : pollutions, réchauffement climatique, ondes / hyper connexion, tabagisme, obésité ;
- Voire une évolution du génome à l'opposé d'une régénérescence des tissus allant dans le sens d'une sélection naturelle plus accrue ;

- Recrudescence d'événements majeurs plus ou moins directement d'origine humaine (guerre, catastrophe naturelle...).

Sans compter que dans les années à venir, l'une des considérations majeures sera probablement celle de l'évolution de l'espérance de vie en bonne santé par rapport à l'espérance de vie globale : serons-nous confrontés à une augmentation de la longévité au profit ou au détriment de l'état de dépendance de l'individu ? D'ailleurs, les mœurs d'un pays peuvent influencer directement sur ces considérations : un pays comme le Japon prône une espérance de vie très élevée et défendra ce précepte quel que soit l'état de dépendance des individus, ce qui contribue à classer aujourd'hui le Japon comme le pays présentant le plus grand nombre de centenaires, et détenteur du record de la « Best Practice Line » depuis quelques années. A contrario, un pays comme le Danemark ne se montre pas nécessairement favorable à l'état de dépendance ni au maintien « à tout prix » dans cet état, ce qui ne le positionne pas comme un pays présentant un record de longévité mais de longévité en bonne santé.

Encore plus aujourd'hui que par le passé, l'Homme est l'un des principaux artisans de ses perspectives de longévité... ! Quoiqu'il en soit, aussi passionnante que soit l'explication par la médecine de l'anticipation de la longévité, cette approche ne fera pas l'objet d'une étude plus approfondie dans ce mémoire.

4- INEGALITES FACE A LA LONGEVITE

Le patrimoine génétique n'expliquerait que 25%⁹ de l'espérance de vie d'un individu. Cette contribution génétique serait presque nulle avant l'âge de 60 ans dans les pays industrialisés, pour ensuite croître après cet âge. Les inégalités face à l'espérance de vie subsistent réellement. Le sexe, l'âge, et la période constituent les facteurs explicatifs majeurs de la mortalité. Cependant, la classe sociale, l'activité professionnelle et notamment le degré de pénibilité du travail, ou encore l'emplacement géographique... sont d'autres facteurs discriminants.

Il est fréquent d'évaluer le SMR ou Standard Mortality Ratio pour comparer deux types de population afin de présenter les inégalités face au taux de mortalité. Le SMR représente en effet le rapport entre le nombre de décès d'une population spécifique et le nombre de décès

⁹ PICARD F. et al. [2013] « Genetic of aging », Médecines Sciences Amérique

qui aurait été constaté sur cette population si la mortalité avait été celle de la population de référence.

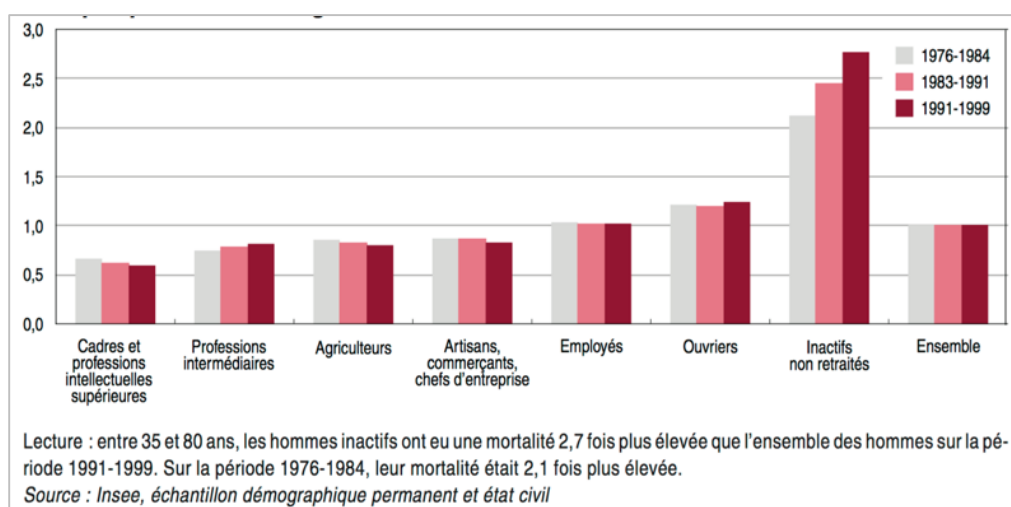
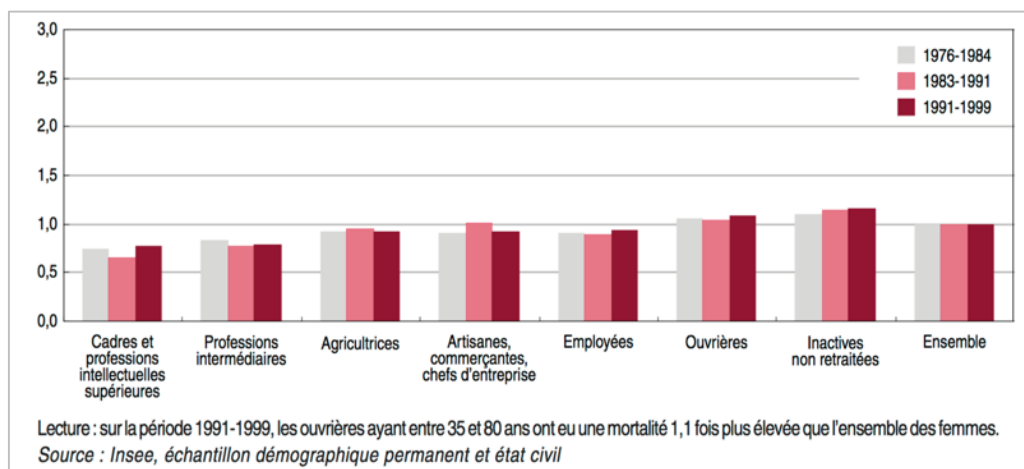


Figure 9 – SMR Femmes (haut) et Hommes (bas) entre 35 et 80 ans par période et par catégorie sociale [2005] INSEE PREMIERE, n° 1025

Bien évidemment, ces inégalités ici présentées au niveau de la France se retrouvent au niveau mondial, selon la qualité d'un pays industrialisé ou d'un pays en voie de développement : chaque pays présente une hétérogénéité qui lui est propre. Par exemple, aux Etats-Unis, la catégorie sociale, compte tenu de l'accès coûteux aux soins, et l'implantation géographique, par le biais des Etats, sont notamment des critères distinctifs très prononcés.

Les inégalités autres que le sexe ne seront pas explicitement abordées dans le cadre de ce mémoire. Néanmoins, pour les assureurs, la problématique de l'anti-sélection est majeure : de l'asymétrie de l'information avec l'assuré au caractère optionnel d'une garantie en passant

par la catégorie socio-professionnelle ou la sélection médicale en souscription d'une garantie.

Le portefeuille qui sera étudié par la suite est constitué du regroupement d'un ensemble de garanties, disposant de conditions de souscription potentiellement différentes et observé sur des périodes distinctes. L'une des problématiques majeures sera d'essayer de gommer au mieux cette hétérogénéité pour constituer un portefeuille agrégé d'assurés sur un historique suffisamment long pour en évaluer la tendance de mortalité. En effet, s'il est reconnu et avéré que la population des assurés présente des taux de mortalité plus faible que la population nationale ($SMR < 1$), qu'en est-il réellement de la tendance de la mortalité ou de la vitesse d'augmentation de l'espérance de vie pour la population des assurés ?

C. MISE EN PERSPECTIVE DES MODELES UTILISES

De manière à comprendre la théorie relative à l'élaboration des tables de mortalité de référence françaises (partie 2), et particulièrement la construction de la tendance du portefeuille agrégé d'assurés étudié (partie 3), il est proposé de recenser les modèles mathématiques sous-jacents utilisés au travers d'un panorama qui s'attache à être aussi pédagogique que possible.

1- MODELES PARAMETRIQUES

Les modèles paramétriques, caractérisés par le fait que la mortalité peut être décrite par une expression mathématique et fonction de paramètres finis identifiés, constituent la genèse des modèles développés de mortalité et offrent la possibilité de la double modélisation (niveau et tendance).

L'un des premiers modèles connus est celui de Gompertz (1825) représentant la force de mortalité en fonction de l'exponentiel de l'âge : $\mu_x(t) = a \cdot e^{b \cdot x}$, avec a et b constantes. L'expression du taux de mortalité qui peut se déduire de celle de la force de mortalité s'exprime également selon une exponentielle en fonction de l'âge (Figure 3). Cependant, de nombreux modèles de mortalité se sont par la suite développés en ne limitant pas le paramétrage à un **unique facteur** : l'âge.

Le modèle de Lee-Carter (1992) est le premier modèle reconnu prenant en compte **deux facteurs** : l'âge et l'année d'observation. Son principe est le suivant : $\ln(\mu_x(t)) = \alpha_x + \beta_x \cdot k_t + \varepsilon_{x,t}$, avec les variables $\varepsilon_{x,t}$ indépendantes et identiquement distribuées selon une loi Normale $\mathcal{N}(0, \sigma^2)$. Le paramètre α_x , s'interprète comme la valeur moyenne des $\ln(\mu_x(t))$ dans le temps, β_x traduit la sensibilité de la mortalité instantanée à l'âge x par rapport à l'évolution générale temporelle de k_t . Il est à noter que cette sensibilité est fixée comme étant constante au cours du temps, ce qui n'est pas nécessairement justifié sur la base des données de populations nationales et mondiales.

Pour évaluer la tendance linéaire d'évolution des taux de mortalité, k_t , les « **séries chronologiques** » (Box et al., [1994]) sont employées. La modélisation ARIMA s'avère simple à élaborer dans un modèle de type Lee-Carter puisqu'elle ne fait intervenir qu'un paramètre (le temps), mais cette même modélisation particulièrement riche permet également d'évaluer les huit paramètres d'un modèle plus sophistiqué à deux facteurs de type Heligman-Pollard. Plus généralement, l'évaluation de l'ensemble des paramètres Lee-Carter (âge et temps)

s'effectue par la méthode des moindres carrés, et la mise en place de contraintes d'identifiabilité en guise d'unicité de solution (voir BROUHNS N. et al. [2001] pour le détail de la mise en œuvre).

Bien que ce modèle constitue, encore aujourd'hui, un réel benchmark, le fait qu'il considère l'hypothèse d'homoscédasticité des taux de mortalité (non justifié notamment aux grands âges où le manque de données induit une augmentation de la variance) ou qu'il ne s'appuie sur aucune justification probabiliste dans le choix des critères optimaux (PLANCHET F., LELIEUR V., [2010]), a conduit BROUHNS N. et al. [2002] à développer le modèle **Log-Poisson**. Ce modèle permet de supprimer les controverses liées aux propriétés des estimateurs en s'appuyant sur un principe de type « maximum de vraisemblance ». Entrer dans des considérations de « maximum de vraisemblance » suppose l'identification d'une loi de probabilité. Il est ainsi estimé au préalable que la distribution des décès se comporte selon une loi de Poisson fonction de l'exposition, $E_{x,t}$, et de la force de mortalité, $\mu_x(t)$: $D_{x,t} \sim \mathcal{P}(E_{x,t} \cdot \mu_x(t))$. (Propriété qui découle de l'hypothèse de constance par morceaux de la force de mortalité évoquée précédemment et qui sera détaillée en partie 2). La forme retenue pour la force de mortalité est ensuite identique à celle du modèle de Lee-Carter ($\ln(\mu_x(t)) = \alpha_x + \beta_x \cdot k_t$). L'évaluation des différents paramètres est dans ce cas réalisée grâce à l'expression obtenue des logarithmes de vraisemblance.

Le modèle Log-Poisson, à l'image de celui de Lee-Carter considère que la sensibilité à la composante temporelle n'est dépendante que de l'âge. Des alternatives et généralisations ont ainsi pu faire leur apparition comme le modèle logistique glissant de Bongaarts (2004) : testé sur un échantillon de 14 pays industrialisés, il s'avère représenter assez justement la mortalité passée.

Plus récemment, des modèles à **trois facteurs** ou APC (Age, Période Cohorte) se sont développés ; l'idée ayant pourtant émanée dès 1965 avec Norman B. Ryder. Une forme générale se retrouve dans le modèle CBD (Cairns, Dowd and Blake) dont un modèle spécifique a ensuite été dérivé par CAIRNS A.J.G. et al. [2008]. RENSHAW A.E. et HABERMAN S. [2006] ou CURRIE I.D. et al. [2006] peuvent également être cités en application d'un modèle de type APC.

Le modèle Log-Poisson fera par la suite l'objet d'un développement et d'une application spécifique. En effet, ce modèle semi-paramétrique¹⁰ présente l'avantage de pouvoir élaborer

¹⁰ *Semi-paramétrique, dans la mesure où le caractère paramétrique vient uniquement de l'interaction âge/année*

une extrapolation **endogène** de la tendance, à partir de la reproduction des données propres de la population étudiée.

2- MODELES SEMI-PARAMETRIQUES : EXEMPLES DES MODELES RELATIONNELS

Extrapoler une évolution de mortalité sur la base d'un modèle paramétrique et par conséquent 100% endogène peut conduire à de fortes irrégularités liées à des fluctuations d'échantillonnage dans un contexte où la quantité de données serait sensiblement faible en regard de la quantité de paramètres à évaluer (LELIEUR V. [2005]). Les modèles paramétriques ont pu être testés et éprouvés sur de larges bases de données comme celles des populations nationales ou mondiales notamment disposant d'un historique suffisamment long. Si cet historique peut remonter jusqu'à 2 siècles¹¹, celui d'une population d'assurés ne dépasse généralement pas les 5 à 15 ans selon les organismes assureurs. La France, contrairement au Royaume-Uni et sa base de données de rentiers, n'a pas mis en place de véritable base d'exploitation des données de mortalité de la population des assurés. En revanche, certains assureurs français contribuent à des études de mortalité de place, menées notamment par l'Institut des Actuaires avec le groupe de travail sur la mortalité, en transmettant les portefeuilles de données de mortalité de leurs assurés. C'est d'ailleurs un groupement de portefeuilles de ce type qui a permis l'étude présentée dans ce mémoire.

Le manque de données de mortalité de place ou pour chaque acteur est la raison pour laquelle depuis toujours, actuaires et démographes utilisent des données de mortalité **exogènes** ou de référence pour améliorer l'estimation de la mortalité de leur population d'intérêt : les modèles semi-paramétriques, parmi lesquels les modèles relationnels, sont également plus souples et permettent une modélisation plus fine de la mortalité.

iv. Modèles relationnels classiques

Le plus connu et le plus simple des modèles relationnels est le modèle de **Cox** (1972). Il s'agit d'un modèle à risques proportionnels entre la population étudiée et une population de référence « *ref* » considérée : le coefficient de proportionnalité θ est considéré comme indépendant de l'âge. Le modèle s'exprime en fonction des forces de mortalité de la manière suivante : $\mu_x = \theta \cdot \mu_x^{ref}$, pour tout x. Ce modèle relationnel particulièrement simple a vu sa

¹¹ Bases de données exploitables notamment via la Human Mortality Database.

validité rapidement mise en doute par de nombreux auteurs. Cependant, il trouve une utilisation pertinente dans un cadre d'hétérogénéité de portefeuille (PLANCHET F. et TOMAS J., [2013]), afin d'harmoniser un portefeuille de données agrégées à partir de différentes sources et sur des périodes d'observation distinctes. Cette application sera d'ailleurs structurante dans le cadre de ce mémoire.

Le modèle relationnel de **Brass** (1971) établit quant à lui une relation entre logit¹² des quotients de mortalité : $logit(q_x) = \theta_1 + \theta_2 \cdot logit(q_x^{ref})$, pour tout x . Au-delà du fait que l'inverse de la fonction logit est explicite, l'avantage de recourir à cette fonction vient de sa bijectivité de $]0 ; 1[$ dans \mathbb{R} , qui permet d'opérer un ajustement plus fiable et robuste d'un nuage de points non plus dans $]0 ; 1[$ mais sur $]-\infty ; +\infty[$. Nous exploiterons par la suite ce modèle pour précisément comparer la tendance extrapolée à partir d'un modèle exogène avec celle obtenue de manière endogène.

Le modèle de Hannerz (2001) complète le tableau des modèles relationnels classiques.

v. Autres modèles relationnels – cadre du GLM

Les modèles relationnels sont également abordés par le biais du cadre des modèles linéaires et additifs généralisés (GLM). Le principe d'application est à nouveau basé sur la constance par morceaux des forces de mortalité et la condition intrinsèque de représentativité des décès par une loi de Poisson : $D_{x,t} \sim \mathcal{P}(E_{x,t} + \mu_x(t))$.

Une application directe au cadre des GLM consiste à supposer une dépendance linéaire à la force de mortalité et d'estimer ainsi les paramètres sur la base des données disponibles. Selon la typologie de la dépendance linéaire considérée, intrinsèquement liée à la quantité et l'historique de données à disposition, la tendance extrapolée pourra être considérée comme :

- Exogène : par exemple un modèle GLM de Poisson qui fixerait une dépendance linéaire selon l'âge et une mortalité de référence
- Endogène : ce même modèle GLM de Poisson établissant en sus une dépendance avec l'année calendaire

Un tel cas endogène fera l'objet d'une étude approfondie dans la suite de ce mémoire.

¹² La fonction logistique ou logit s'écrit : $logit(x) = \ln\left(\frac{x}{1-x}\right)$

3- MODELES NON PARAMETRIQUES

La considération de la distribution de Poisson des décès offre aussi une ouverture au monde du non paramétrique via notamment le lissage par vraisemblance locale. Ces modèles non paramétriques, appliqués par DELWARDE et al. [2004] ou plus récemment par GSCHLOSSL et al. [2011] ont également été utilisés dans le cadre d'un lissage multidimensionnel par TOMAS J. et PLANCHET F. ([2011] et [2013]).

Les modèles non paramétriques s'inscrivent dans un cadre particulièrement flexible au plus éloigné de l'analytique. Mais ils en perdent la qualité prédictive et ne permettent donc que des extrapolations limitées.

Ce type de modèle ne sera donc pas exploité dans le cadre de ce mémoire. Mais notons que les tables de mortalité « *best estimate* » élaborées dans le cadre du groupe de travail sur la mortalité de l'Institut des Actuaires ont fait appel à un lissage fin par vraisemblance locale avec application exogène des améliorations, et notamment de la tendance d'une table INSEE prospective.

CONCLUSION PARTIE 1

L'objectif de cette première partie était de disposer de l'ensemble des notions nécessaires à l'étude que nous nous proposons de mener sur la tendance de la mortalité propre d'un portefeuille agrégé d'assurés.

L'extrapolation de cette tendance ne peut être élaborée qu'à partir d'un historique suffisamment profond, ce qui est particulièrement rare dans le monde de l'assurance. De plus, au delà du fait qu'il existe un biais dans l'estimation de la tendance d'un échantillon, aucun élément factuel ne peut, en soit, attester de la justesse d'une tendance prospective par rapport à une autre. C'est la raison pour laquelle les tables de mortalités assurantielles de référence s'appuient toutes par défaut sur la dérive de la population nationale.

La seconde partie explicitera et challengera l'omniprésence de cette tendance nationale dans le monde de l'assurance, grâce à l'opportunité donnée dans le cadre de ce mémoire de disposer d'une base agrégée d'informations de mortalité de place disponible sur une large période d'observation.

PARTIE 2 : TABLES PROSPECTIVES ASSURANTIELLES ET OMNIPRESENCE DE LA TENDANCE DE LA POPULATION NATIONALE

Maîtriser le risque de longévité passe par une « bonne » estimation de la dynamique de la mortalité de demain. Cette anticipation se matérialise par des tables de mortalité prospectives présentant les taux de mortalité par âge et années calendaires futures (tables transversales) ou par âge et générations (tables longitudinales). La méthode employée pour élaborer une telle table peut s'appuyer sur un avis d'expert (projection expectative), ou sur une analyse de causes (projection explicative), ou encore sur des modèles tels qu'explicités dans la liste non exhaustive présentée en première partie (projection par extrapolation). Les tables prospectives hommes et femmes de la population nationale française, élaborées par l'INSEE¹³, couplent une projection expectative et explicative.

Plusieurs tables prospectives réglementaires ont été élaborées dans le monde de l'assurance pour imposer un cadre minimum aux assureurs dans la tarification et le provisionnement de leurs différentes garanties. Pour chacune de ces tables, le positionnement reflète le niveau inférieur de mortalité reconnu pour la population des assurés par rapport à la population nationale mais la tendance reste extrapolée sur celle de la population nationale, notamment du fait d'un manque de données de place.

Ces tables réglementaires introduisent une marge de prudence qu'il n'est plus nécessaire de considérer dans un calcul « au plus juste » des provisions *best estimate* sous Solvabilité 2. En effet, le choc de longévité de la formule standard permet d'évaluer le besoin en fonds propres pour faire face aux risques de longévité sous-jacents et la marge pour risque procure la vision prudentielle. Un autre type de table de référence a ainsi récemment fait son apparition : les tables *best estimate*¹⁴, déclinées pour les hommes et les femmes. Si le niveau de ces tables est légèrement ajusté par rapport aux tables réglementaires, leur tendance a également été extrapolée à partir de celle de la population nationale.

L'omniprésence de la tendance de la population nationale pour la population des assurés est donc bien avérée. Mais, pourrait-elle être remise en question ?

¹³ BLANPIN N., CHARDON O. [2010] « Projections de population 2007-2060 pour la France métropolitaine : méthode et principaux résultats », DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES, - N°F1008

¹⁴ Créées en 2013 dans le cadre du groupe de travail sur la mortalité de l'Institut des Actuaire

Une première réponse sera apportée dans cette partie avec la mise en application d'un ajustement par le modèle relationnel de Brass sur un portefeuille d'assurés agrégé, rendu homogène par application de la régression de Cox, observé sur une large période de 16 ans (1996-2011). Mais, commençons par un état des lieux plus détaillé des tables prospectives de référence.

A. TABLES PROSPECTIVES ET REFERENCES

1- LES TABLES DE REFERENCE DE LA POPULATION NATIONALE

En 2011, l'INSEE (BLANPIN N. et CHARDON O. [2010]) a proposé une mise à jour des travaux de 2006 de projection démographique nationale de la population française basée sur des hypothèses de fertilité, de mortalité et de migration. Un ensemble de 27 scénarii a été établi, comprenant une vision centrale et des variantes utiles à l'analyse des effets d'un changement d'hypothèse sur les résultats projetés. Le scénario central prend en compte une fertilité assez élevée (1,95 enfants par femme sur la durée de projection contre 1,9 lors des projections de 2006), ainsi qu'une hypothèse de mortalité plus optimiste pour les plus âgés toujours par rapport à 2006 (hommes de plus de 85 ans et femmes de plus de 91 ans). Quant à la migration, elle retient un solde migratoire de 100 000 personnes par an, sans changement par rapport à 2006.

Pour la mortalité, les hypothèses haute, centrale et basse sont le fruit de l'échange de 18 experts (dont la plupart expriment des positions personnelles qui n'engagent pas l'organisme qui les emploie, sauf pour la Drees¹⁵ ou le Pôle de Service de l'Action Régionale, PSAR, Emploi-Population de l'INSEE) et s'illustrent graphiquement à l'aide des espérances de vie à la naissance. L'hypothèse centrale estime une espérance de vie à la naissance en 2060 de 91,1 ans pour les femmes et 86 ans pour les hommes.

¹⁵ Drees – Direction de la recherche, des études, de l'évaluation et des statistiques

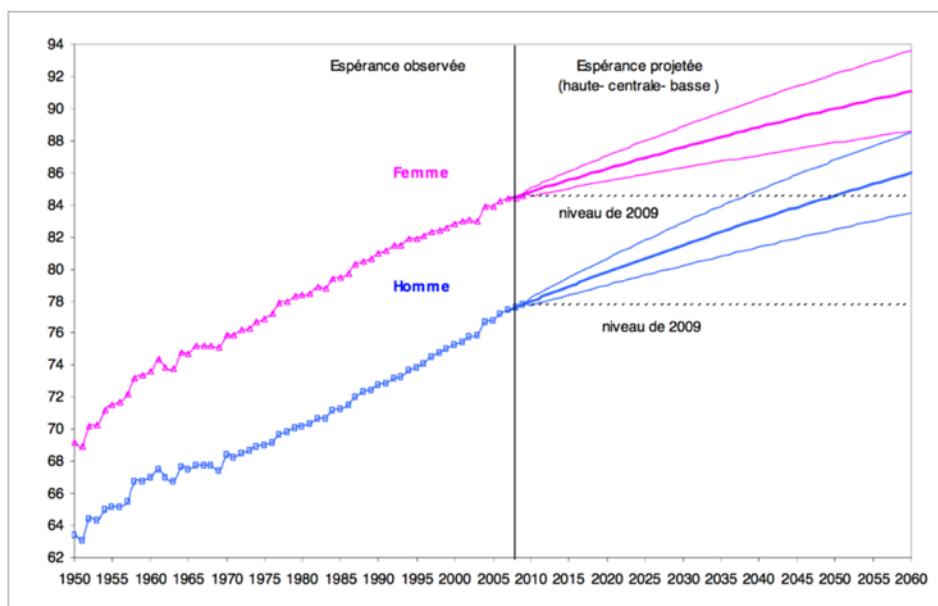
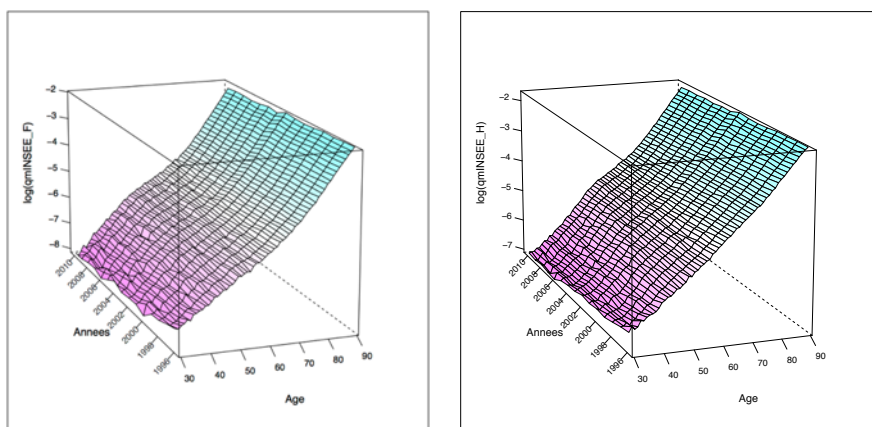


Figure 10 - Espérance de vie à la naissance observée et projetée de la population française
 Source - Projections de population 2007-2060 pour la France métropolitaine : méthode et principaux résultats

i. Tables INSEE brutes

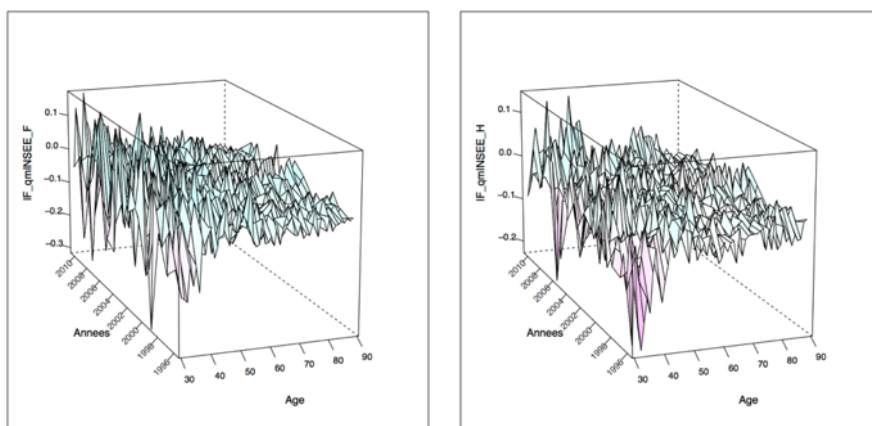
Une table de mortalité brute pour chaque sexe découle de ces projections (qmINSEE hommes et femmes¹⁶). Le caractère brut de ces tables se matérialise par une surface de mortalité irrégulière sur la période 1995-2011 mais surtout par des facteurs d'améliorations des taux de mortalité à âge fixé : $IF(x, t) = \frac{q_x(t+1)}{q_x(t)}$, notés IF¹⁷, particulièrement accidentés.



Nappe 1 – Logarithmes de mortalité Femmes / Hommes – Tables INSEE brutes

¹⁶ Source - <http://www.ressources-actuarielles.net/gtmortalite> pour les tables qmINSEE_F et qmINSEE_H

¹⁷ IF pour Improvement Factor

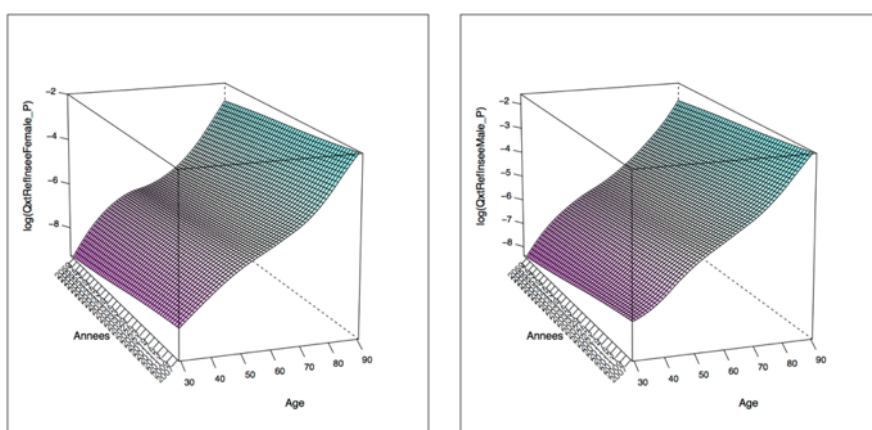


Surface 1 – Facteurs d'amélioration des taux de mortalité Femmes / Hommes – tables INSEE brutes

Ces tables seront exploitées comme tables de référence pour ajuster les taux bruts du portefeuille agrégé d'assurés avec le modèle de Brass (Partie 2) et le modèle de type GLM Poisson (Partie 3).

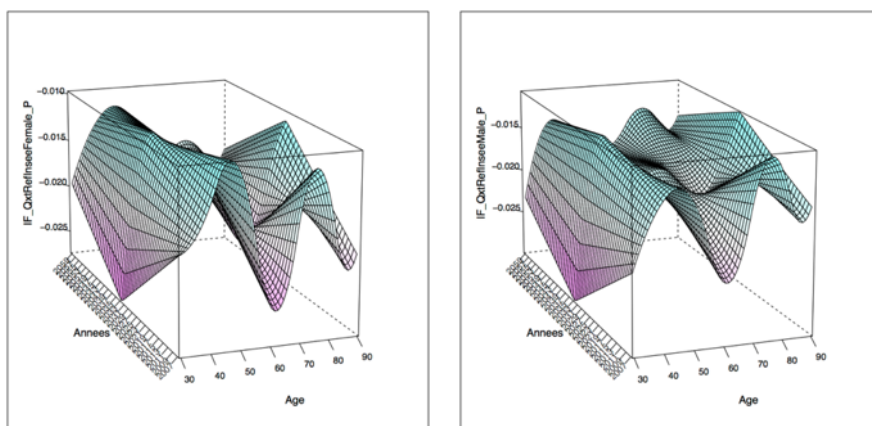
ii. Tables INSEE lissées et ajustées

L'INSEE fournit également des tables équivalentes aux tables brutes, mais lissées sur cette même période 1995-2011. Malgré ce lissage, il subsiste de fortes variations aux âges élevés, ce qui a conduit PLANCHET F. et TOMAS J. [2013]¹⁸ à pratiquer un ajustement complémentaire selon un modèle non paramétrique de p-splines (ou penalised-splines).



Nappe 2 – Logarithmes de mortalité Femmes / Hommes – Tables INSEE lissées et ajustées

¹⁸ Document I11291-11 v1.4, « Construction et validation des références de mortalité de place »



Surface 2 – Facteurs d'amélioration des taux de mortalité Femmes / Hommes – Tables INSEE lissées et ajustées

La représentation graphique prospective de ces tables ajustées¹⁹ illustre cette régularité et ce lissage.

Notons que les tables prospectives de l'INSEE sont également utilisées par le COR (Conseil d'Orientation des Retraites) dans la projection des flux de trésorerie futurs potentiels des régimes AGIRC et ARRCO pour évaluer le degré de couverture : ces mêmes études ont conduit à d'importantes réformes des retraites.

2- LES TABLES DE REFERENCES ASSURANTIELLES

D'un point de vue assurantiel, les tables prospectives sont primordiales pour tarifer ou provisionner une garantie en lien à la longévité ou à la mortalité d'un assuré.

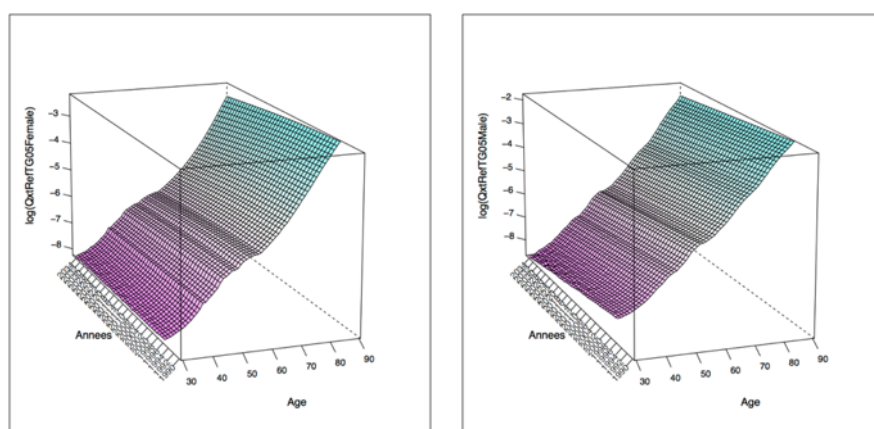
i. Les tables réglementaires

Les tables requises pour la tarification et le provisionnement sont encadrées par la réglementation. L'article A-335-1 du Code des Assurances en précise le cadre : les tables TH et TF00-02 pour les assurances en cas de décès, ces mêmes tables utilisées avec un décalage d'âge pour les assurances en cas de vie, à l'exclusion des rentes qui s'appuient sur les tables TGH et TGF05. La notation H (respectivement F) désigne les hypothèses relatives aux hommes (respectivement aux femmes).

¹⁹ Source <http://www.ressources-actuarielles.net/gtmortalite> pour les tables femmes *QxtRefInseeFemale* et hommes *QxtRefInseeMale*

Les tables TH et TF00-02 sont des tables à un seul facteur (l'âge) représentées par une unique colonne, pour lesquelles un mécanisme de décalage d'âge permet de considérer un second facteur (le temps). Ces tables ont été établies à partir de données de l'INSEE issues des observations réalisées entre 2000 et 2002.

Les tables TGH et TGF05 sont des tables générationnelles prospectives (1900-2005), élaborées pour le marché français de l'assurance et fournies par l'Institut des Actuaires (Planchet, 2006). Celles-ci ont été obtenues à partir de la mortalité de la population des bénéficiaires de contrats de rentes observée sur la période 1993-2005 et de données sur la population générale (INSEE) de 1962 à 2000. Une représentation graphique des taux de mortalité transversaux²⁰, sur la période 2007-2035 et pour les âges 30 à 90 ans, est proposée ci-dessous.



Nappe 3 – Logarithmes des taux de mortalité – Femmes / Hommes – Tables TG05

Ainsi, si nous devons faire une rétrospective récapitulative des tables réglementaires : jusqu'en 1993, les tables homologuées étaient instantanées et basées sur les observations de la population nationale. Les premières tables de générations en France (les TPG93) sont prospectives et ont été établies à partir d'une extrapolation des taux de mortalité de la population féminine fournie par l'INSEE. Les tables TGH/F05 les ont remplacées à partir d'août 2006.

L'important à retenir reste que ces tables réglementaires ont toutes été élaborées à l'appui de la tendance de la population générale (INSEE).

²⁰ Grâce à la relation: $\mu(x, t) = \mu_{\text{gen}}(x, t - x)$, avec μ_{gen} , la force de mortalité longitudinale ou générationnelle

Notons que la succession des tables de mortalité réglementaires²¹ depuis les années 80 traduit notamment une sous-estimation répétée de cette tendance...

ii. Les tables d'expérience et principe de construction

Le même article A-335-1 du Code des Assurances autorise les assureurs à faire certifier une table de mortalité d'expérience qui s'avèrerait alors ajustée à un portefeuille spécifique (population, contrat, risque, garantie...) : ce type de table, dont la validité est de 5 ans, doit permettre la « constitution de provisions suffisantes et prudentes ».

Nous ne nous intéresserons pas ici au processus de certification et de suivi, mais à la méthode de construction qui distingue deux étapes majeures :

1. **L'estimation des taux bruts** par âge et période ;
2. Puis **l'ajustement et l'extrapolation** de cette table sur la base d'un modèle de mortalité.

En pratique, l'élaboration d'une table d'expérience dépend principalement de la quantité et de la fiabilité des données à disposition. Usuellement, le manque de données conduit à utiliser la tendance d'une référence pour rendre la table d'expérience prospective. Dans la suite, l'objectif sera bien d'extrapoler la tendance propre du portefeuille agrégé d'assurés étudié.

Un dernier sujet important dans la construction d'une table d'expérience est celui de la fermeture de la table aux âges élevés. En effet, les données, aussi étendues soient-elles, ne sont généralement pas suffisantes et fiables pour les grands âges (supérieurs à 90 ans par exemple). Deux types d'extrapolation sont alors usuels pour compléter la table à ces âges : une extrapolation linéaire sur la base des taux de mortalités ajustés aux âges 80 à 90 ans avec une contrainte imposée à 120 ou 130 ans d'avoir un taux de mortalité égal à 1, ou une extrapolation en logarithme quadratique avec une condition équivalente au taux de mortalité unitaire à 130 ans et une tangente nulle de la mortalité à ce même âge, développée par DENUIT M. et GODERNIAUX A.C. [2005].

Les tables qui seront par la suite construites ne seront pas « fermées ». La comparaison des résultats aux tables de référence sera faite à iso périmètre, à savoir sur la tranche d'âge de 30 à 90 ans. Notons simplement que les résultats qui seront présentés seront non comparables

²¹ La succession des tables réglementaires de référence sont répertoriées sur le site de l'Institut des Actuaire.

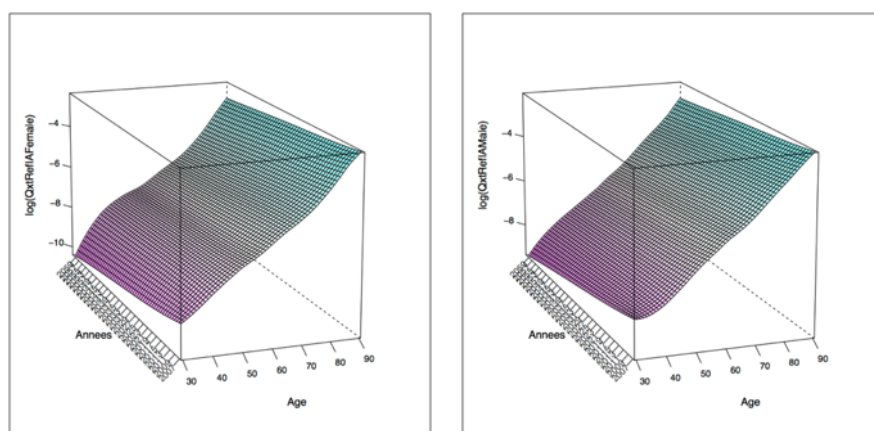
(inférieurs en l'occurrence pour les espérances de vie par exemple) aux ordres de grandeur évoqués en première partie.

iii. Les tables best estimate 2013

Le calcul des provisions *best estimate* imposé dans le cadre de la Directive Solvabilité 2 doit être réalisé au plus juste du risque sous-jacent. En effet, les chocs de la formule standard et notamment le choc de longévité est sensé caractériser la marge de prudence adaptée aux risques sous-jacents à la longévité pour l'évaluation des fonds propres.

Dans cette perspective, le groupe de travail sur la mortalité de l'Institut des Actuaire a publié en 2013 deux tables prospectives de référence, notées *QxtRefIAFemale* et *QxtRefIAMale*, complétées d'un panel d'outils permettant à tout assureur de construire sa propre table *best estimate*, la plus ajustée à son propre portefeuille, à la disponibilité de données propres près.

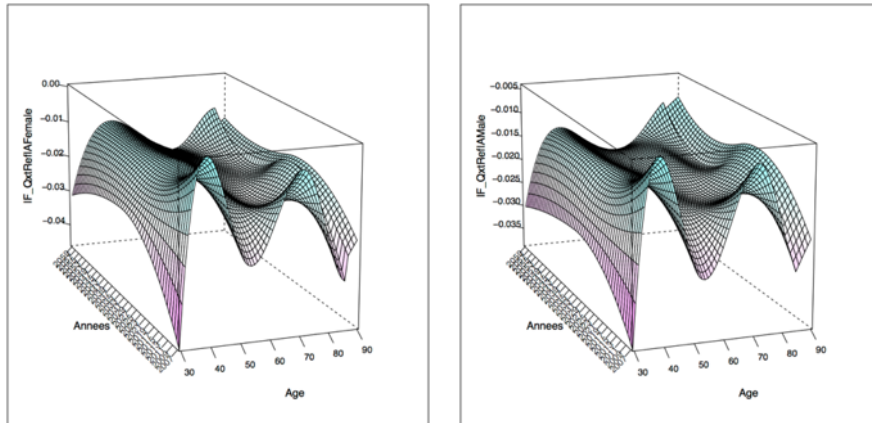
Ces tables *best estimate* ont été construites sur la base d'un ensemble de données de décès de place transmises par près de 15 organismes, au titre de différentes garanties (vie, emprunteur, retraite...). Un lissage non paramétrique par vraisemblance locale a été employé sur une période d'observation commune à l'ensemble des données, à savoir entre 2007 et 2011. Une extrapolation de la tendance a ensuite été faite en référence aux tables INSEE de la population nationale.



Nappe 4 – Logarithmes des taux de mortalité – Femmes / Hommes – Tables IA

Les nappes de mortalité justifient d'un niveau de mortalité inférieur à celui de la population nationale et les nappes de facteurs d'amélioration reflètent une forme équivalente à celle des tables INSEE.

Notons que les données brutes ici utilisées et détaillées par la suite sont précisément celles ayant servies à la construction de ces tables *best estimate*. La comparaison des résultats obtenus à ces tables n'en sera que d'autant plus pertinente.



Surface 3 – Facteurs d'amélioration des taux de mortalité Femmes / Hommes – Tables IA

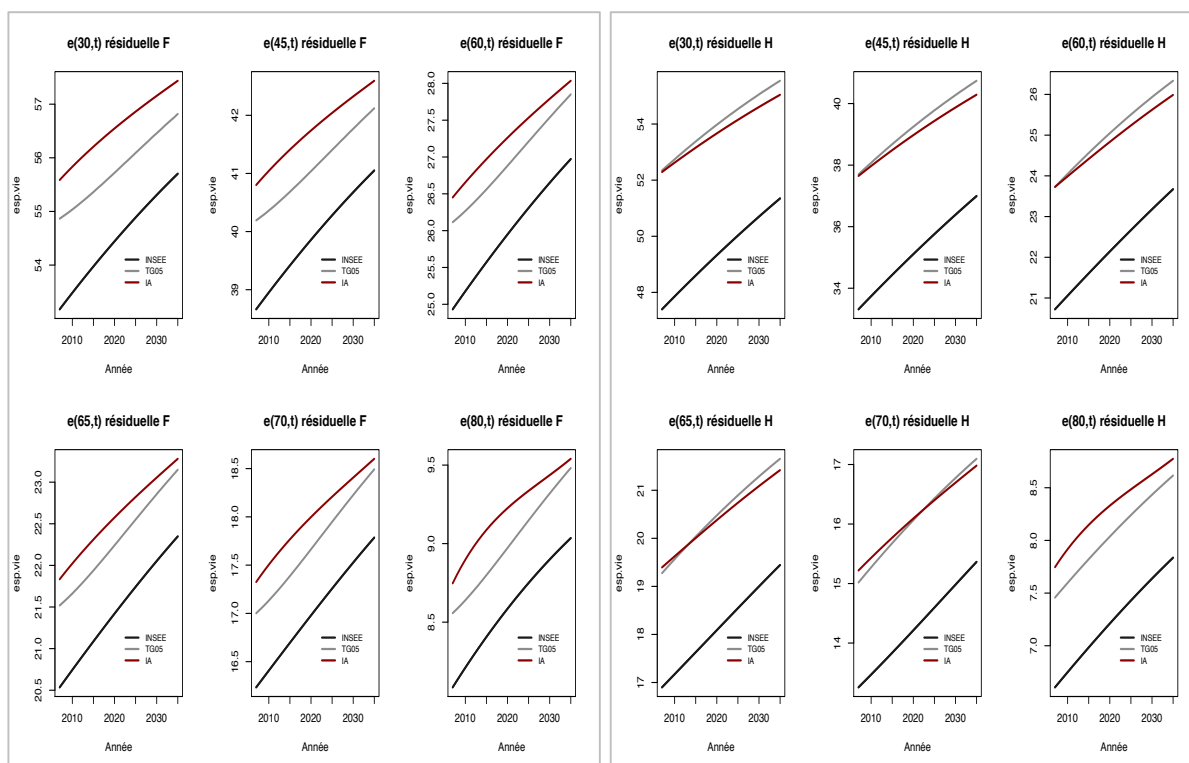
3- COMPARAISON DES TABLES ASSURANTIELLES ET NATIONALES

Nous proposons de conclure cet état des lieux par la représentation graphique de trois indicateurs pertinents pour la comparaison de la tendance :

- L'espérance de vie résiduelle à différents âges en fonction de l'année d'observation ;
- La moyenne de la vitesse d'espérance de vie à chaque âge, évaluée par moyenne sur la période d'observation considérée, ici 2007 à 2035, de la dérivée de l'espérance de vie résiduelle à chaque âge ;
- Le facteur d'amélioration, ou $IF(x, t)$, précédemment défini.

Ces mêmes indicateurs seront utilisés pour analyser les résultats obtenus. Pour être cohérent, la tranche d'âge 30 à 90 ans a donc dès à présent été considérée.

i. Espérance de vie résiduelle par âge



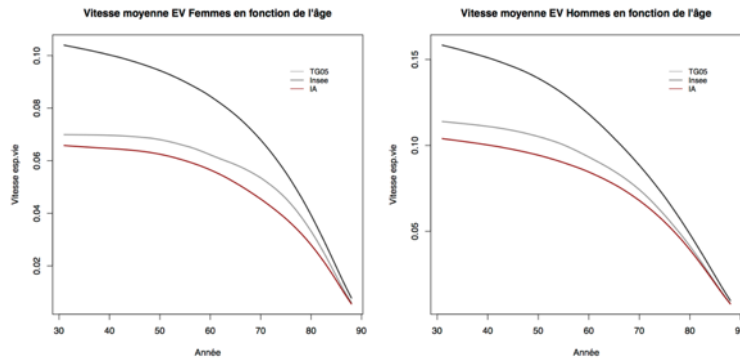
Graphique 1 – Espérance de vie résiduelle - Femmes / Hommes

Ces courbes illustrent l'ajustement flexible obtenu avec un lissage par vraisemblance locale pour les tables *best estimate* (ou tables IA) pour lesquelles la forme de l'espérance de vie résiduelle se montre très variable en fonction de l'âge et de l'année d'observation, contrairement aux tables INSEE ou TG05.

ii. Vitesse annuelle de l'espérance de vie résiduelle par âge

La vitesse annuelle d'espérance de vie à un âge donné se lit comme la moyenne sur une période d'observation du coefficient directeur de la tangente à la courbe d'espérance de vie résiduelle à ce même âge. Mais elle peut également s'exprimer comme un gain d'espérance de vie à la naissance de 3 mois par an constaté à ce jour, soit 0,23. Cette même vitesse est estimée à 0,1 à l'âge de 60 ans, soit un gain d'environ 1 mois par an de l'espérance de vie résiduelle à 60 ans.

Moyennées sur la période d'observation 2007 à 2035, les vitesses relatives aux tables TG05 et IA peuvent être considérées comme quasi équivalentes. Celles des tables INSEE apparaissent en revanche encore supérieures.

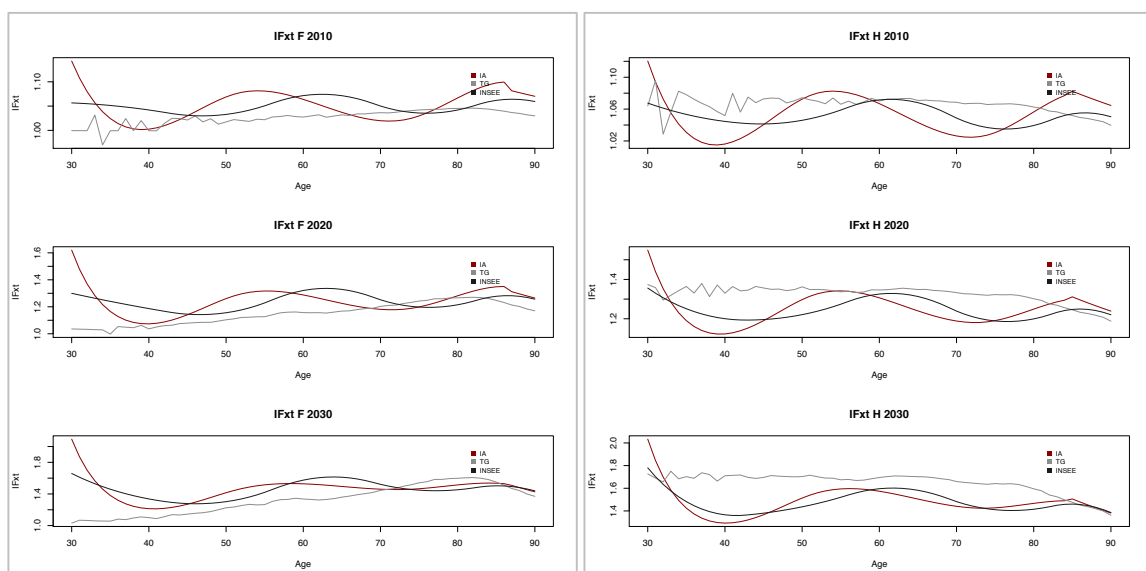


Graphique 2 – Vitesse moyenne d'espérance de vie Femmes et Hommes

Ce résultat vient de la construction mathématique de la tendance des tables TG05 et IA, pour lesquelles il a été supposé que la mortalité des assurés tendrait à terme vers la mortalité de la population nationale. Les taux de survie étant supérieurs en niveau pour les assurés, leur évolution se doit d'intégrer une pente et donc une vitesse légèrement inférieure à celle de la population nationale.

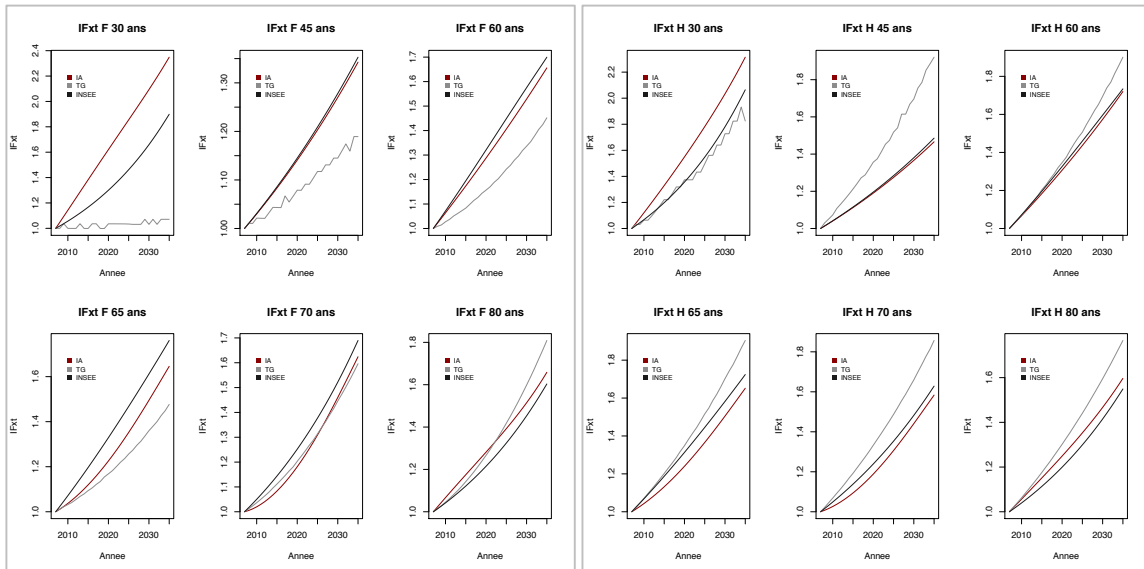
iii. Facteurs d'amélioration des taux de mortalité

L'analyse des facteurs d'amélioration à année d'observation fixée met en évidence la prudence (en regard de la longévité) pour les femmes et jusqu'à 70 ans de la table TG05. Pour les hommes, il s'agit alternativement de la table best estimate ou de la table INSEE qui présente des facteurs d'amélioration plus faibles.



Graphique 3 - Facteurs d'amélioration en fonction de l'âge et par années d'observation – Femmes/Hommes

Aux âges élevés, nous pouvons remarquer une convergence des facteurs d'amélioration des différentes tables avec l'année d'observation, ce qui rejoint la remarque précédemment faite.



Graphique 4 - Facteurs d'amélioration en fonction de l'année d'observation et différents âges – Femmes/Hommes

L'analyse de ces mêmes facteurs d'amélioration par année d'observation à âges fixés illustre les similitudes entre les tables INSEE et les tables *best estimate* à partir de 45 ans. La table TG05 présente des divergences certaines.

Notons que les facteurs d'amélioration permettent une analyse et une comparaison plus fine et précise de la tendance de la longévité par âge et année d'observation. Cependant, en première approche, nous privilégierons l'exploitation de la vitesse d'espérance de vie qui offre une vision plus immédiate de la tendance générale.

4- ENJEU DES ASSUREURS EN REGARD DU RISQUE DE LONGEVITE

La succession des tables réglementaires surévaluant la mortalité prospective illustre un premier risque de longévité important pour les assureurs. Même si l'assureur dispose d'un délai de 15 ans pour lisser le provisionnement de ses rentes en portefeuille²², il n'en reste pas moins que le changement de table induit un besoin de provisionnement supplémentaire qui peut être conséquent : ce coût a été en moyenne évalué à 8% (BARRIEU P. et al. [2010]) des provisions pour le passage de la table TPG 93 aux tables TG05.

²² Article A331-1-2 du Code des Assurances

Plus précisément, les risques sous-jacents à l'élaboration d'une table sont nombreux :

- Le risque d'échantillonnage
- Le risque opérationnel
- Les risques systématiques que l'on retrouve à travers :
 - Le risque de modèle
 - Le risque de positionnement
 - Le risque de tendance

C'est pour cette raison que par la suite, les ajustements de mortalité qui seront opérés à l'aide des différents modèles seront rattachés à des intervalles de confiance, reflétant une estimation de l'incertitude sous-jacente au risque d'échantillonnage dans l'évaluation d'une table de mortalité.

Le risque opérationnel ne sera pas étudié en détail mais sera tout de même pris en compte par le biais de l'analyse critique des données brutes utilisées, en amont de l'application des modèles mais également lors de l'exploitation des résultats obtenus.

Les risques systématiques feront effectivement partie de l'analyse, dans la mesure où l'étude mettra en perspective les résultats obtenus par le biais de deux modèles endogènes distincts, et qu'une étude de plusieurs sensibilités sera menée.

Par ailleurs, l'assureur dispose de différents moyens pour couvrir son risque de longévité, comme notamment celui de transférer ce risque aux réassureurs ou aux marchés financiers, par le biais de swap de longévité ou, moins avéré, par la titrisation. Le principe de couverture et ses différentes perspectives ne seront pas développés dans le cadre de ce mémoire.

B. CONSTRUCTION DE LA TABLE DES MOMENTS DU « PORTEFEUILLE DES ASSURES »

1- LES DONNEES

i. Données initiales - généralités

L'étude ici menée n'aurait pas été possible sans un portefeuille suffisamment large de données d'assurés : les informations de décès transmises par 15 organismes dans le cadre de la construction des tables *best estimate* ont ainsi offert une belle opportunité.

En effet, en sus de proposer une période temporelle de plus de 15 ans exploitable, le regroupement de l'ensemble de ces portefeuilles justifie d'une certaine exhaustivité des données de place :

- Les trois fédérations (Sociétés d'Assurance, Institutions de Prévoyance, Mutuelles) sont représentées ;
- Les organismes sont de tailles variées ;
- Les garanties peuvent être tout autant de type retraite, vie entière, emprunteur que décès ;
- Les contrats sont individuels, collectifs à adhésion obligatoire ou collectifs à adhésion facultative ;
- Ces contrats présentent pour certains un critère de sélection médicale et pour d'autres non.

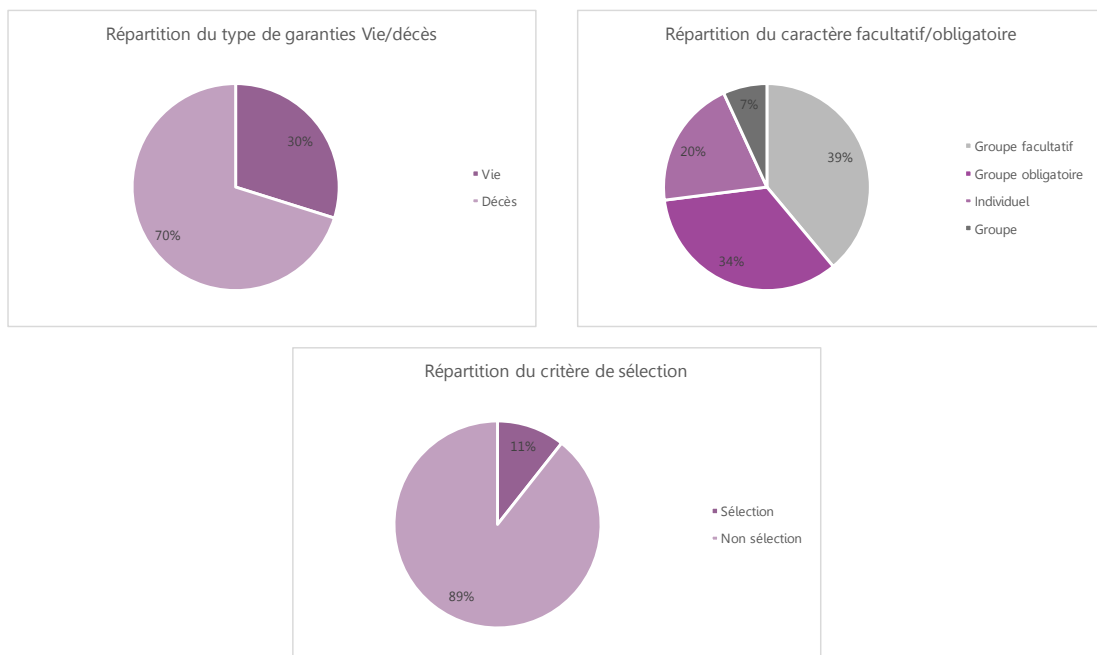


Figure 11 – Répartition des données brutes selon différents critères

Cette base de données ne saurait être considérée comme une parfaite représentation des « assurés » de place mais permet une première analyse de la tendance. Il est à noter que l'échantillon de données utilisé pourrait induire une surestimation des décès par rapport à « la » population des assurés, dans la mesure où une majorité de garanties est relative aux décès, en non sélection, et au sein de contrats principalement facultatifs. Ce biais, considéré comme minime, sera néanmoins pris en compte dans l'analyse des résultats.

ii. Données initiales – ligne à ligne

Plus précisément, le groupe de travail « mortalité » de l'Institut des Actuaire a donné accès, dans le cadre de ce mémoire, à version anonymisée (un numéro de référence par organisme) de la base de données²³ utilisée en 2013 lors de la construction des tables *best estimate*. En l'état, cette base répertorie plus de 8 millions de lignes s'étalant de 1952 à 2011. Une ligne représente un assuré notamment par sa date de naissance, son sexe, sa date d'entrée dans l'organisme, sa date de sortie et son statut correspondant à une indicatrice de la survenance du décès ou non de l'individu durant sa présence dans la base (1 si le sinistre a eu lieu, 0 sinon).

Sur la base de cette indicatrice, notée δ_i , il est possible d'évaluer l'exposition au risque d'un individu, ou le temps τ_i exact durant lequel l'individu est observé. Sont ainsi évalués respectivement l'exposition au risque et les décès survenus au global d'un portefeuille pour tout âge x et année d'observation t :

$$\sum_i \tau_i = E_{x,t} \text{ et } \sum_i \delta_i = D_{x,t}$$

Chaque individu est par ailleurs référencé non seulement à un organisme mais également à une période d'observation. Notons qu'un organisme peut présenter plusieurs périodes d'observation, du fait de l'existence notamment de différents outils de gestion de contrats.

iii. Retraitements

Bien que de premiers retraitements aient été réalisés en amont dans la base commune transmise (et non définitive pour l'élaboration des tables IA), une nouvelle étude de

²³ TOMAS, J. et PLANCHET, F. [2011] « Construction et validation des références de mortalité de place », Institut des Actuaire, II 1291-11 v1.4 explicite le contenu non confidentiel de ces données.

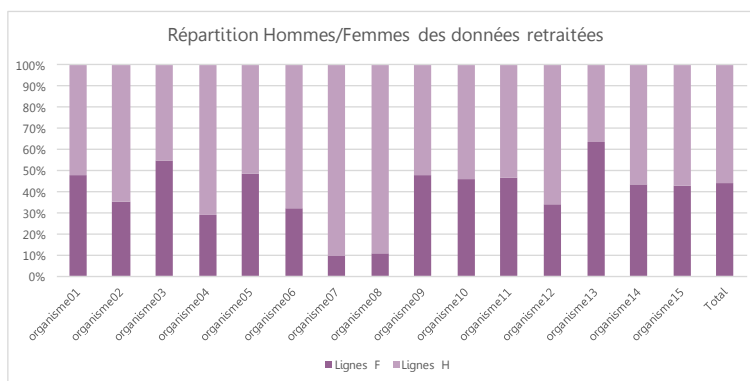
cohérence²⁴ a ici été menée et a conduit à la suppression d'environ 500 000 lignes, soit 6% de la base.

Suite à ces retraitements, les 7 711 010 restantes, dont 56% de lignes relatives aux hommes, sont synthétisées dans les tableaux ci-dessous.

Femmes							
Organisme	Date début observation organisme	Date fin observation organisme	Moyenne âge entrée	Moyenne âge sortie	Moyenne durée exposition	Moyenne âge décès	Nombre de lignes
1	01/01/1952	31/12/2011	34,57	41,05	6,48	54,99	12 960
	01/01/1967	31/12/2011	31,73	36,27	4,54	53,32	338 502
	01/01/1998	31/12/2011	67,37	79,04	11,67	83,16	743
	01/01/2009	31/12/2011	75,83	78,61	2,77	85,98	7 996
2	01/01/2004	31/12/2011	71,13	76,97	5,85	86,89	807
	01/01/2007	31/12/2011	65,78	69,68	3,90	70,91	385
	01/01/2008	31/12/2011	64,17	67,85	3,68	77,42	776
3	01/01/2000	31/12/2011	31,36	36,23	4,87	69,05	49 093
4	01/01/1999	31/12/2011	32,80	38,27	5,47	50,58	56 147
5	01/01/2006	31/12/2011	36,43	39,01	2,59	54,67	12 567
6	01/01/2001	31/12/2010	59,82	67,85	8,03	87,66	268 472
7	01/01/1997	31/12/2011	56,46	67,19	10,73	73,71	20 301
	01/01/2004	31/12/2011	73,07	79,50	6,43	87,92	4 831
8	01/01/1990	13/04/2011	64,52	76,90	12,38	85,96	13 961
	01/01/2005	13/04/2011	67,39	73,40	6,01	74,26	10 097
9	26/12/1979	30/06/2011	36,68	42,11	5,43	50,73	2 597
	01/12/1989	30/06/2011	45,11	52,99	7,88	51,07	549
	01/06/1994	30/06/2011	36,80	39,72	2,92	46,80	2 070
	01/01/1998	30/06/2011	41,54	48,72	7,18	73,47	603 856
10	31/12/1998	31/12/2011	58,00	69,60	11,60	78,61	29 759
11	01/01/2009	30/06/2011	37,99	39,91	1,92	59,28	352 269
12	01/01/2000	31/12/2011	33,91	37,80	3,89	55,72	292 899
13	01/01/1989	31/12/2011	30,17	35,25	5,08	46,37	766 741
	01/01/2008	31/12/2011	38,21	40,34	2,13	51,35	148 342
14	01/07/2000	31/12/2011	40,63	46,75	6,12	74,78	317 982
15	01/01/2000	31/12/2011	46,30	53,59	7,29	78,33	91 355

Hommes							
Organisme	Date début observation organisme	Date fin observation organisme	Moyenne âge entrée	Moyenne âge sortie	Moyenne durée exposition	Moyenne âge décès	Nombre de lignes
1	01/01/1952	31/12/2011	33,23	41,67	8,44	57,65	53 473
	01/01/1967	31/12/2011	32,55	37,55	5,00	54,31	326 436
	01/01/1998	31/12/2011	68,11	79,73	11,62	83,56	846
	01/01/2009	31/12/2011	75,75	78,45	2,70	83,22	6 676
2	01/01/2004	31/12/2011	70,47	77,11	6,64	82,55	1 746
	01/01/2007	31/12/2011	65,99	69,80	3,81	72,06	1 815
	01/01/2008	31/12/2011	61,33	65,32	4,00		18
3	01/01/2000	31/12/2010	35,67	41,50	5,83	69,68	40 654
4	01/01/1999	31/12/2011	35,76	42,05	6,30	53,11	135 137
5	01/01/2006	31/12/2011	36,60	39,11	2,51	59,07	13 407
6	01/01/2001	31/12/2010	46,91	53,80	6,89	82,27	563 826
7	01/01/1997	31/12/2011	58,62	70,35	11,73	69,94	231 221
	01/01/2004	31/12/2011	68,98	75,59	6,61	82,08	5 379
8	01/01/1990	13/04/2011	58,47	71,78	13,31	77,50	184 475
	01/01/2005	13/04/2011	68,05	73,94	5,89	74,66	7 118
9	26/12/1979	30/06/2011	37,79	43,42	5,63	51,08	5 450
	01/12/1989	30/06/2011	45,17	53,09	7,92	54,68	766
	01/06/1994	30/06/2011	40,26	43,48	3,22	47,06	1 408
	01/01/1998	30/06/2011	41,63	49,37	7,76	71,55	654 321
10	31/12/1998	31/12/2011	57,44	68,45	11,01	77,56	34 928
11	01/01/2009	30/06/2011	37,96	39,89	1,93	58,03	402 111
12	01/01/2000	31/12/2011	35,98	40,63	4,64	54,92	570 141
13	01/01/1989	31/12/2011	32,02	35,44	3,42	48,42	380 550
	01/01/2008	31/12/2011	39,17	41,30	2,13	62,24	143 287
14	01/07/2000	31/12/2011	39,51	45,94	6,43	70,46	417 063
15	01/01/2000	31/12/2011	44,29	51,62	7,33	70,59	122 701

Tableau 1 - Récapitulatif des données retraitées Femmes/Hommes par organisme



Graphique 5 – Répartition Hommes/Femmes par organisme des données retraitées

iv. Période d'observation

Usuellement, pour limiter le caractère hétérogène induit par l'évolution des taux de mortalité au cours du temps, une période d'observation restreinte d'exploitation est utilisée : c'est ce qui a été fait pour la construction des tables *best estimate* avec une plage d'observation réduite entre 2007 et 2011. Sur cette période de 5 ans, les décès et l'exposition ont alors été comptabilisés sans retraitement particulier.

²⁴ La liste exhaustive des retraitements et résultats est présentée en annexe

Dans la perspective d'évaluer la tendance propre à un portefeuille d'assurés, il est primordial de disposer d'une période d'observation la plus large possible. C'est dans ce cadre que le modèle de Cox a été appliqué.

2- PRISE EN COMPTE DE L'HETEROGENEITE PAR APPLICATION DU MODELE DE COX

i. Théorie

La principale hypothèse de cette méthode, exploitée par Tomas (2013), consiste à considérer que l'hétérogénéité de chaque portefeuille et notamment leur observation à des périodes différentes peut être décrite à l'aide d'un modèle à hasard proportionnel.

A partir de cette considération, l'idée est donc d'estimer le risque relatif de chaque portefeuille à partir d'un modèle de Cox sans avoir besoin de faire une nouvelle hypothèse sur la forme de la fonction de hasard. Chaque fonction de hasard d'une population p s'exprime de la manière suivante :

$$\mu_x^p(t) = \alpha_p \cdot \mu_x^0(t)$$

avec $\mu_x^0(t)$, la fonction de hasard inconnue de référence et $\alpha_p = \exp(z_p^T \cdot \theta_p)$. Dans notre cas, θ_p mesure l'influence d'appartenir à la population p sur l'intensité, et z_p^T représente le vecteur des covariables pour chaque individu, ou, en d'autres termes, une variable fictive indiquant l'appartenance ou non d'un individu à la population p . Nous verrons que les coefficients α_p obtenus seront ensuite utilisés pour ajuster le poids de chaque portefeuille dans le calcul de l'exposition globale.

La seconde hypothèse, déjà évoquée en partie 1, consiste à considérer que la force de mortalité est constante par morceaux. Selon les mêmes notations introduites précédemment, la vraisemblance s'écrit alors :

$$\mathcal{L}(\mu_x(t)) = \exp(-E_{x,t} \cdot \mu_x(t)) \cdot (\mu_x(t))^{D_{x,t}}$$

Et la log-vraisemblance :

$$\ell(\mu_x(t)) = \log[\mathcal{L}(\mu_x(t))] = -E_{x,t} \cdot \mu_x(t) + D_{x,t} \cdot \log \mu_x(t)$$

Cette expression fait apparaître la proportionnalité de la log-vraisemblance à une vraisemblance de Poisson du nombre de décès qui s'exprimerait de la manière suivante : $D_{x,t} \sim \mathcal{P}(E_{x,t} \cdot \mu_x(t))$. En effet, maximiser la log-vraisemblance donne par ailleurs l'estimation

$$\hat{\mu}_x(t) = \frac{D_{x,t}}{E_{x,t}}.$$

Ainsi, en partant du principe que le modèle de Poisson des décès s'applique pour chaque portefeuille p :

$$D_{x,t}^p \sim \mathcal{P}(E_{x,t}^p \cdot \mu_x^p(t))$$

Par somme,

$$\sum_p D_{x,t}^p \sim \mathcal{P}\left(\sum_p E_{x,t}^p \cdot \mu_x^p(t)\right)$$

Et par intégration du modèle de Cox,

$$\sum_p D_{x,t}^p \sim \mathcal{P}\left(\sum_p (\alpha_p \cdot E_{x,t}^p) \cdot \mu_x^0(t)\right)$$

L'application du modèle de Cox revient donc à considérer un portefeuille agrégé global dont les décès $D_{x,t}^f$ sont évalués par somme des décès de l'ensemble des portefeuilles p le composant, $D_{x,t}^f = \sum_p D_{x,t}^p$ et suivent une loi de poisson de type :

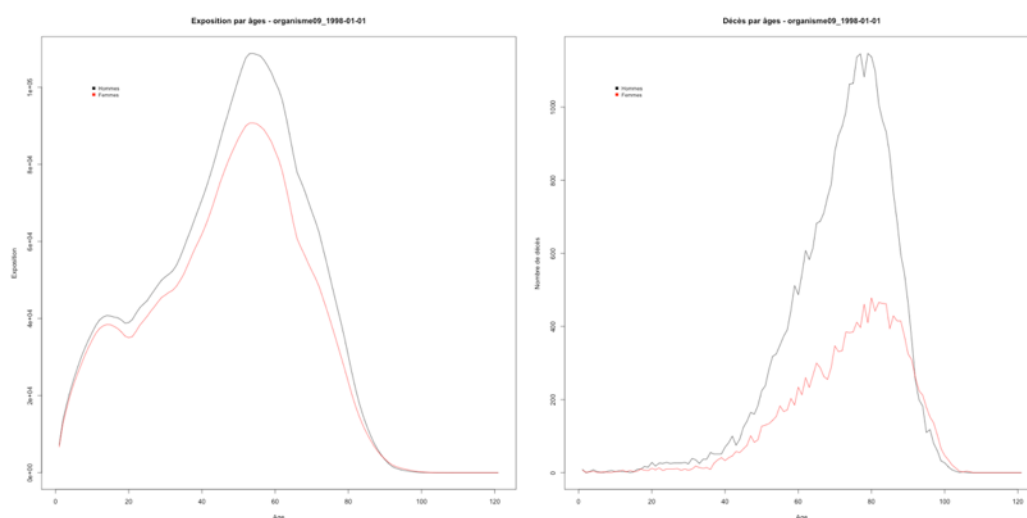
$$D_{x,t}^f \sim \mathcal{P}(E_{x,t}^f \cdot \mu_x^0(t))$$

avec $E_{x,t}^f$, l'agrégation de l'exposition des portefeuilles p pondérée des coefficients de Cox $E_{x,t}^f = \sum_p (\alpha_p \cdot E_{x,t}^p)$, et $\mu_x^0(t)$, la fonction de hasard inconnue, que nous allons désormais chercher à évaluer.

ii. Application aux données

En pratique, dans la mesure où un organisme peut présenter plusieurs périodes d'observation, il a été considéré qu'un portefeuille serait caractérisé par un organisme et une période d'observation donnée : ce qui porte à 26 le nombre de portefeuilles p pris en compte, tel qu'explicité en Tableau 1.

Pour chacun de ces portefeuilles p , l'exposition et le décès ont été évalués par âge et par année d'observation. Ces calculs font bien évidemment intervenir le principe de censure à droite intrinsèque aux données issues des différents organismes, dans la mesure où le sinistre d'un individu peut être ultérieur à sa période de présence ; tout comme le principe de troncature gauche liée aux éventuelles franchises des contrats.



Graphique 6 – Exposition et décès sommés par âge sur la période d'observation du portefeuille 9 et la période d'observation débutant le 01/01/1998

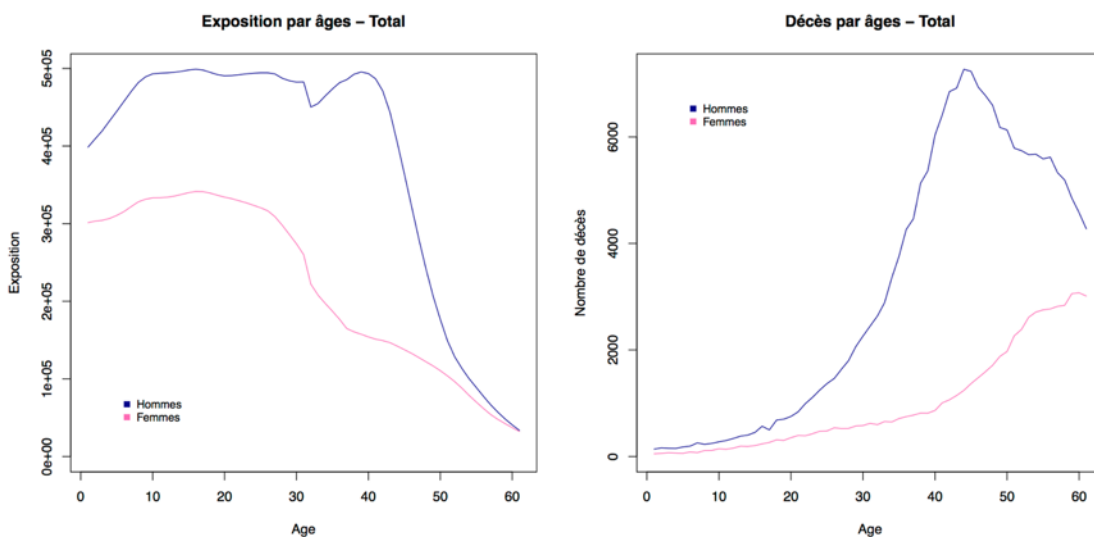
Parallèlement, le modèle de Cox a été appliqué et les coefficients obtenus sont détaillés dans le Tableau 2 ci-dessous. L'évaluation des coefficients a été faite avec le logiciel R, comme l'ensemble des calculs et analyses de ce mémoire, par le biais d'un algorithme de type Newton-Raphson²⁵. Le portefeuille 9 dont l'année d'observation débute au 01/01/1998 a par ailleurs été choisi comme portefeuille de référence pour les deux sexes dans la mesure où il est celui présentant le plus grand nombre de lignes.

²⁵ Modèle développé dans le package « Survival » (fonctions utilisées *coxph* et *surv*)

Portefeuille	Femmes	Hommes
organisme01-01/01/1952	2,41348	4,19605
organisme01-01/01/1967	0,43165	0,51621
organisme01-01/01/1998	0,41655	0,34326
organisme01-01/01/2009	1,23275	1,39363
organisme02-01/01/2004	0,65701	0,52672
organisme02-01/01/2007	1,45000	0,74455
organisme02-01/01/2008	1,50284	0,00151
organisme03-01/01/2000	1,36342	1,38975
organisme04-01/01/1999	1,17694	1,34699
organisme05-01/01/2006	2,32624	1,64333
organisme06-01/01/2001	0,93227	0,90747
organisme07-01/01/1997	1,09925	1,08021
organisme07-01/01/2004	1,17494	1,15277
organisme08-01/01/1990	1,02778	1,03864
organisme08-01/01/2005	0,94487	0,87479
organisme09-12/26/1979	0,65230	0,71247
organisme09-12/01/1989	1,04358	1,06632
organisme09-06/01/1994	1,00855	0,37984
organisme09-01/01/1998	1,00000	1,00000
organisme10-12/31/1998	1,01319	1,02925
organisme11-01/01/2009	1,52795	1,54612
organisme12-01/01/2000	1,23189	1,10444
organisme13-01/01/1989	0,76922	1,06015
organisme13-01/01/2008	0,45900	0,58034
organisme14-07/01/2000	0,78453	0,72987
organisme15-01/01/2000	1,48698	1,27996

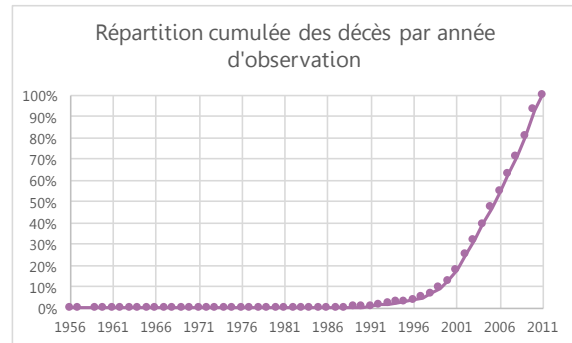
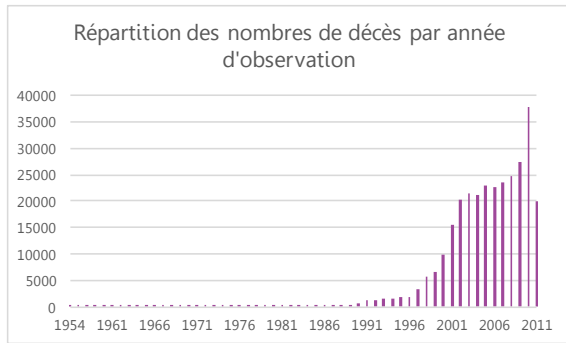
Tableau 2 – Coefficients α^p de Cox Hommes et Femmes obtenus

Par somme selon les principes théoriques exposés précédemment, le portefeuille agrégé de décès et d'exposition est ainsi élaboré.



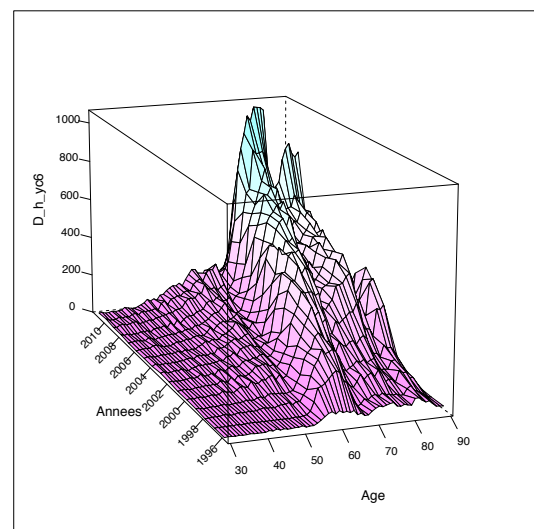
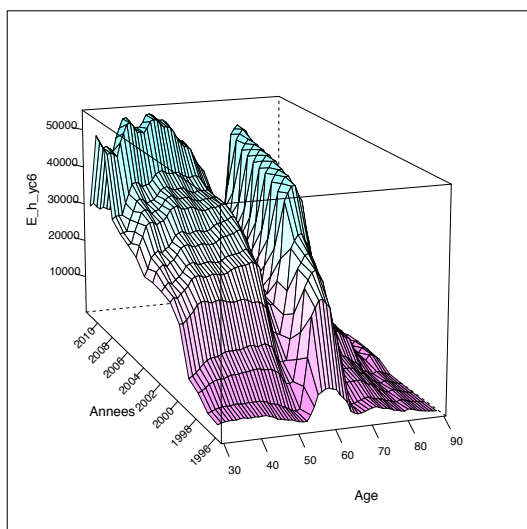
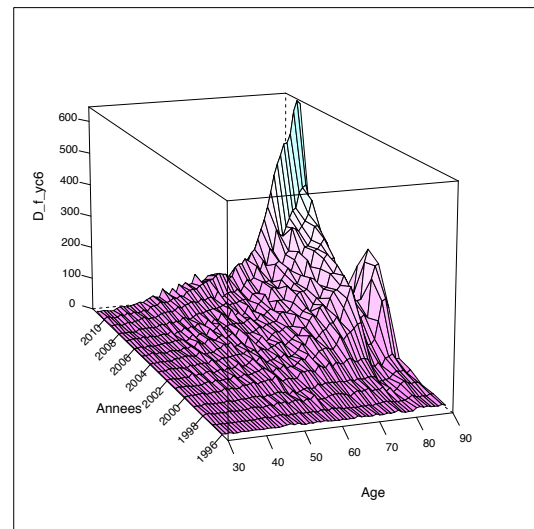
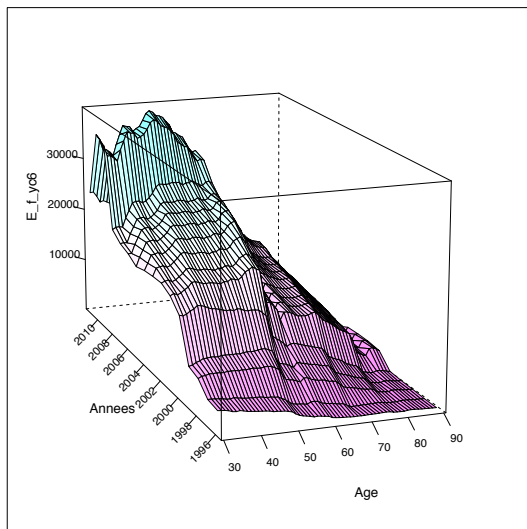
Graphique 7 – Exposition et décès sommés par âge sur la période d'observation du portefeuille agrégé

De manière à disposer d'une base de données d'exposition et décès exploitable, nous décidons de restreindre le portefeuille agrégé à la période 1996 – 2011. En effet, avant cette période, le nombre de décès comptabilisés, évalué à moins de 3% de l'ensemble des décès observés entre 1952 et 2011, est bien trop faible, comme l'illustrent les graphiques ci-dessous.



Graphique 8 – Répartition des décès par année d'observation du portefeuille agrégé

Sur cette période restreinte, les surfaces de décès et exposition du portefeuille agrégé font apparaître pour les hommes un pic autour de 60 ans qui correspond à un portefeuille significatif.



Graphique 9 – Surface des expositions (E) et décès (D) par âge et année d'observation du portefeuille agrégé Hommes (h) et Femmes (f)

iii. **Elaboration de la table Femmes et Hommes des moments**

Nous avons choisi d'estimer les taux bruts de mortalité à l'aide de l'estimateur des moments de Hoem. Cet estimateur a pour avantage d'être sans biais et de prendre en compte les censures et troncatures (Planchet & Therond, 2006). L'exposition $E_{x,t}$ illustrée précédemment a ainsi été calculée en ne prenant en compte que les périodes exactes de présence de chaque individu i pour l'âge x et une année t d'observation, donnés.

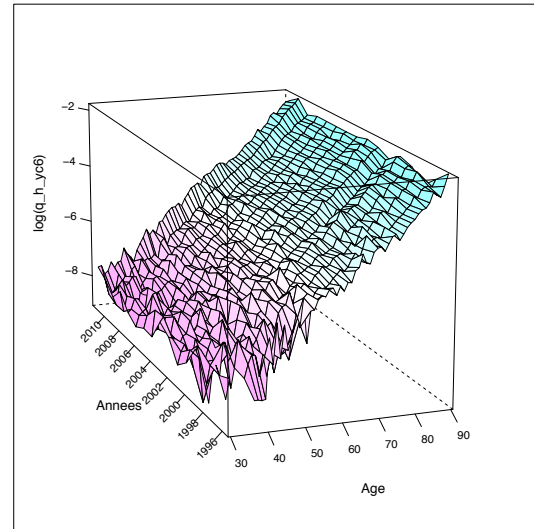
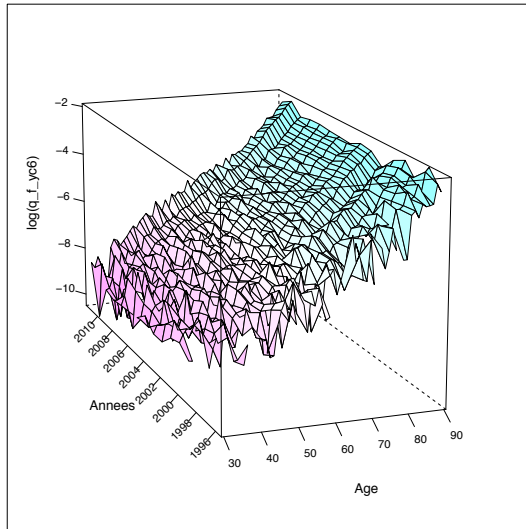
Considérons τ_i^d (respectivement τ_i^f), la date de début d'observation de l'assuré d'âge x (la date de fin d'observation de l'assuré d'âge x par sinistre, par censure ou par passage à l'âge $x+1$). En reprenant les notations utilisées au **ii-Données initiales – ligne à ligne**, la durée d'observation τ_i de chaque individu i dans l'intervalle $[x, x + 1]$, s'écrit donc $\tau_i = \tau_i^f - \tau_i^d$ et l'expression exacte de l'exposition : $E_{x,t} = \sum_i \tau_i = \sum_i \tau_i^f - \tau_i^d$.

Par ailleurs, l'indicatrice δ_i représentant l'état sinistré ou non de l'individu i suit une loi de Bernoulli de paramètre $\frac{\tau_i^f - \tau_i^d}{\tau_i} q_{x+\tau_i^d}$ ou encore de paramètre $(\tau_i^f - \tau_i^d) \cdot q_x$, soit $\tau_i \cdot q_x$, en considérant que la probabilité d'un individu de décéder d'un âge à un autre est une fonction linéaire du temps. $D_{x,t} = \sum_i \delta_i$ suit une loi binomiale de probabilité $\tau_i \cdot q_x$.

L'estimateur de Hoem \hat{q}_x de q_x s'évalue donc de la manière suivante :

$$\hat{q}_x = \frac{\sum_i \delta_i}{\sum_i \tau_i} = \frac{D_{x,t}}{E_{x,t}}$$

La représentation de la nappe du logarithme de mortalité du portefeuille agrégé reflète des taux de mortalité plutôt constants et non croissants comme attendu aux âges élevés proches de 90 ans. De plus, un pic important pour l'année 2010 et un autre moins conséquent pour l'année 2011, quel que soit l'âge et quel que soit le sexe, est visible. Cette forme particulière de tables des moments sera étudiée plus en détail dans la troisième partie, lors de l'évaluation de la tendance propre à ce portefeuille agrégé d'assurés construit.



Nappe 5 – Logarithme du taux de mortalité Femmes (f) et Hommes (h) du portefeuille agrégé

Pour clore cette 2^{ème} partie, démontrons dans un premier temps par l'absurde que la tendance du portefeuille étudié est bien différente de celle de la population nationale.

C. APPLICATION DU MODELE DE BRASS – JUSTIFICATION D’UNE TENDANCE

« ASSURES » PROPRE

1- THEORIE

Le modèle de Brass (1971) est un modèle relationnel permettant l’ajustement de taux de mortalité bruts d’une population, q , à partir d’une table de référence, q^{ref} , par le biais de la fonction logistique (logit). La méthode qu’il est ici proposé d’appliquer est une variante semi paramétrique, qui exprime les taux de mortalité bruts ajustés \tilde{q} de la manière suivante :

$$\text{logit}({}_t\tilde{q}_{x^*}) = \theta_1 + \theta_2 \cdot \text{logit}({}_tq_{x^*}^{ref}) + \varepsilon_{x^*,t^*}$$

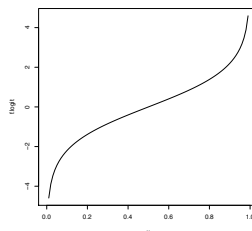
avec x^* et t^* respectivement les plages d’âges et période d’observation communes avec la référence de mortalité et ε_{x^*,t^*} , le terme d’erreur.

Le paramètre θ_1 est un indicateur de mortalité qui impacte tous les âges, alors que le paramètre θ_2 modifie cet effet avec l’âge. L’estimation est effectuée par une minimisation de la distance pondérée de l’exposition entre les probabilités de décès observés et estimés sur la période d’étude : $\sum |\varepsilon_{x^*,t^*} * ({}_t\tilde{q}_{x^*} - {}_tq_{x^*})|$. Les estimateurs de θ_1 et θ_2 , respectivement $\widehat{\theta}_1$ et $\widehat{\theta}_2$, permettent d’exprimer les taux de mortalités ajustés :

$${}_t\tilde{q}_x = \frac{\exp(\widehat{\theta}_1 + \widehat{\theta}_2 \cdot \text{logit}({}_tq_x^{ref}))}{1 + \exp(\widehat{\theta}_1 + \widehat{\theta}_2 \cdot \text{logit}({}_tq_x^{ref}))} \quad (1)$$

Dans la mesure où le modèle semi-paramétrique relationnel a l’avantage d’intégrer l’estimation et l’extrapolation des tendances de mortalité (PLANCHET F. et THEROND P., [2011]), cette formule peut s’appliquer pour extrapoler la mortalité au-delà de la tranche d’âge utilisée pour l’ajustement mais aussi au-delà de la période d’observation.

Il est à noter que ce type de modèle conduit usuellement à sous évaluer la mortalité du portefeuille dès lors que les taux de mortalités sont inférieurs à 1/2. En effet, la fonction logit est concave sur $]0; 1/2[$, comme le montre le graphique ci-dessous, ce qui induit, par l’inégalité de Jensen la relation suivante : $E(\text{logit}({}_t\tilde{q}_x)) \leq \text{logit}({}_t\tilde{q}_x)$.



Graphique 10 - Représentation graphique de la fonction logistique (logit)

Ce modèle sera par la suite noté « Modèle de Brass » ou modèle « M2²⁶ ».

2- TABLE DES MOMENTS AJUSTEE

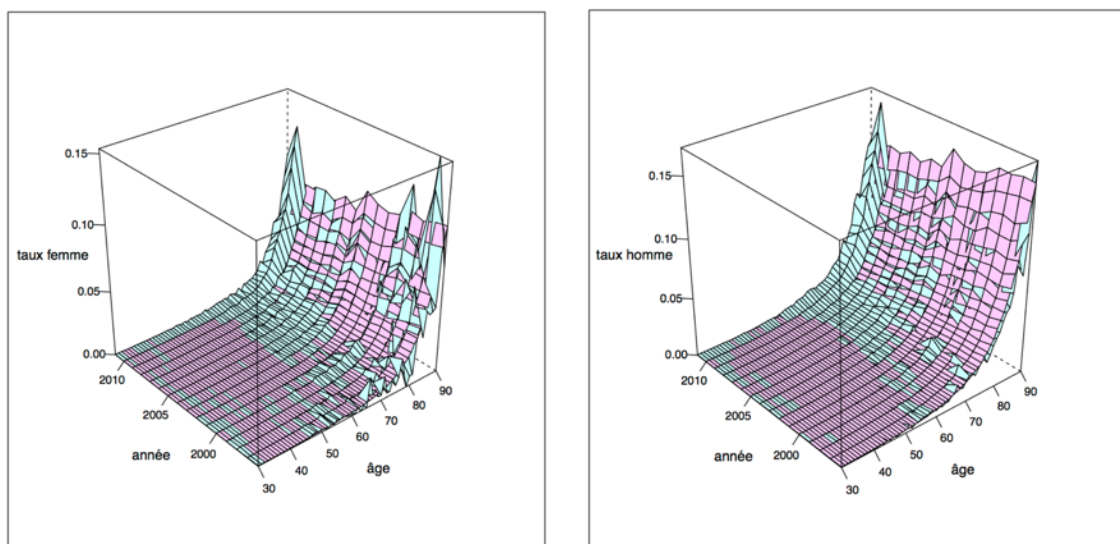
L'ajustement a été réalisé sur la période 1996-2011, pour la tranche d'âges compris entre 30 et 90 ans avec pour référence les taux bruts de l'INSEE (tables qmINSEE)²⁷.

	$\widehat{\theta}_1$	$\widehat{\theta}_2$
Femmes	-0,1468	1,0781
Hommes	0,0222	1,1609

Tableau 3 - Paramètres d'ajustement du modèle de Brass M2

i. Comparaison des taux bruts et des taux ajustés

La comparaison de la surface de mortalité ajustée (surface rose) obtenue à partir des paramètres d'ajustement explicités dans le tableau ci-dessus et de l'équation (1) avec la surface de mortalité observée (surface bleue) fait à nouveau apparaître le pic de mortalité de taux bruts pour l'année 2010 pour les hommes et pour les femmes.

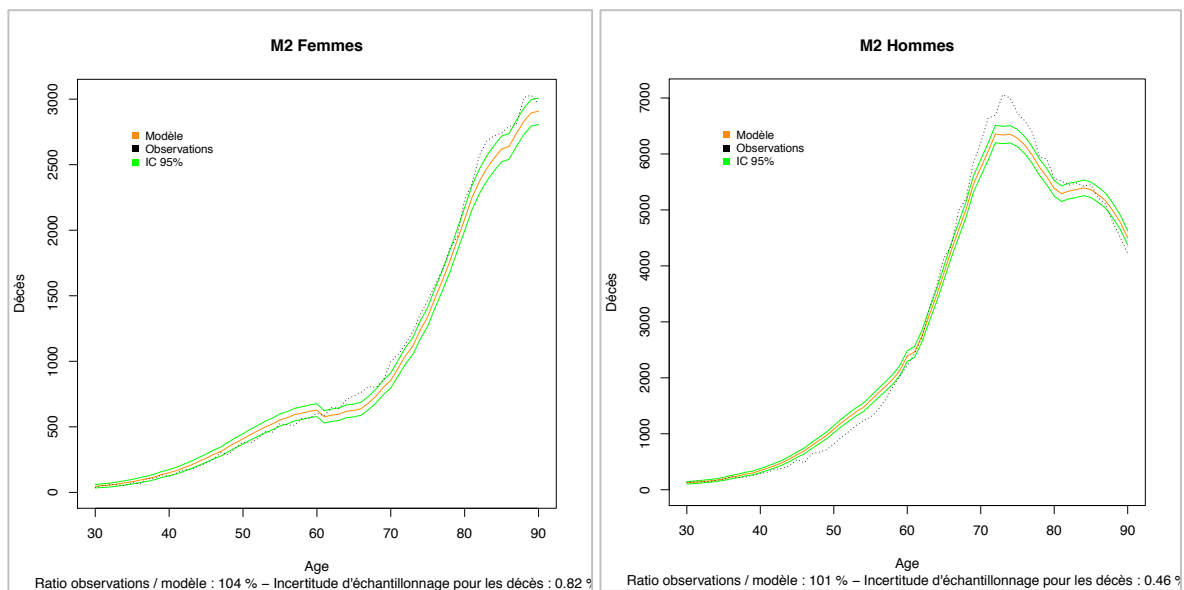


Surface 4 – Comparaison des taux de mortalités Observés (bleus) / Ajustés (roses) - modèle M2

²⁶ « M2 » en référence à la Méthode 2 explicitée dans la note II1291-12 v1.7 « Méthodes de positionnement : aspects méthodologiques » de l'Institut des Actuaire

²⁷ Modélisation sous R de l'optimisation de la distance (avec l'utilisation de la fonction « constrOptim »)

Ce pic principalement concentré sur les âges assez élevés, et qui n'est pas reproduit sur les taux ajustés se retrouve également dans le graphique cumulé par âge ci-dessous. Le modèle sous-estime les observations en deçà de l'intervalle de confiance à 95% (IC 95%). Cette sous-estimation est particulièrement prononcée pour les femmes entre 60 et 70 ans puis entre 80 et 90 ans, et pour les hommes entre 65 à 75 ans, et ne peut s'expliquer que par la seule utilisation de la fonction logit du modèle. A contrario, entre 45 et 60 ans le modèle surévalue les décès observés pour les hommes, ce qui conduit à une surestimation globale moyenne du modèle de seulement 1% pour les hommes contre 4% pour les femmes.



Graphique 11 – Comparaison des décès observés et ajustés et intervalle de confiance à 95 % - modèle M2

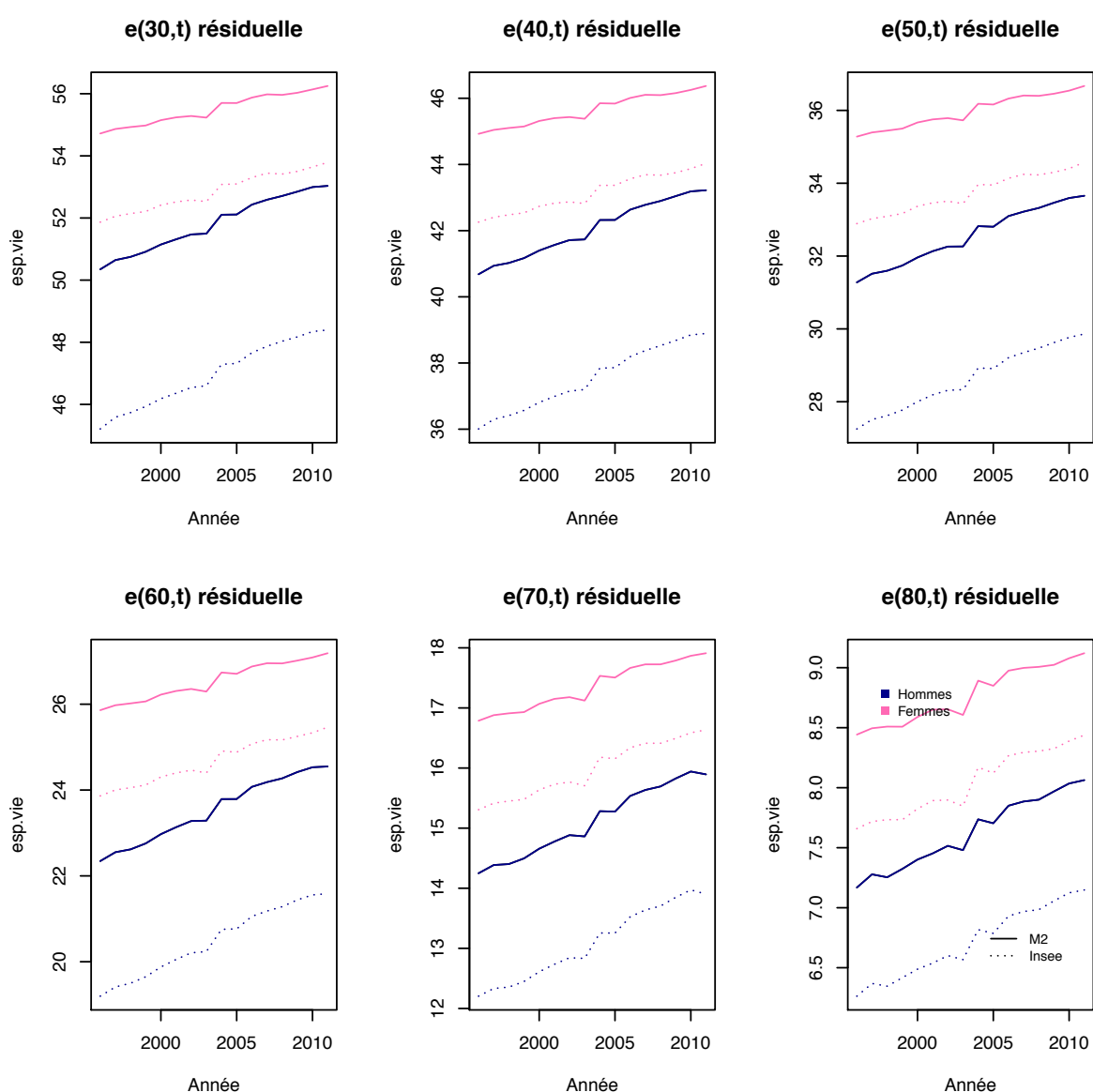
L'intervalle de confiance a été évalué selon le principe usuel évoqué de considérer que les décès $D_{x,t}$ suivent une loi binomiale de paramètre $(E_{x,t}, {}_tq_x)$. En effet, à l'appui du Théorème Central Limite, la loi des décès peut ainsi être approchée par une loi normale d'espérance $E_{x,t} \cdot {}_tq_x$ et de variance $E_{x,t} \cdot {}_tq_x \cdot (1 - {}_tq_x)$. En conséquence, l'intervalle de confiance à 95% s'écrit :

$$IC_{95\%} = \left[\tilde{D}_{x,t} - 1,96 \sqrt{E_{x,t} \cdot {}_t\tilde{q}_x \cdot (1 - {}_t\tilde{q}_x)}; \tilde{D}_{x,t} + 1,96 \sqrt{E_{x,t} \cdot {}_t\tilde{q}_x \cdot (1 - {}_t\tilde{q}_x)} \right]$$

avec $\tilde{D}_{x,t}$ et ${}_t\tilde{q}_x$, respectivement le nombre de décès et les taux de décès obtenus après ajustement.

ii. Comparaison à la table de référence Insee

La comparaison des taux ou décès bruts et ajustés reflète bien la forte capacité de « lissage » du modèle relationnel semi-paramétrique construit. Il reste assez peu flexible sur les données brutes et supprime donc les spécificités potentiellement majeures du portefeuille initial au profit du comportement de la population de référence. L'analyse de l'espérance de vie résiduelle à différents âges sur la période d'ajustement conforte cette remarque : l'évolution des espérances de vie pour les hommes et pour les femmes est équivalente à celle des tables INSEE de référence. La non régularité de l'espérance de vie vient du fait que les taux ont été ajustés sur la table INSEE brute.



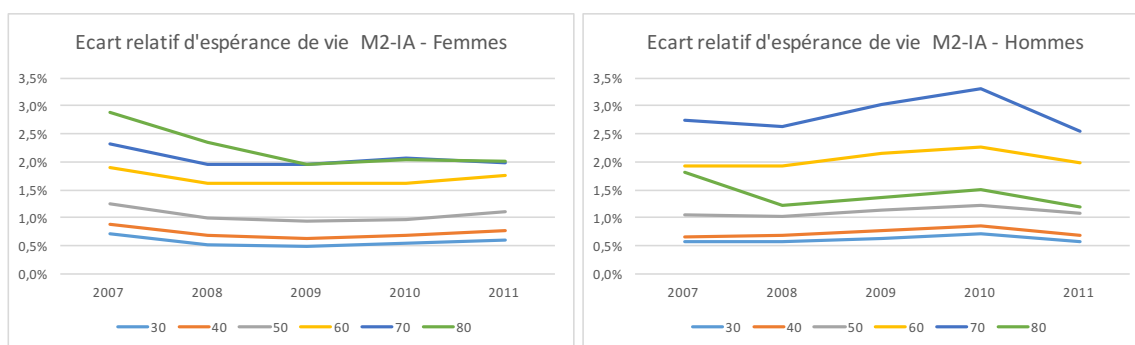
Espérance de vie résiduelle aux âges 30, 40, 50, 60, 70 et 80 ans en fonction de l'année calendaire

Pour rappel, l'espérance de vie résiduelle à chaque âge est évaluée jusqu'à un âge de 90 ans (tables non fermées). Les ordres de grandeur ici obtenus sous-estiment donc l'espérance de vie réelle²⁸.

Les taux ajustés par le modèle de Brass sont donc représentatifs de la population des assurés en termes de positionnement ou niveau mais ont capté la tendance du portefeuille de référence de la population nationale.

iii. Comparaison à la table best estimate IA

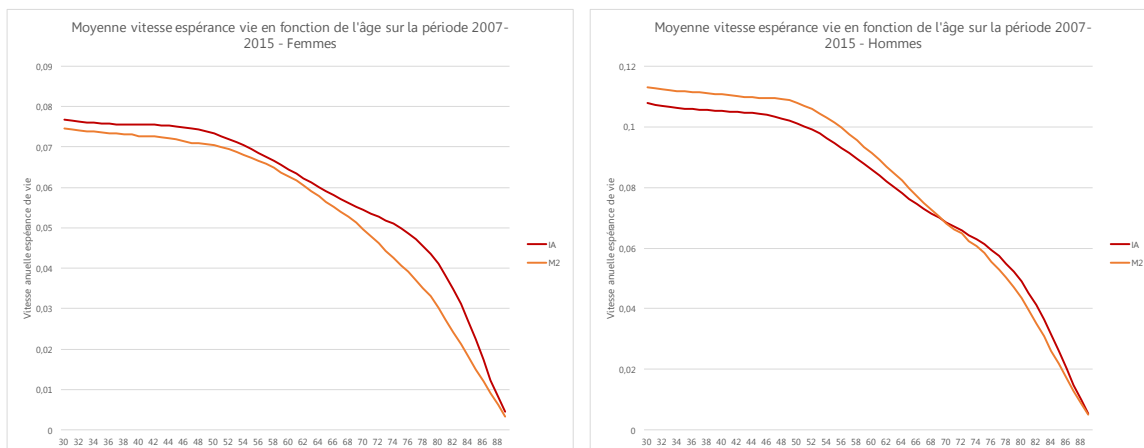
L'espérance de vie des taux ajustés est légèrement supérieure à celle des tables *best estimate*. En revanche, l'écart relatif apparaît stable dans le temps aux différents âges, ce qui est rassurant quant à la pertinence de l'application du modèle de Cox pour réduire l'hétérogénéité des données.



Graphique 12 – Ecart relatif d'espérance de vie entre l'ajustement de Brass M2 et les tables best estimate IA

La vitesse moyenne d'espérance de vie sur 2007-2015 peut effectivement être considérée comme quasiment équivalente tant pour les femmes que pour les hommes.

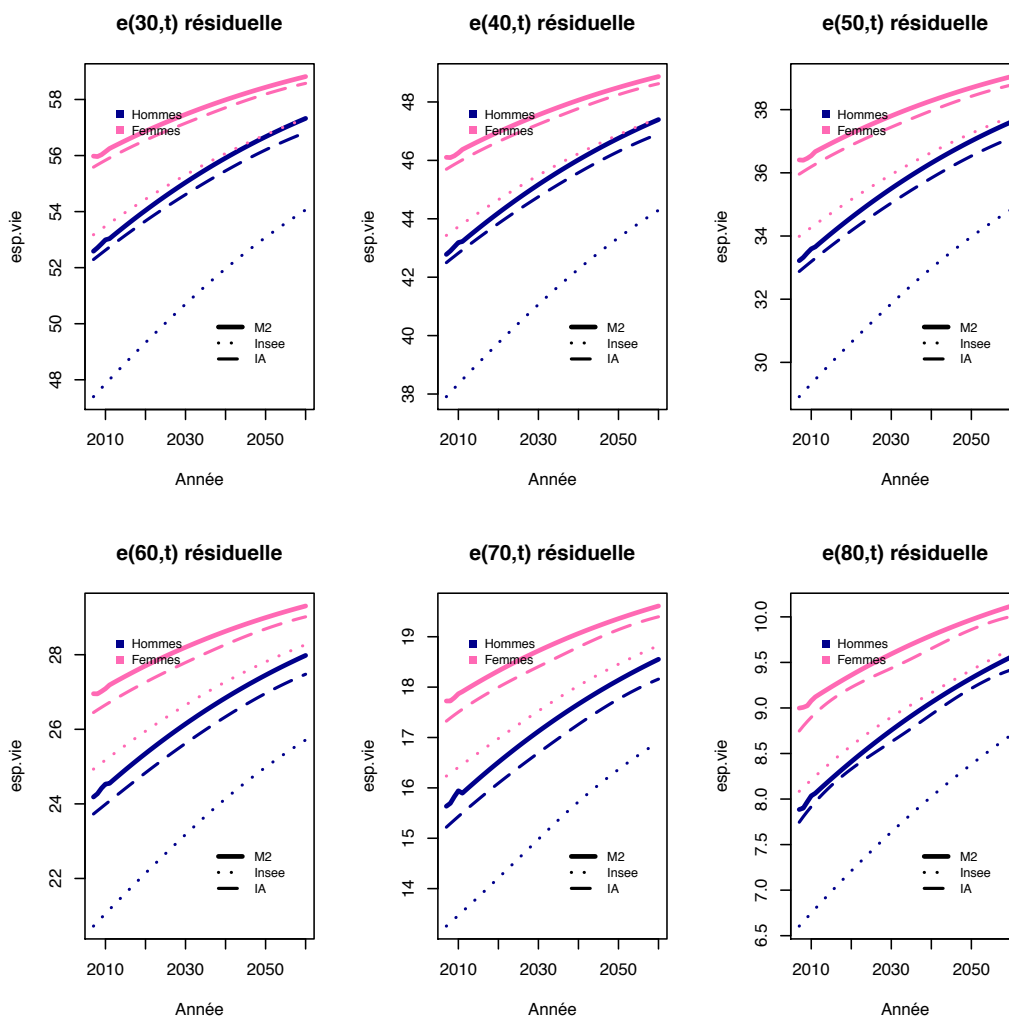
²⁸ En 2010, l'espérance de vie à 60 ans était de 22,7 ans pour les hommes et 27,4 pour les femmes – source <http://www.insee.fr/fr/themes/series-longues.asp?indicateur=esperance-vie-60ans>



Graphique 13 – Moyenne des vitesses d'espérance de vie sur 2007-2015 Brass (M2) – best estimate (IA)

3- TABLE PROSPECTIVE

Le quasi parallélisme des courbes d'espérances de vie résiduelle extrapolées jusqu'en 2060, confirme les constatations faites sur le modèle de Brass, à savoir que la tendance des taux de mortalité ajustés s'appuie sur celle de la table de référence. Cette tendance de la table de référence se retrouve tout autant dans les tables INSEE que dans les tables IA *best estimate*, du fait des modalités de construction de la table IA.

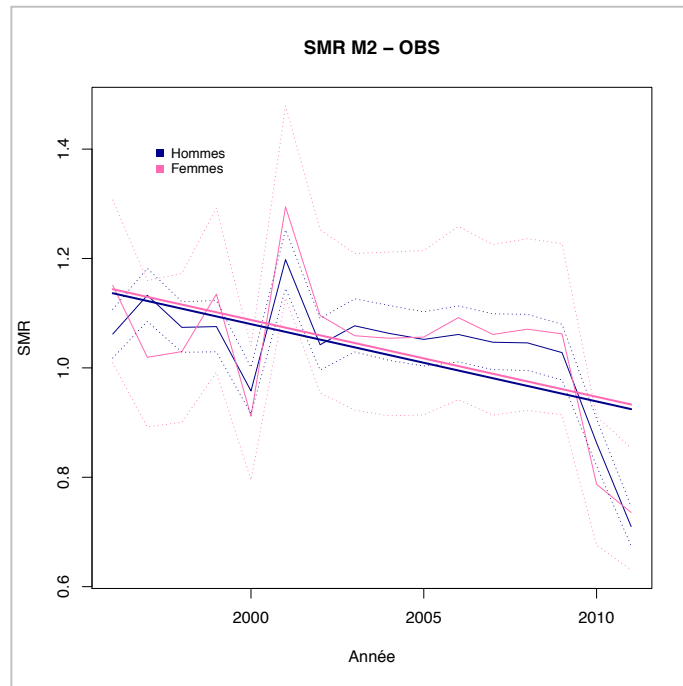


Graphique 14 – Comparaison des espérances de vie résiduelle à différents âges – Brass et références

4- PERTINENCE DE LA TENDANCE ISSUE DES TAUX AJUSTES

L'étude du SMR, représentant le ratio entre le nombre de décès annuels calculé sur les taux ajustés sur le nombre de décès observés dans le portefeuille pour chaque année de la période d'ajustement, montre que la tendance des taux ajustés ne reflète pas celle du portefeuille initial. Un modèle parfaitement ajusté sur la tendance du portefeuille devrait présenter un SMR constant, qui plus est proche et légèrement inférieur à 1, du fait de la sous-estimation de la mortalité intrinsèque à l'utilisation de la fonction logistique.

Du fait de la décroissance constatée de la régression linéaire du SMR, la dynamique d'amélioration de la mortalité du portefeuille semble plus rapide que celle issue de la table ajustée par le modèle de Brass.



Graphique 15 – SMR des décès ajustés par le modèle de Brass (M2) sur décès observés et IC 95%

CONCLUSION PARTIE 2

Nous disposons désormais d'un portefeuille agrégé de données d'assurés rendu homogène sur une période de 16 ans grâce à la régression de Cox et pour lequel le modèle relationnel de Brass a permis de démontrer l'existence d'une tendance propre et différente de celle de la population nationale.

La troisième partie a pour objectif l'évaluation de cette tendance propre à l'aide d'un modèle endogène GLM Poisson, d'en confirmer les résultats à l'aide d'un modèle de type Log-Poisson, variante plus robuste d'un modèle Lee-Carter tout en émettant des réserves sur les biais d'estimation.

PARTIE 3 : EVALUATION DE LA TENDANCE DU PORTEFEUILLE D'ASSURES AGREGE

Avant d'analyser plus en détail la tendance propre du portefeuille d'assurés agrégé, nous allons évaluer les tables des moments ajustées selon deux types de modèles endogènes.

A. TABLE DES MOMENTS AJUSTEE AVEC UN MODELE ENDOGENE DE TYPE GLM POISSON

Le premier modèle envisagé est un modèle de type GLM Poisson avec interaction de l'âge et de l'année calendaire. Ce modèle sera par la suite appelé « modèle GLM Poisson » ou modèle « M3»²⁹.

1- THEORIE

Le modèle GLM Poisson repose sur la représentativité préalablement abordée des décès par une loi de Poisson, $D_{x,t} \sim \mathcal{P}(E_{x,t} \cdot \mu_x(t))$. Il consiste ensuite à fixer une forme à la force de mortalité $\mu_x(t)$ par le biais d'un modèle linéaire généralisé, ou plus précisément d'un modèle log-linéaire, compte tenu de la fonction de lien logarithmique associé à un modèle de Poisson.

Les variables déterministes choisies du modèle log-linéaire sont : l'âge, l'année calendaire et la mortalité d'une référence, à savoir les tables INSEE brutes (qm_INSEE).

Ainsi, pour x^* et t^* , respectivement la plage d'âge et la période d'observation communes entre les données et la référence, le modèle log-linéaire s'écrit :

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

avec $\mu_{x^*}^{ref}(t^*) = -\log(1 - qm_{INSEE_{x^*}}(t^*))$, la force de mortalité de référence.

Bien que ce modèle fasse intervenir une référence, la structure du GLM considéré intégrant la dépendance en âge et année calendaire permet de capter la tendance propre du portefeuille étudié.

²⁹ « M3 » en référence à la Méthode 3 explicitée dans la note II1291-12 v1.7 « Méthodes de positionnement : aspects méthodologiques » de l'Institut des Actuaire

La force de mortalité ajustée s'obtient ensuite à l'aide des paramètres obtenus du modèle log-linéaire, par la formule suivante :

$$\hat{\mu}_x = \exp \left[\hat{\beta}_0 + \hat{\beta}_1 \cdot \log \left(\mu_x^{ref}(t) \right) + \hat{\beta}_2 \cdot x + \hat{\beta}_3 \cdot t + \hat{\beta}_4 \cdot x \cdot t \right]$$

et les taux de mortalité ajustés découlent de la relation : $\hat{q}_x = 1 - \exp(-\hat{\mu}_x)$.

2- RESULTATS

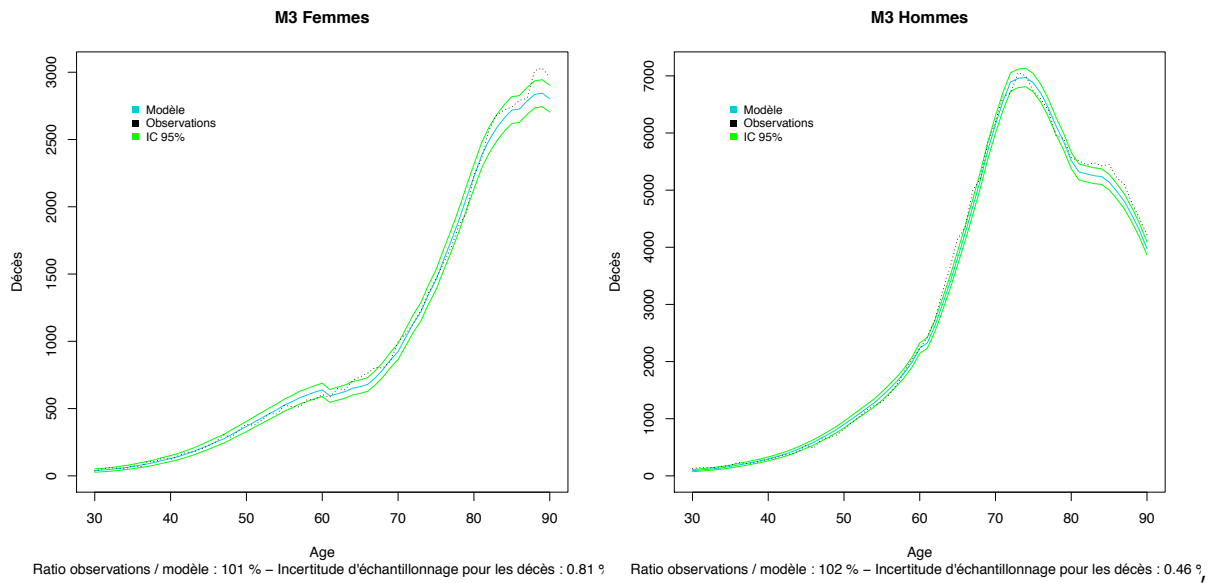
Le modèle linéaire généralisé est ainsi appliqué sur les données du portefeuille d'assurés agrégé sur la période commune 1996-2011, avec pour variable offset, $\log(E_{x,t})$, de manière à équilibrer le modèle.

Les résultats obtenus du modèle log-linéaire manifestent une excellente significativité de l'ensemble des variables, tant pour les femmes que pour les hommes (p value << 0,05)

Femmes					Hommes				
Coefficient	Estimate	Std. Error	z value	p value	Coefficient	Estimate	Std. Error	z value	p value
β_0	47,0400	13,6400	3,448	0,000564	β_0	51,0800	8,1870	6,239	4,41E-10
β_1	0,8201	0,0195	42,047	< 2e-16	β_1	0,4997	0,0204	24,549	< 2e-16
β_2	-1,3980	0,1801	-7,765	8,15E-15	β_2	-0,8835	0,1111	-7,949	1,88E-15
β_3	-0,0251	0,0068	-3,680	0,000233	β_3	-0,0288	0,0041	-7,030	2,07E-12
β_4	0,0007	0,0001	7,920	2,38E-15	β_4	0,0005	0,0001	8,493	< 2e-16

Tableau 4 – Résultats du modèle GLM Poisson exploité sur 1996-2011

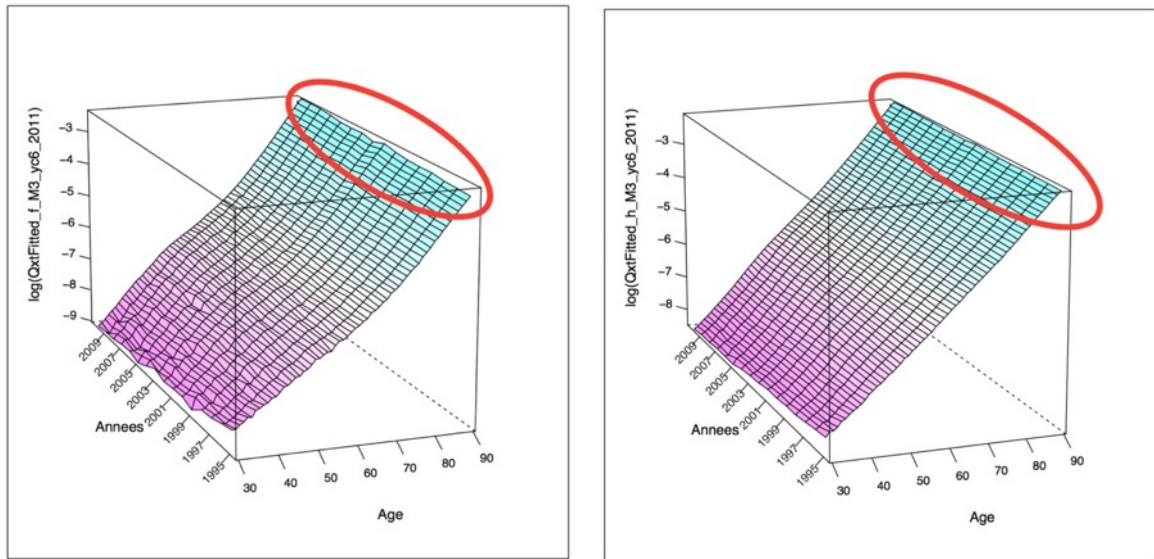
L'étude de l'adéquation entre les données observées et les données obtenues démontre également d'une bonne qualité d'ajustement, avec une incertitude inférieure à 1 % pour les femmes et inférieure à 0,5 % pour les hommes.



Graphique 16 - Comparaison des décès observés et ajustés et intervalle de confiance à 95 % - modèle M3

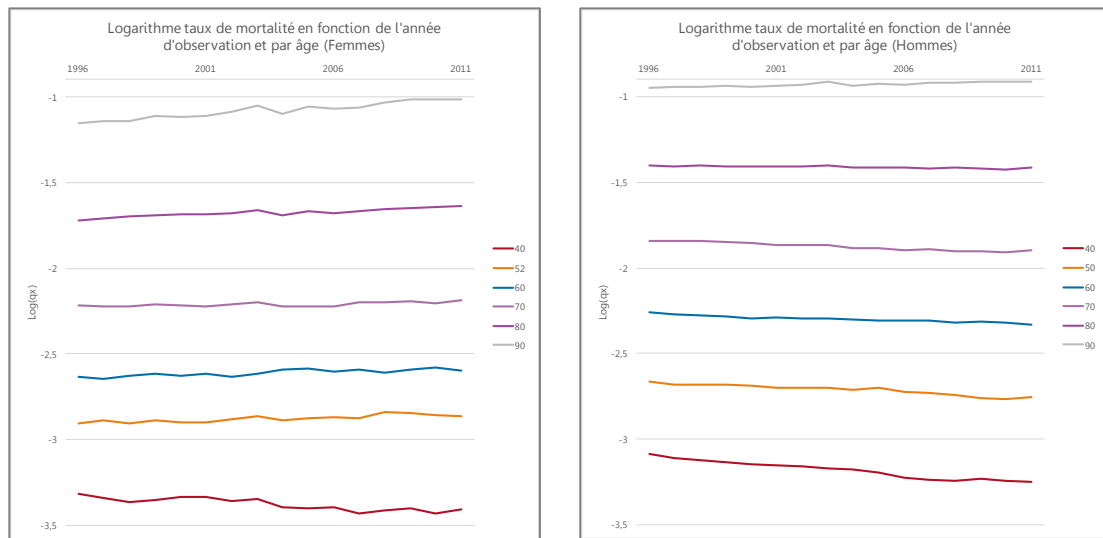
Quel que soit l'âge, les données observées sont quasiment incluses dans l'intervalle de confiance à 95 % du modèle, ce qui apparaît bien meilleur que le modèle de Brass détaillé en 2^{ème} partie. Par sa structure, le modèle GLM Poisson s'ajuste et reflète naturellement mieux l'information contenue dans les données observées.

En revanche, la représentation des nappes de mortalité ajustées fait apparaître une tendance surprenante qui conduirait à penser qu'avec l'année d'observation et à un âge donné, la durée de vie diminue. Les taux de mortalité ajustés sont en effet croissants avec l'année calendaire pour les âges avancés. Cette observation est surtout flagrante pour les femmes dans les graphiques ci-dessous.



Nappe 6 – Logarithme des taux de mortalité ajustés avec le modèle GLM Poisson exploité sur 1996-2011

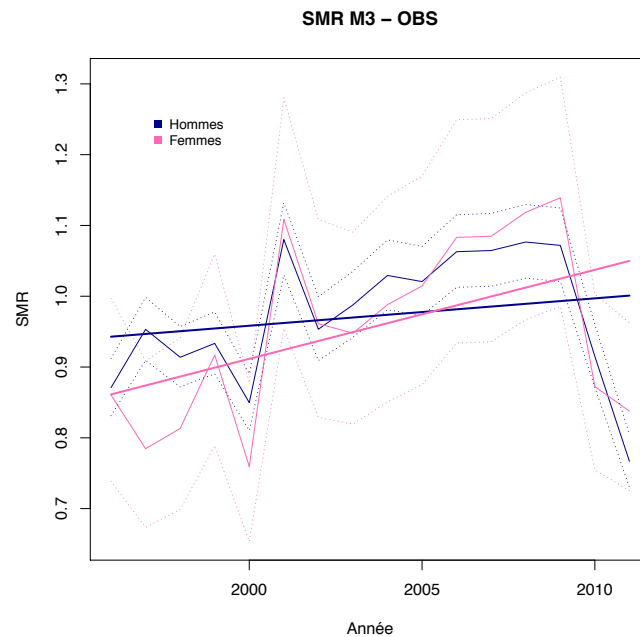
Une analyse par âge reflète cette même croissance inattendue des taux de mortalités, plus précisément pour les femmes de plus de 50 ans et pour les hommes de plus de 60 ans.



Graphique 17 – Logarithme des taux de mortalité à différents âges en fonction de l'année du modèle M3 exploité sur 1996-2011

Le SMR (ratio entre décès ajustés et décès observés) est non constant en moyenne et illustre ainsi le fait que le modèle ne capte pas la tendance intégrée dans les observations. Le nombre de décès ajustés sous-estime les observations en début de période pour les

surestimer en fin de période. Cette observation est à nouveau plus marquée pour les femmes que pour les hommes.

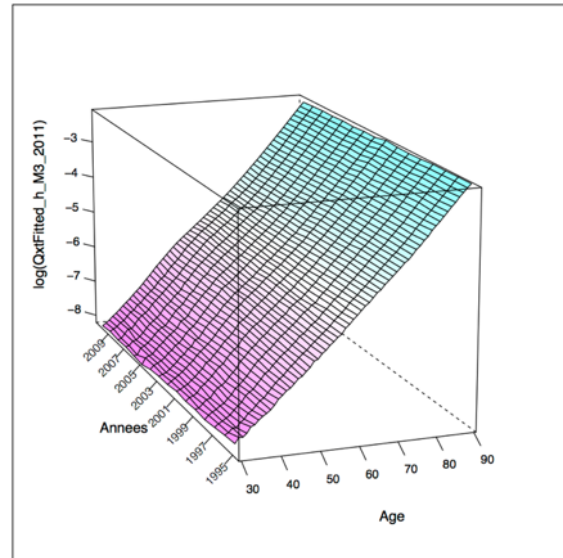
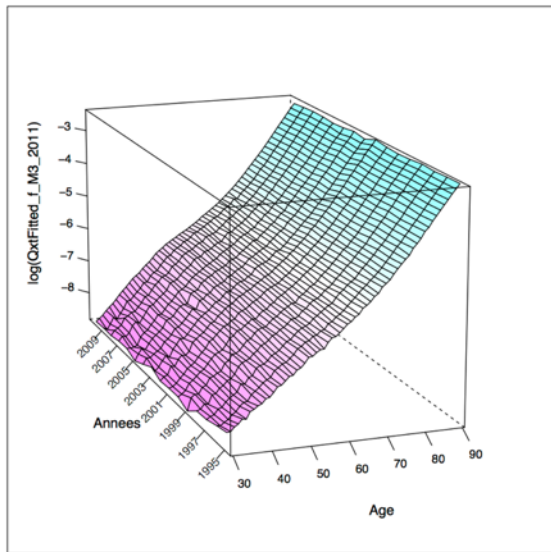


Graphique 18 – SMR des décès ajustés par le modèle GLM Poisson (M3) sur décès observés et IC 95%

3- ÉTUDE D'IMPACT DES PORTEFEUILLES UTILISÉS POUR L'AGREGATION

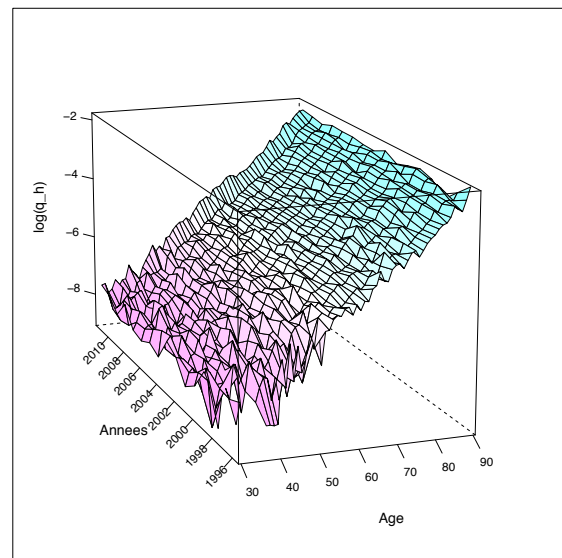
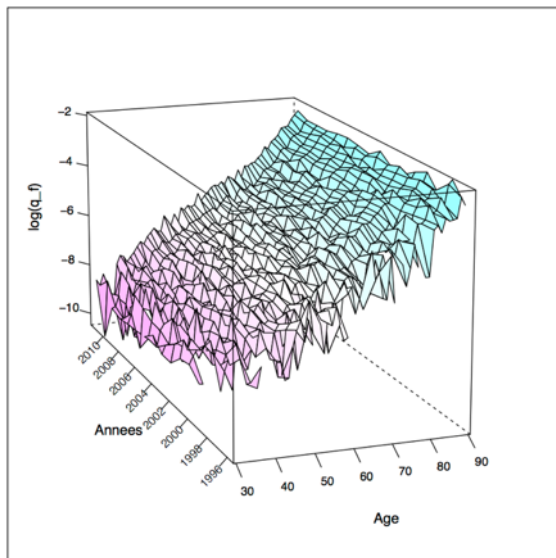
Ce constat a conduit à une analyse d'impact des différents portefeuilles utilisés dans la constitution du portefeuille agrégé.

C'est ainsi que le portefeuille 6 (détaillé en Tableau 1), s'est révélé comme particulièrement influençant sur la tendance observée de la mortalité. En effet, la seule suppression du portefeuille 6 dans le portefeuille agrégé conduit à une forme de mortalité a priori plus cohérente. Curieusement, il s'agit du seul portefeuille qui ait un tel impact sur le résultat : sans ce portefeuille, la mortalité ajustée fait bien état d'une décroissance en fonction de l'année calendaire, quelque soit l'âge.



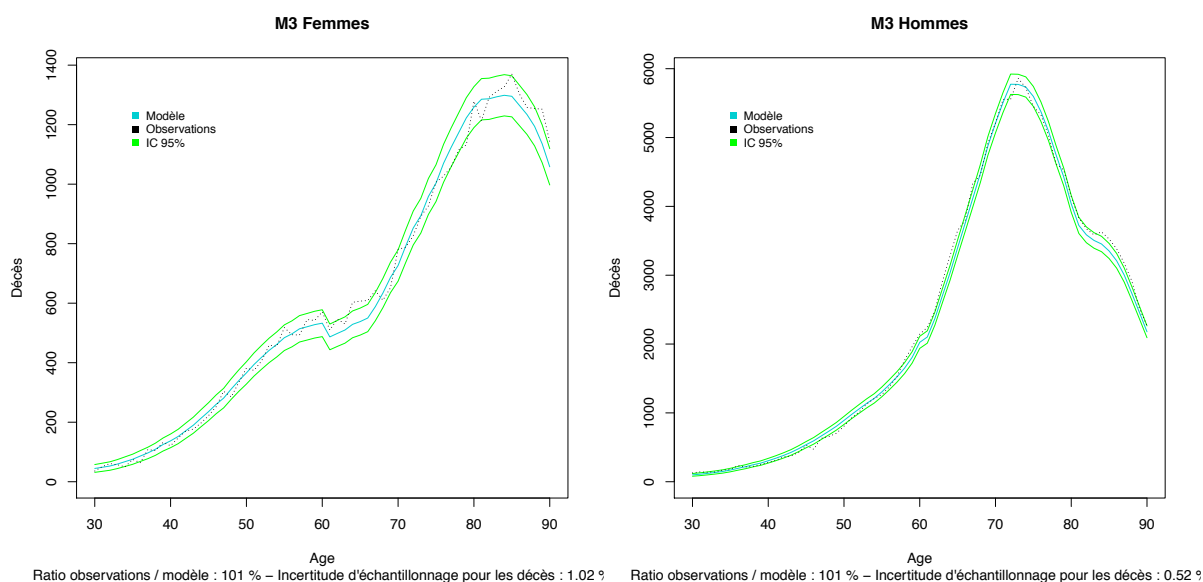
Nappe 7 – Logarithme des taux de mortalité ajustés du modèle GLM Poisson exploité sur 1996-2011- hors p6

En outre, la suppression du portefeuille 6 n'impacte pas uniquement les taux de mortalité ajustés mais également la table des moments brutes obtenues avec l'estimateur de Hoem et la régression de Cox. Le pic sur 2010, mis en évidence en 2^{ème} partie sur la Nappe 5 reflétant l'ensemble du portefeuille agrégé, disparaît avec la suppression du portefeuille 6.



Nappe 8 - Logarithme du taux de mortalité Femmes (f) et Hommes (h) du portefeuille agrégé – hors p6

Par ailleurs, l'adéquation du modèle aux données observées reste parfaitement acceptable avec au plus 1% d'incertitude tant pour les femmes que pour les hommes.



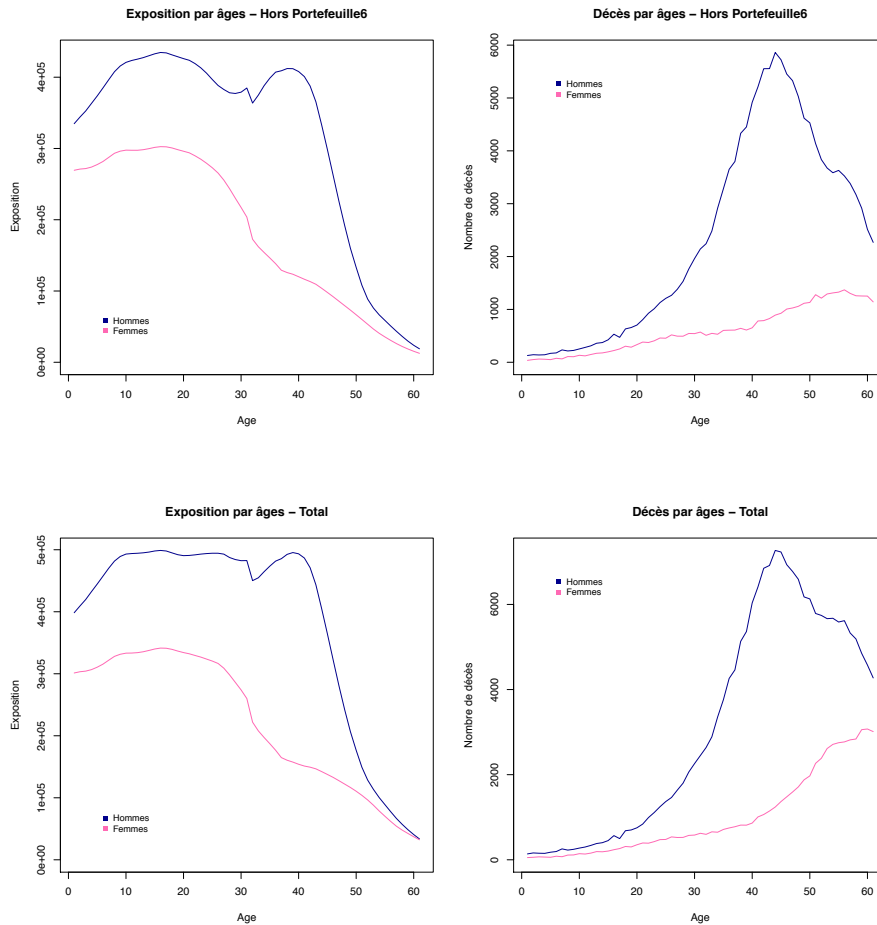
Graphique 19 - Comparaison des décès observés et ajustés et intervalle de confiance à 95 % - modèle M3 hors p6

En revanche, l'évaluation des coefficients du modèle log-linéaire démontre une significativité moins parfaite. Pour les hommes, les p-values restent bien inférieures à 0,05 % tandis que pour les femmes, les p-values relatives aux paramètres de l'âge (β_2) et de l'âge corrélé à l'année calendaire (β_4) dépassent les 0,05 %.

Femmes					Hommes				
	Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)
β_0	-38,6700	14,3800	-2,688	0,00718	β_0	-25,2400	8,4320	-2,994	2,75E-03
β_1	0,8611	0,0240	35,917	< 2e-16	β_1	0,6253	0,0232	26,986	< 2e-16
β_2	0,2621	0,1972	1,329	1,84E-01	β_2	0,2549	0,1164	2,191	2,85E-02
β_3	0,0180	0,0072	2,511	0,01202	β_3	0,0100	0,0042	2,373	1,77E-02
β_4	-0,0001	0,0001	-1,225	2,21E-01	β_4	-0,0001	0,0001	-1,795	7,27E-02

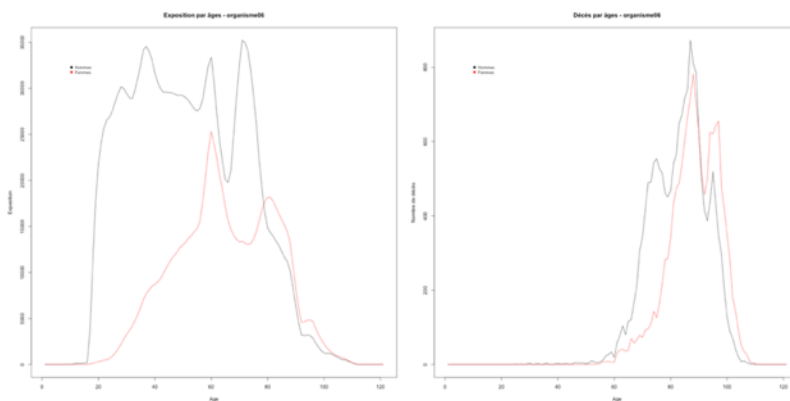
Tableau 5 - Résultats du modèle GLM Poisson exploité sur 1996-2011 hors p6

Cette baisse de significativité s'explique notamment par l'augmentation du risque d'échantillonnage lié à la suppression du portefeuille 6 : le nombre de décès des femmes est quasiment diminué de moitié, alors que le portefeuille 6 ne représente que 6% de l'exposition des femmes et par ailleurs 13% de l'exposition des hommes. La forme de l'exposition et du décès du portefeuille agrégé avec et sans le portefeuille 6 n'est en revanche pas modifiée.



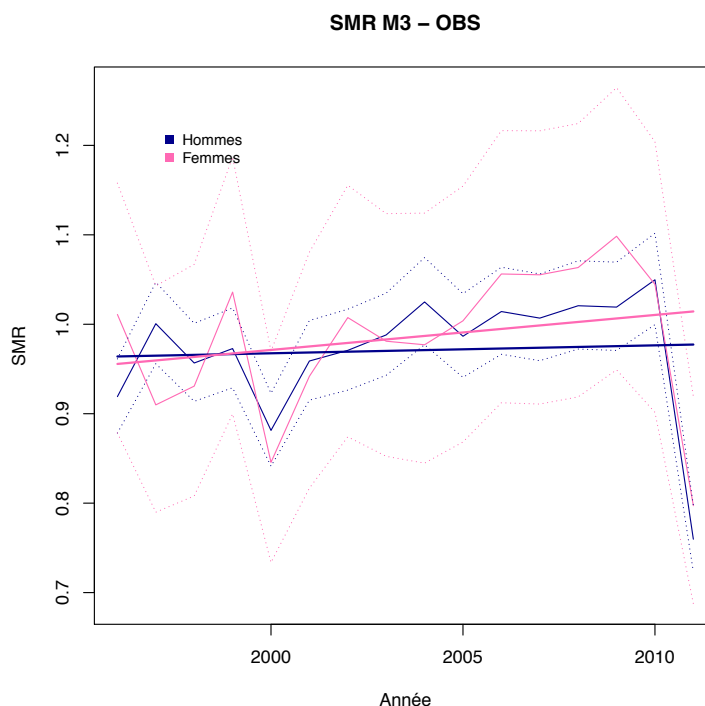
Graphique 20 – Comparaison de l'exposition et décès par âge du portefeuille agrégé avec (bas) et sans p6 (haut)

La distribution de décès cumulés sur la période est centrée sur des âges plutôt élevés, autour de 80 ans. Ceci s'explique notamment par le fait que le portefeuille 6 correspond à une garantie de type rente.



Graphique 21 – Exposition et décès par âge du portefeuille 6

La suppression du portefeuille 6, observé entre 2001 et 2010, permet au modèle GLM de bien mieux refléter la tendance propre du portefeuille ainsi agrégé, comme le montre la représentation du SMR ci-dessous. Les résultats relatifs aux femmes restent à mettre en parallèle du risque d'échantillonnage plus important constaté précédemment.

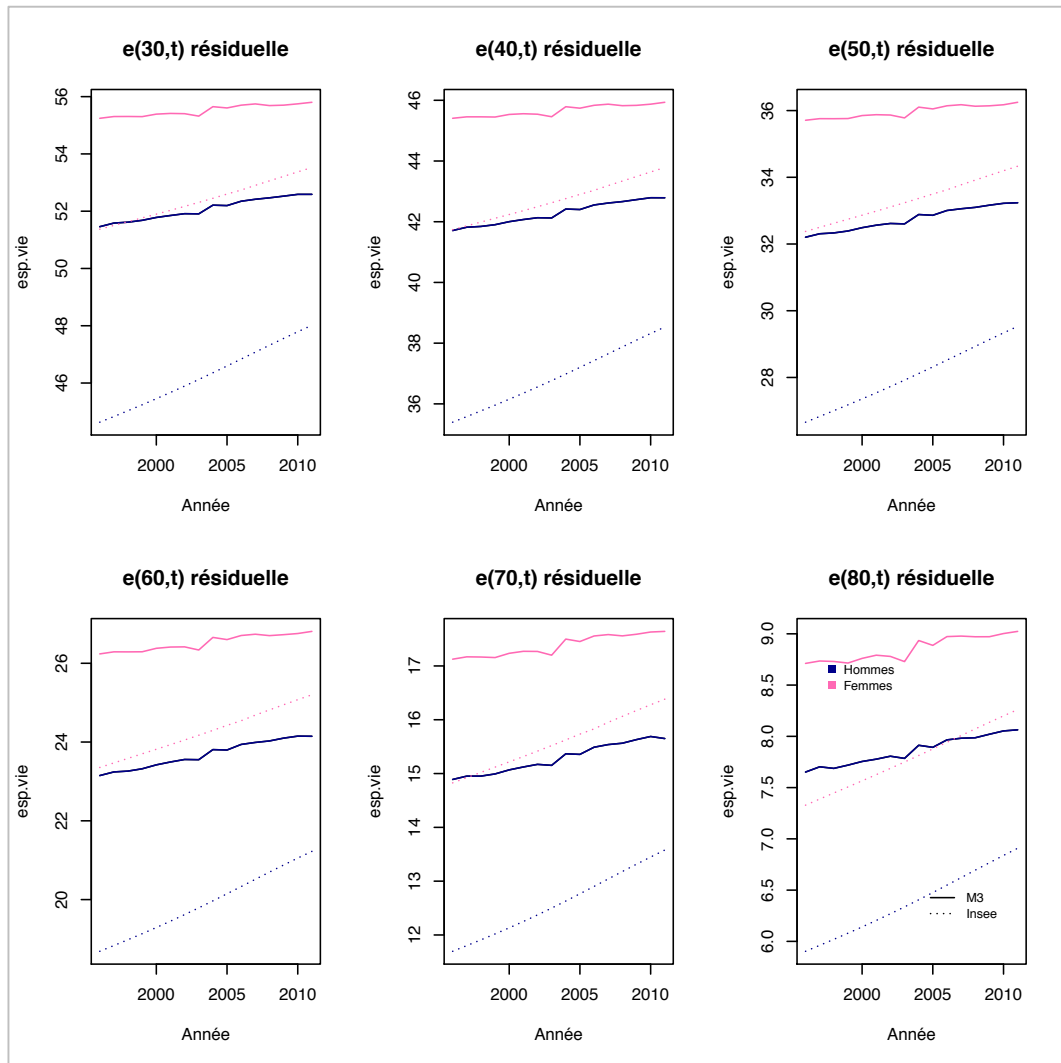


Graphique 22 – SMR des décès ajustés par le modèle GLM Poisson (M3) sur décès observés.

4- CONCLUSION

L'application du modèle GLM Poisson et l'étude d'impact sur les portefeuilles sous-jacents au portefeuille agrégé a permis de mettre en évidence une limite de la régression de Cox dans la perspective de supprimer l'hétérogénéité du portefeuille. En effet, l'extrême particularité du portefeuille 6 n'a pu être « gommée » par la seule application des coefficients de Cox.

Même sans le portefeuille 6, l'analyse des espérances de vie du modèle M3 fait état d'une tendance de longévité bien plus faible que celle de la population nationale. L'espérance de vie reste assurément supérieure pour les assurés mais l'écart diminue de manière prononcée sur la période 1996 à 2011 quel que soit l'âge d'évaluation. Par ailleurs, la mortalité des hommes semble converger plus fortement vers celle des femmes pour le modèle M3 comparé aux tables INSEE lissées, comme l'illustre le graphique ci-dessous.



Graphique 23 – Comparaison de l'espérance de vie résiduelle à différents âges, en fonction de l'année calendaire, pour le modèle GLM Poisson (M3) et la table INSEE lissée

Challengeons ces résultats par l'application d'un autre modèle endogène, un modèle de type Log Poisson, classique et analogue au modèle de Lee-Carter mais démontrant une plus large robustesse statistique.

B. TABLE DES MOMENTS AJUSTEE AVEC UN MODELE ENDOGENE DE TYPE LOG POISSON

Ce modèle sera par la suite appelé « modèle Log Poisson » ou modèle « LP ».

1- THEORIE

Ce modèle repose aussi sur le fait que les décès s'expriment selon une loi de Poisson, et s'appuie sur une forme de la force de mortalité $\mu_x(t)$ identique au modèle classique de Lee-Carter :

$$D_{x,t} \sim \mathcal{P}(E_{x,t} \cdot \mu_x(t)) \text{ avec } \mu_x(t) = \exp(a_x + b_x \cdot k_t)$$

Le paramètre a_x , s'interprète comme la valeur moyenne des $\ln(\mu_x(t))$, b_x traduit la sensibilité de la mortalité instantanée à l'âge x par rapport à l'évolution générale de k_t .

Sa vraisemblance s'écrit de la manière suivante :

$$\begin{aligned} \mathcal{L}(a,b,k) &= \sum_{x,t} [-E_{x,t} \cdot \mu_x(t) + D_{x,t} \cdot \log \mu_x(t)] + \text{constante} \\ &= \sum_{x,t} [D_{x,t} \cdot (a_x + b_x \cdot k_t) - E_{x,t} \cdot \exp(a_x + b_x \cdot k_t)] + \text{constante} \end{aligned}$$

La résolution d'un tel modèle passe par l'estimation du maximum de vraisemblance, pour lequel il n'existe pas d'expression analytique. Les estimateurs $(\hat{a}, \hat{b}, \hat{k})$ peuvent en revanche être évalués par itération, selon un principe de généralisation à deux dimensions (x,t) de la méthode unidimensionnelle de Newton (Brouhns et Al, 2002), à savoir en posant $\theta = (a, b, k)$:

$$\left\{ \begin{array}{l} \hat{\theta}^{(v+1)} = \hat{\theta}^{(v)} - \frac{\partial \mathcal{L}^{(v)} / \partial \theta}{\partial^2 \mathcal{L}^{(v)} / \partial \theta^2}, \quad \text{avec } \mathcal{L}^{(v)} = \mathcal{L}^{(v)}(\hat{\theta}^{(v)}) \text{ pour toute itération } v \\ \sum_t k_t = 0 \text{ et } \sum_x b_x = 1, \quad \text{pour garantir l'unicité des solutions} \end{array} \right.$$

En développant à l'ensemble des trois paramètres de notre modèle, le système d'itération s'écrit selon le schéma suivant :

$$\left\{ \begin{array}{l} \hat{a}_x^{(v+1)} = \hat{a}_x^{(v)} - \frac{\sum_t [D_{x,t} - \hat{D}_{x,t}^{(v)}]}{-\sum_t \hat{D}_{x,t}^{(v)}}, \quad \hat{b}_x^{(v+1)} = \hat{b}_x^{(v)}, \quad \hat{k}_t^{(v+1)} = \hat{k}_t^{(v)}, \\ \hat{k}_t^{(v+2)} = \hat{k}_t^{(v+1)} - \frac{\sum_x [(D_{x,t} - \hat{D}_{x,t}^{(v+1)}) \cdot \hat{b}_x^{(v+1)}]}{-\sum_x \hat{D}_{x,t}^{(v+1)} \cdot (\hat{b}_x^{(v+1)})^2}, \quad \hat{a}_x^{(v+2)} = \hat{a}_x^{(v+1)}, \quad \hat{b}_x^{(v+2)} = \hat{b}_x^{(v+1)}, \\ \hat{b}_x^{(v+3)} = \hat{b}_x^{(v+2)} - \frac{\sum_t [(D_{x,t} - \hat{D}_{x,t}^{(v+2)}) \cdot \hat{k}_t^{(v+2)}]}{-\sum_t \hat{D}_{x,t}^{(v+2)} \cdot (\hat{k}_t^{(v+2)})^2}, \quad \hat{a}_x^{(v+3)} = \hat{a}_x^{(v+2)}, \quad \hat{k}_t^{(v+3)} = \hat{k}_t^{(v+2)}, \end{array} \right.$$

$$\text{avec } \hat{D}_{x,t}^{(v)} = E_{x,t} \cdot \exp(\hat{a}_x^{(v)} + \hat{b}_x^{(v)} \cdot \hat{k}_t^{(v)})$$

2- RESULTATS POUR L'ENSEMBLE DU PORTEFEUILLE

Ce système d'équations a été développé sous R, en considérant l'itération 0 égale

$$\text{à } \begin{cases} \hat{a}_x^{(0)} = 0, \\ \hat{k}_t^{(0)} = 0, \text{ et selon un critère d'arrêt de } 10^{-6}. \\ \hat{b}_x^{(0)} = 1 \end{cases}$$

solutions a ensuite permis d'évaluer les paramètres d'estimation bruts finaux³⁰.

Tout comme le modèle de Lee-Carter, le lissage des paramètres d'estimation est usuel pour ce type de modèle. Nous avons donc appliqué un tel lissage, en considérant les variables a_x et b_x comme des fonctions quadratiques de l'âge et k_t , comme une fonction linéaire du temps. Les résultats graphiques des paramètres bruts et lissés pour les femmes et les hommes sont présentés ci-dessous (pour un âge compris entre 30 et 90 ans et une période d'observation 1996-2011).

³⁰ Entre 5672 (femmes) et 6308 (hommes) itérations ont été nécessaires pour un critère d'arrêt de 10⁻⁶. A noter qu'un critère d'arrêt plus faible n'a ici que peu d'impact sur la valeur des paramètres obtenus.

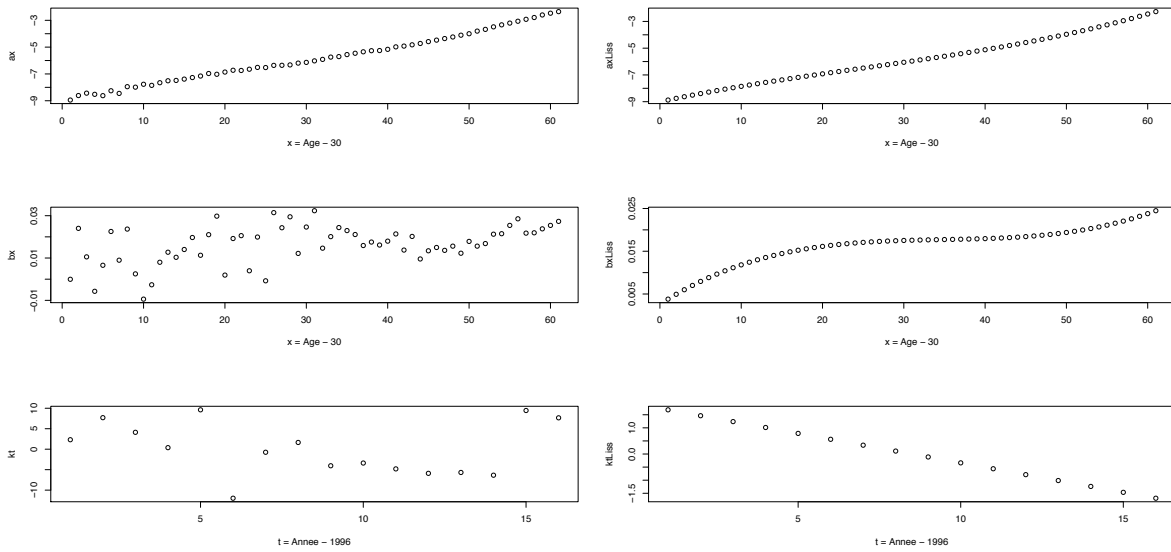


Figure 12 – Paramètres a_x , b_x et k_t du modèle Log Poisson brut (gauche) et lissé (droite) - Femmes

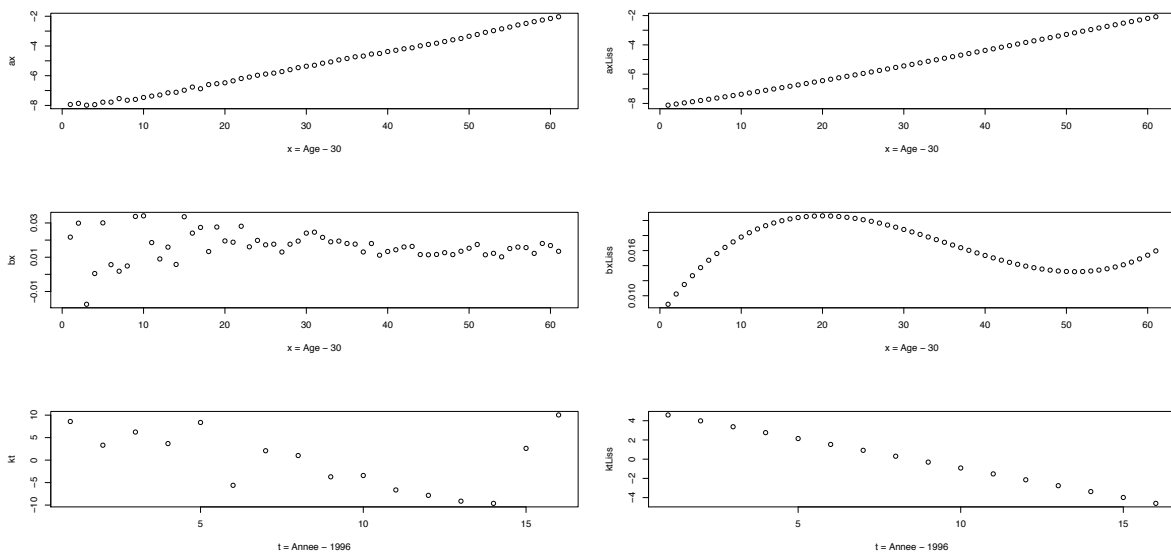
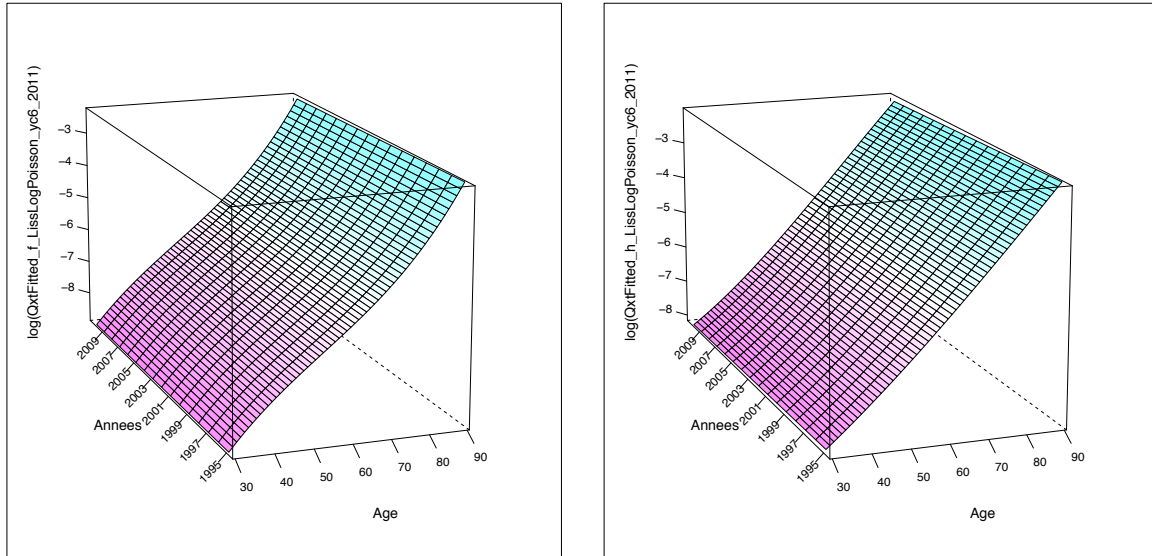


Figure 13 – Paramètres a_x , b_x et k_t du modèle Log Poisson brut (gauche) et lissé (droite) - Hommes

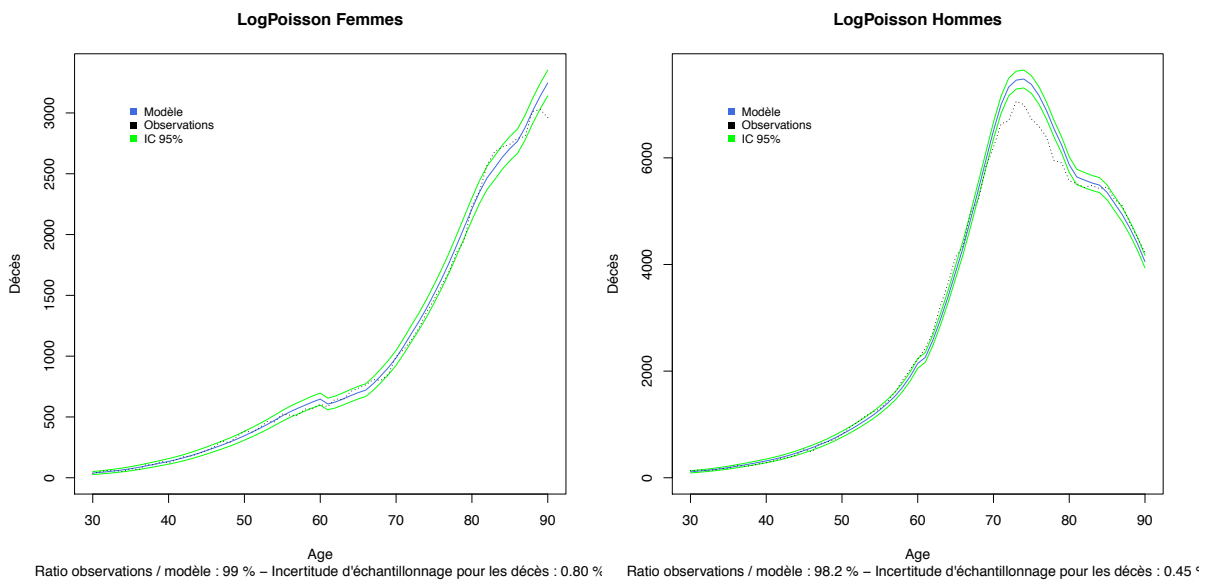
Les pics des années 2010 et 2011 sont particulièrement visibles sur le graphique des paramètres k_t bruts, hommes et femmes : pics qui disparaissent naturellement suite au lissage.

L'observation inattendue des taux de mortalité aux grands âges avec le modèle M3 prenant en compte l'ensemble du portefeuille n'est donc pas reproduite pour les nappes de mortalité du modèle LP lissé, mais la diminution des taux de mortalité aux grands âges avec l'année d'observation reste tout de même relativement faible.



Nappe 9 - Logarithme des taux de mortalité ajustés du modèle Log Poisson lissé exploité sur 1996-2011

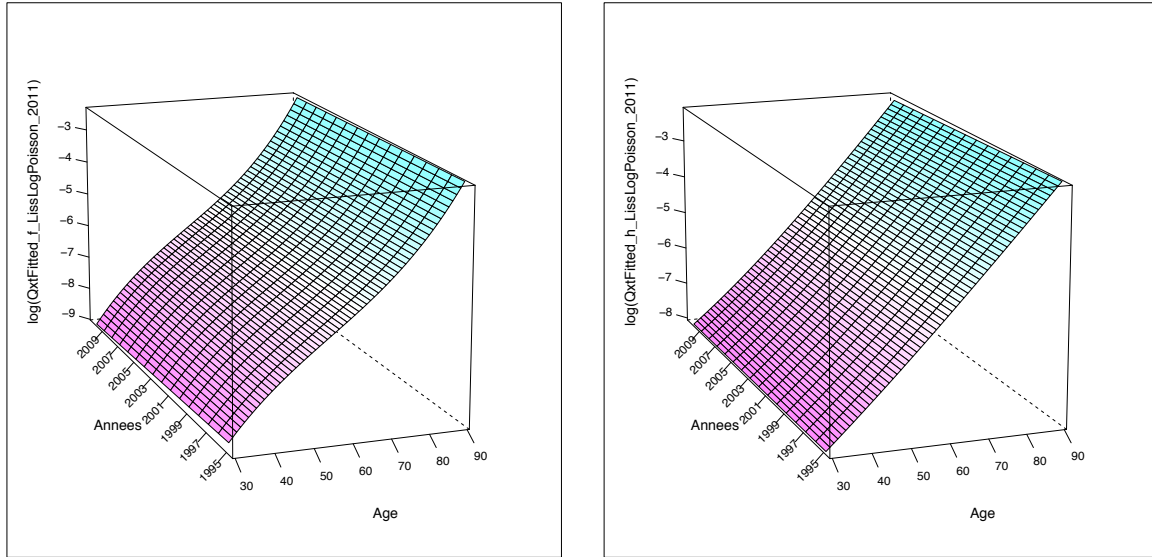
La comparaison des taux de mortalité observés/ajustés du modèle LP lissé illustre la surévaluation de la mortalité induite par le lissage, notamment pour les hommes entre 65 et 80 ans.



Graphique 24 - Comparaison des décès observés et ajustés et intervalle de confiance à 95 % - modèle LP

3- RESULTATS AJUSTES DE L'IMPACT DU PORTEFEUILLE 6

La représentation des nappes de mortalité obtenues selon un modèle LP lissé, à partir d'un portefeuille agrégé excluant le portefeuille 6 ne fait apparaître que très peu de différence avec celles initialement élaborées (Nappe 9).



Nappe 10 – Logarithme des taux de mortalité ajustés du modèle Log Poisson exploité sur 1996-2011- hors p6

L'impact de la suppression du portefeuille 6 est en revanche bien visible en vision brute de du paramètre temporel k_t , puisque le pic en année 10 disparaît.

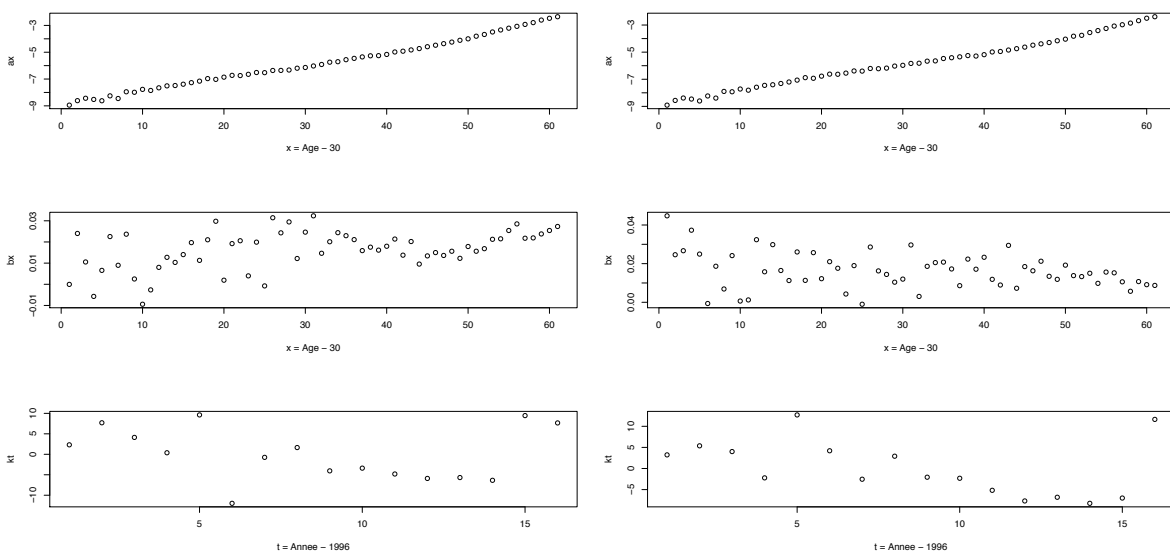


Figure 14 – Paramètres a_x , b_x et k_t du modèle LP brut élaboré sur l'ensemble du portefeuille (gauche) et en excluant le portefeuille 6 (droite) - Femmes

Elle l'est également après lissage à l'observation de la forme de la sensibilité lissée, b_x , tant pour les hommes que pour les femmes, qui devient alors concave aux âges élevés (à partir de 60 ans).

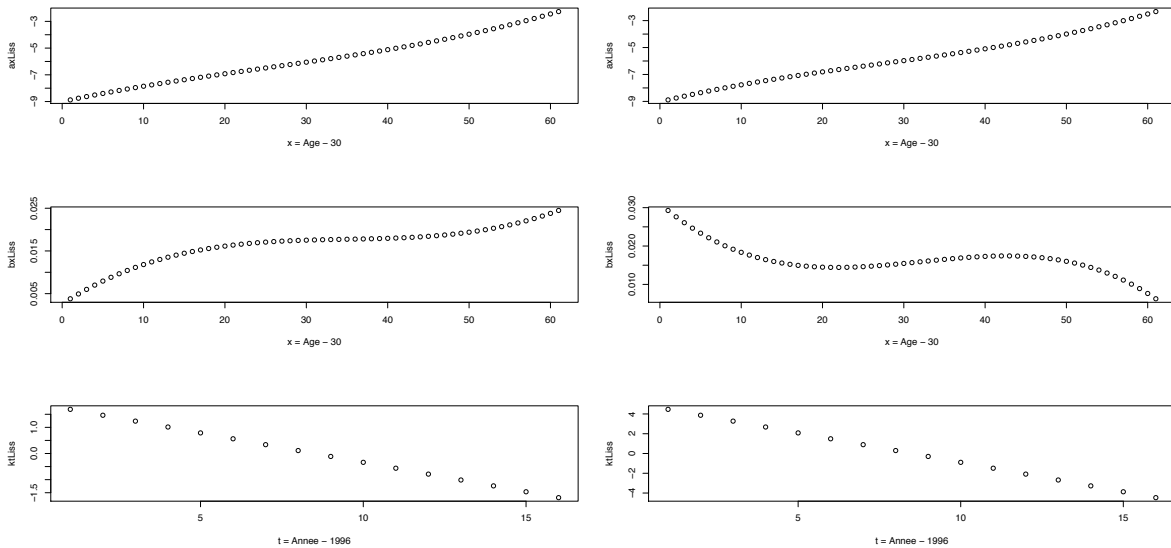


Figure 15 -- Paramètres a_x , b_x et k_t du modèle LP lissé élaboré sur l'ensemble du portefeuille (gauche) et en excluant le portefeuille 6 (droite) – Femmes

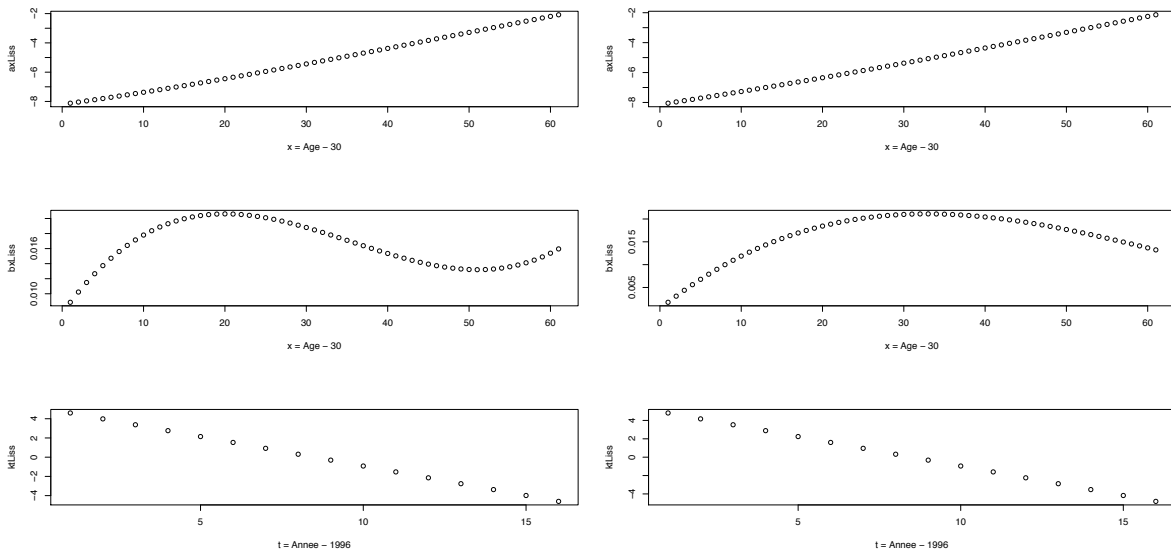


Figure 16 – Paramètres a_x , b_x et k_t du modèle LP lissé élaboré sur l'ensemble du portefeuille (gauche) et en excluant le portefeuille 6 (droite) - Hommes

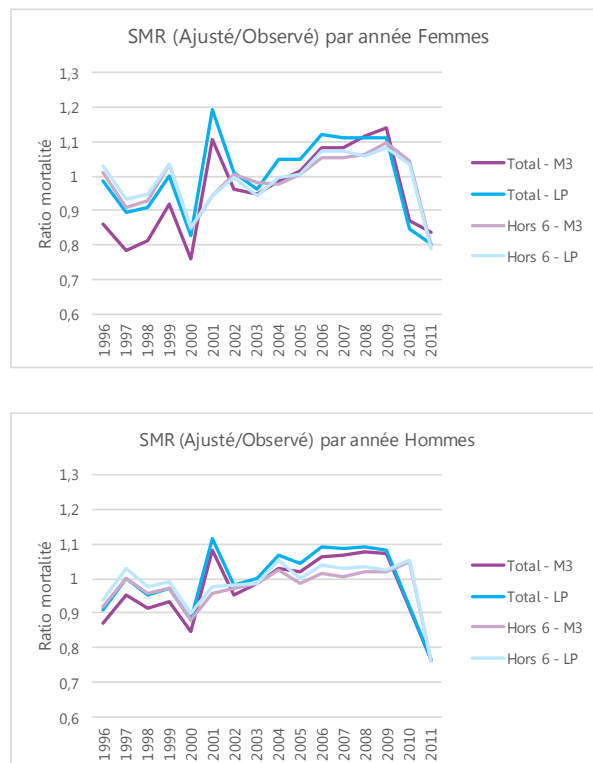
Dans la mesure où cette concavité aux âges élevés est la forme la plus usuellement observée, cela conduit à justifier à nouveau la suppression du portefeuille 6 pour la suite de l'étude de la tendance du portefeuille d'assurés.

4- COMPARAISON ET CONCLUSION

L'analyse des coefficients directeurs des ajustements linéaires des différents SMR (ajustés/observés) pour les deux modèles, Log Poisson (LP) et GLM Poisson (M3), et pour les deux groupements portefeuilles, fait également apparaître une meilleure représentativité à la tendance intrinsèque aux données lorsque le portefeuille 6 est exclu.

	Portefeuille total		Hors Portefeuille 6	
	GLM Poisson (M3)	Log Poisson (LP)	GLM Poisson (M3)	Log Poisson (LP)
Femmes	1,3%	0,3%	0,4%	0,2%
Hommes	0,4%	0,1%	0,1%	0,0%

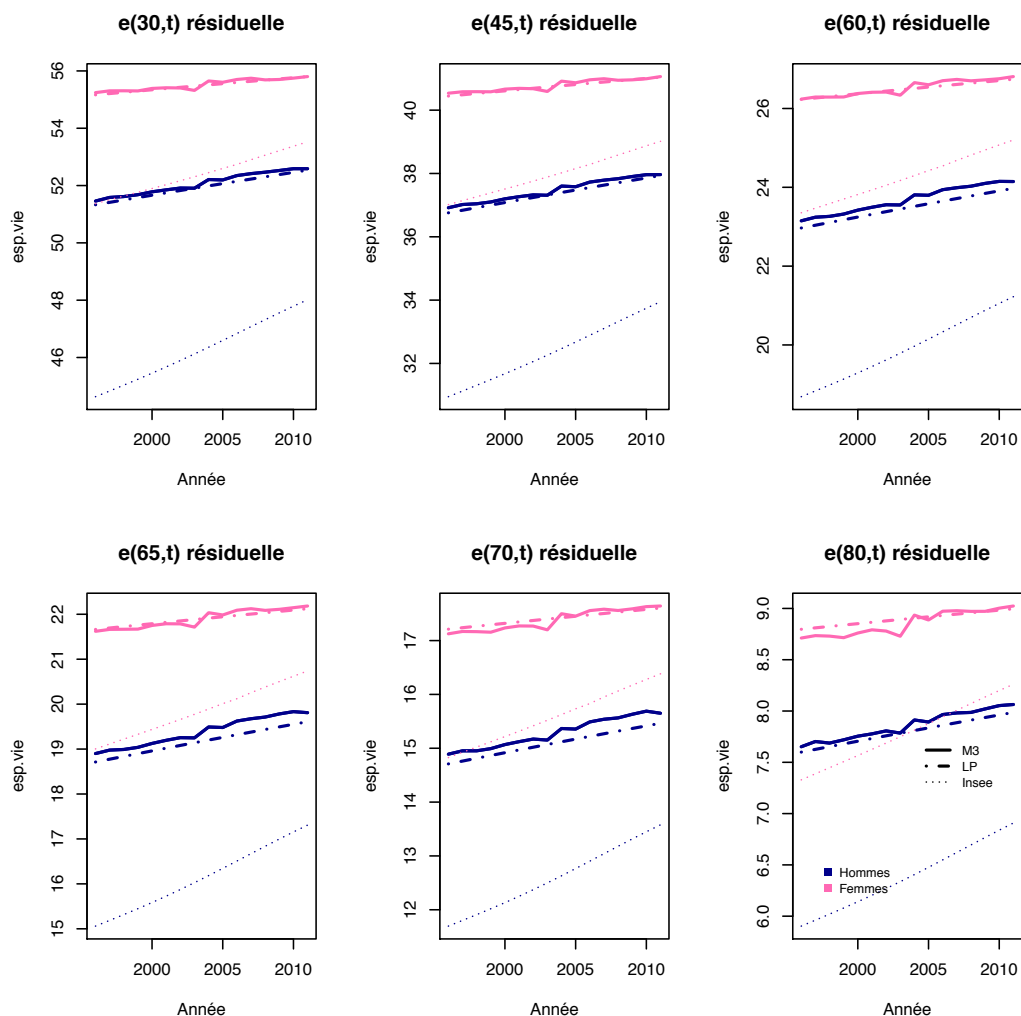
Tableau 6 - Coefficient directeur des ajustements linéaires des SMR des différents modèles



Graphique 25 – SMR des décès ajustés par les différents modèles sur décès observés sur le portefeuille global et hors portefeuille 6

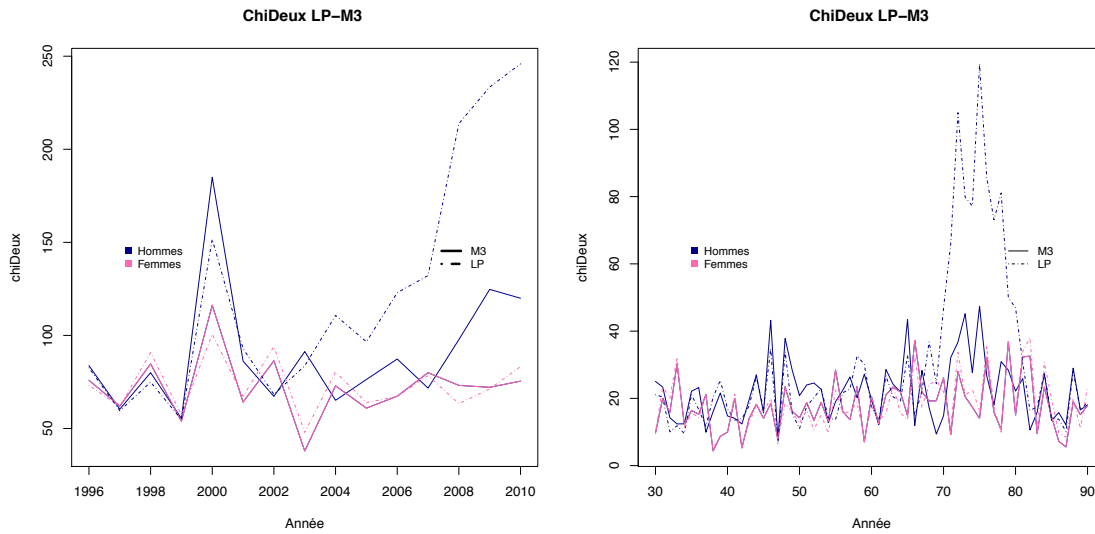
Dans le cas où le portefeuille 6 n'est pas intégré, cette même analyse ne révèle pas d'écarts significatifs entre les deux modèles GLM Poisson et Log Poisson.

La représentation des espérances de vie résiduelles à différents âges de ces deux modèles montre d'ailleurs une quasi égalité, et en tout cas un parallélisme certain de la tendance propre du portefeuille.



Graphique 26 – Comparaison de l'espérance de vie résiduelle à différents âges, en fonction de l'année calendaire pour les modèles LP, M3, hors p6 et la table Insee lissée

En revanche, l'étude du Chi-Deux en fonction de l'âge présente le GLM Poisson comme le meilleur modèle, notamment pour les âges 70 à 80 ans (âges influant majoritairement sur le pic de mortalité subsistant en 2010 sur les données brutes, et supprimé suite au lissage par le modèle Log-Poisson), mais aussi quel que soit l'année d'observation.



Graphique 27 – Comparaison des paramètres du Chi-Deux en fonction de l'âge ou de l'année calendaire des modèles LP et M3, hors p6

En conclusion, les deux modèles endogènes étudiés sont relativement comparables et ont permis, l'un comme l'autre, de mettre en évidence la problématique engendrée par le portefeuille 6, pour lequel la particularité est telle que le modèle de Cox n'est pas suffisant pour en gommer l'hétérogénéité par rapport aux autres portefeuilles étudiés.

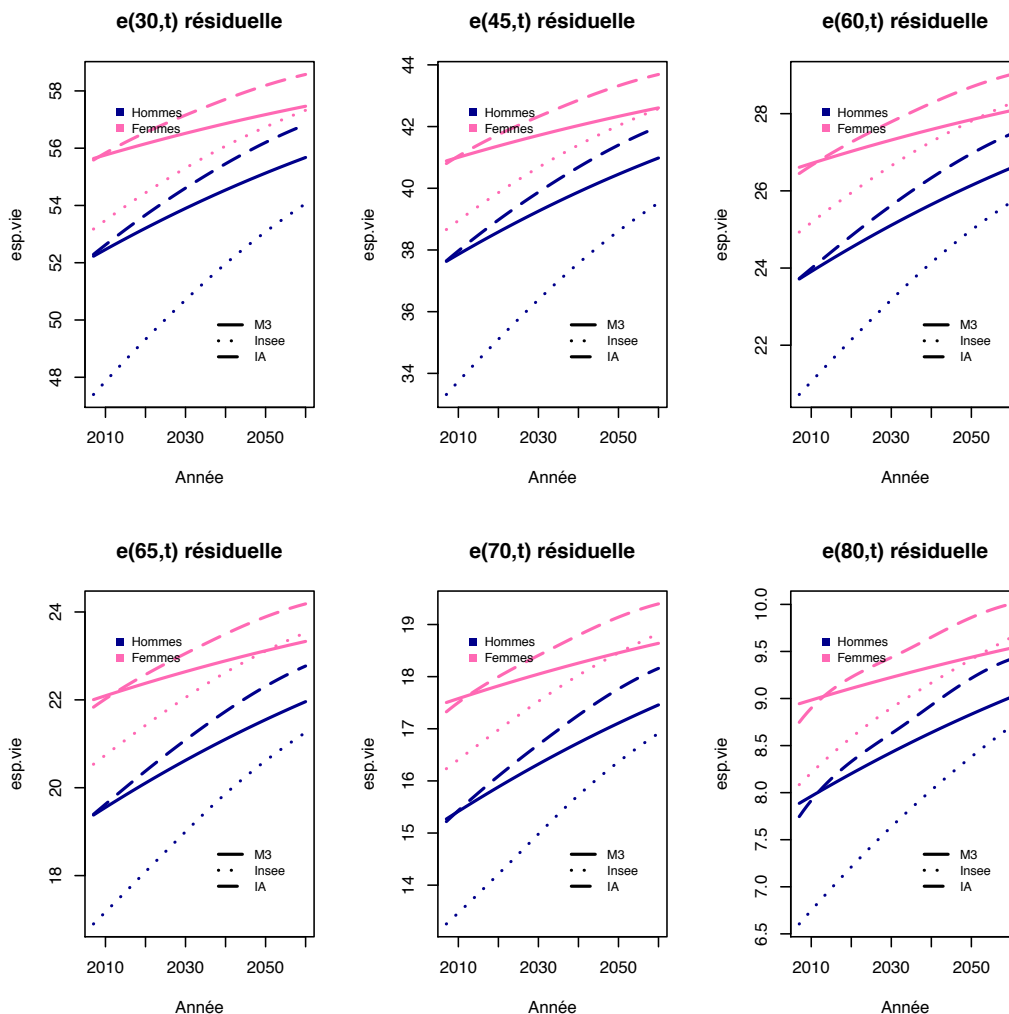
Pour la suite de l'étude, nous nous concentrerons sur les valeurs de mortalité ajustées avec le modèle GLM Poisson, ou M3, avec exclusion du portefeuille 6.

C. COMPARAISON DES TABLES PROSPECTIVES OBTENUES PAR RAPPORT AUX REFERENCES

La table prospective est obtenue par extrapolation jusqu'en 2060 des taux de mortalités ajustés par le modèle GLM Poisson (M3) pour les âges 30 à 90 ans, pour les femmes d'une part et pour les hommes d'autre part.

1- COMPARAISON DES TENDANCES

La comparaison des tables prospectives présente une tendance de longévité bien moins importante pour la population « assurés » par rapport à la population nationale (table INSEE, ou table IA).



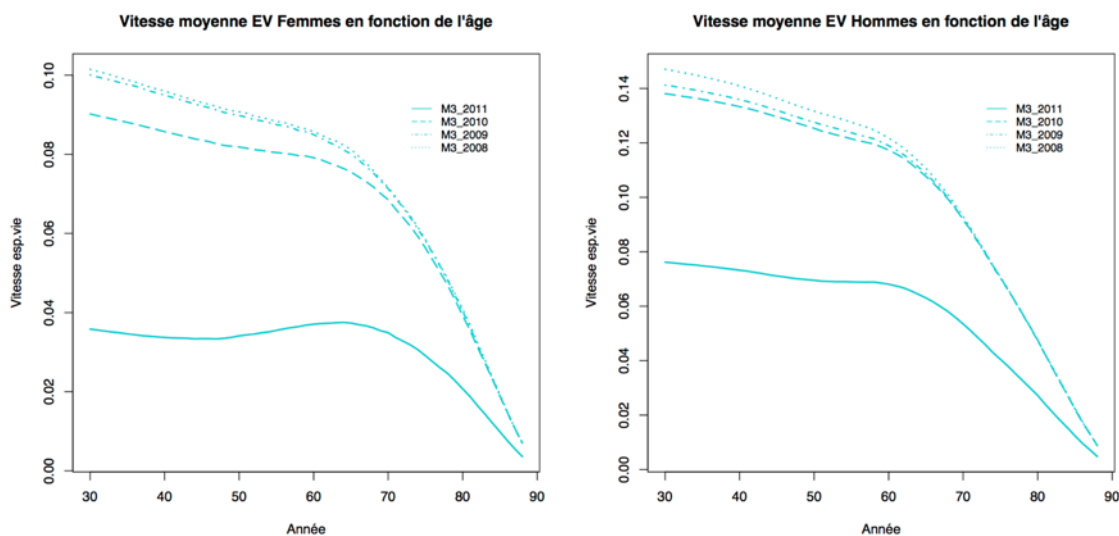
Graphique 28 – Comparaison de l'espérance de vie résiduelle à différents âges, en fonction de l'année calendaire pour le modèle GLM Poisson (M3) hors p6, la table INSEE lissée et la table IA

La quasi égalité des espérances de vie instantanées en 2010, aux différents âges aussi bien pour les hommes que pour les femmes, évaluée sur la table M3 et sur la table IA, confirme le caractère a priori pertinent du modèle M3 par rapport à la référence IA.

Cependant, avant de conclure quant à la tendance endogène évaluée, une dernière analyse de sensibilité a été faite selon la période d'observation considérée dans l'ajustement du modèle. En effet, la prise en compte du pic de mortalité de l'année 2011, persistant malgré la suppression du portefeuille 6, pourrait être particulièrement perturbant.

Dans cette perspective, un ajustement pratiqué selon la méthodologie précédemment détaillée a été élaboré avec le modèle GLM Poisson, pour les périodes d'observation restreintes [1996-2008], [1996-2009] et [1996-2010].

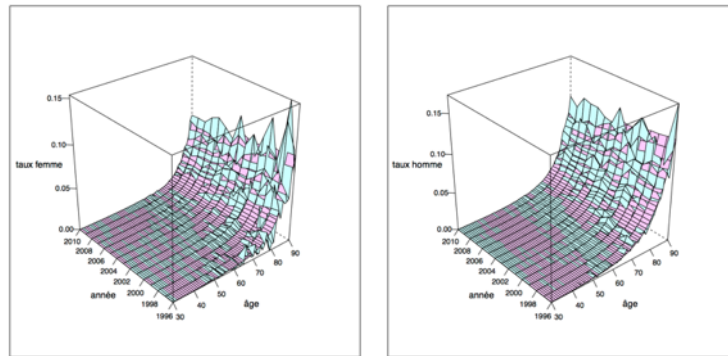
La comparaison de la vitesse des espérances de vie moyennées sur la période [1996-2008] pour chacun des ajustements signale une rupture nette de tendance dès lors que l'année 2011 est intégrée dans les données utilisées pour élaborer l'ajustement au modèle.



Graphique 29 – Vitesse moyenne sur la période [1996-2008] des espérances de vie instantanées en fonction de l'âge, obtenues pour les différentes périodes d'ajustement du modèle M3 hors p6

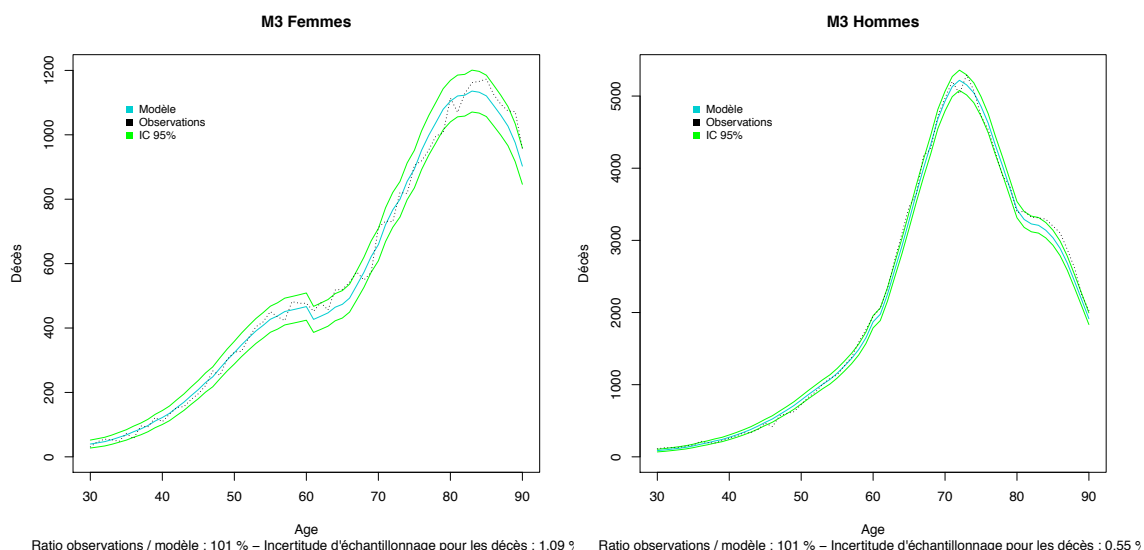
2- COMPARAISON DES TENDANCES SUR UNE PERIODE D'AJUSTEMENT RESTREINTE

La non prise en compte de l'année 2011 dans les données initiales supprime donc tous les pics de mortalités constatés (le pic de 2010 étant principalement lié au portefeuille 6, également non considéré) : le pic « vert » de données observées a disparu sur les surfaces de taux observés hommes et femmes présentées ci-dessous.



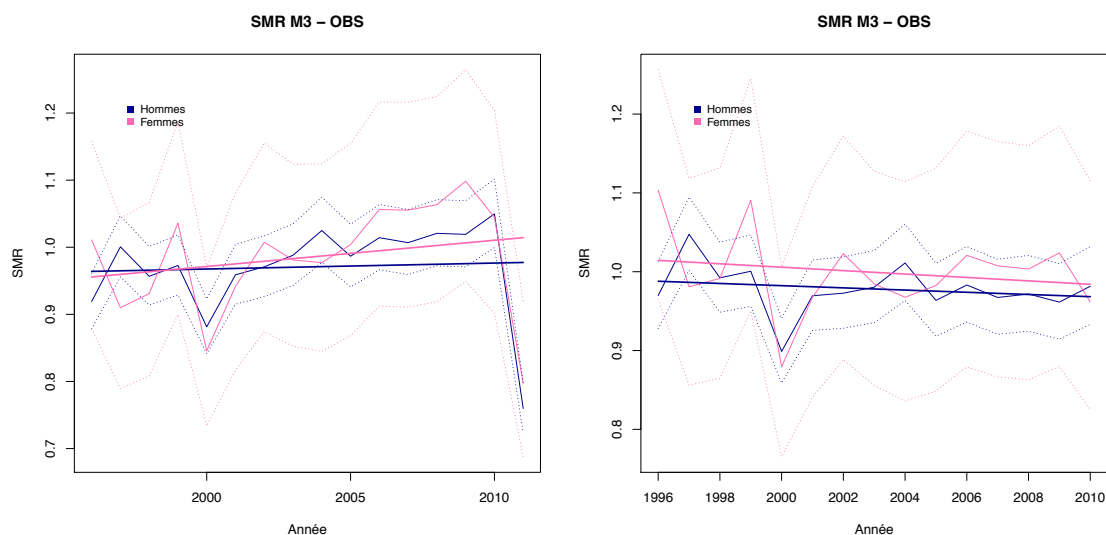
Surface 5 – Comparaison des taux de mortalités Observés (bleus) / Ajustés (roses) – modèle M3, 1996-2010 hors p6

La pertinence de l'ajustement du modèle aux données reste forte en supprimant l'année 2011. Nous pouvons remarquer que l'incertitude d'échantillonnage pour les femmes est légèrement supérieure à celle incluant l'année 2011, qui était alors de 1,02% (Graphique 16) : cela vient du fait que la suppression d'une année induit à nouveau une diminution du nombre d'observations de décès qui influe directement sur le risque d'échantillonnage.



Graphique 30 – Comparaison des décès observés et ajustés et intervalle de confiance à 95 % - modèle M3 1996-2010 hors 6

Les décès ajustés ont tendance à sous-estimer les décès observés dans le temps contrairement à ce qu'induisait la prise en compte de l'année 2011 et de son pic de mortalité.



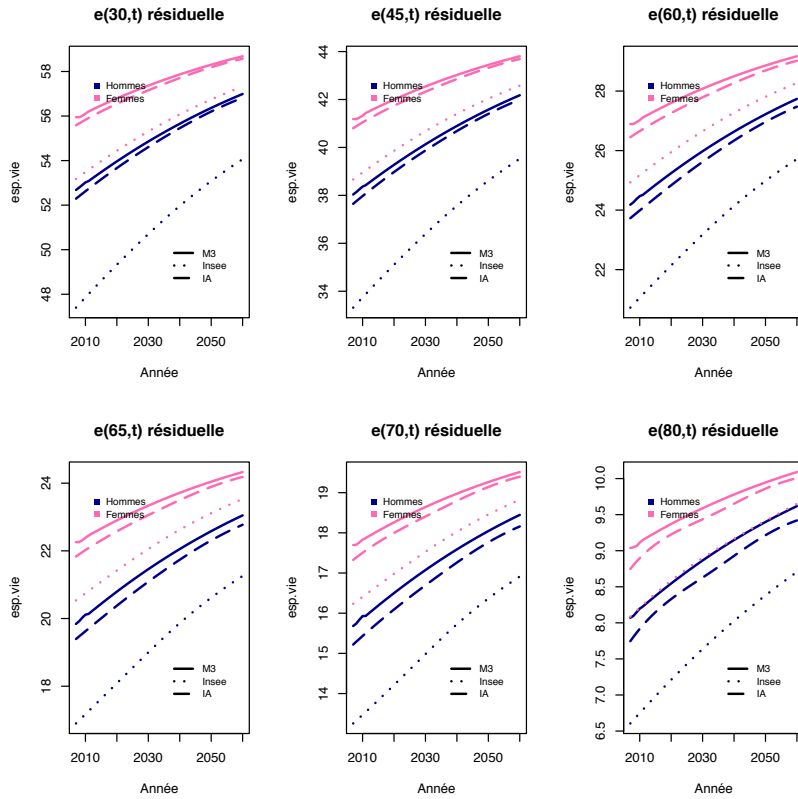
Graphique 31 – SMR des décès ajustés par le modèle GLM Poisson (M3) sur décès observés sur la période 1996-2011 (gauche) 1996-2010 (droite), hors p6

La comparaison des coefficients directeurs des régressions linéaires des SMR justifie d'une tendance de modèle équivalente en valeur absolue à celle des données observées quelque soit la période d'observation retenue pour l'ajustement.

Modèle M3 – hors p6	Période 1996-2010	Période 1996-2011
Femmes	-0,217%	0,391%
Hommes	-0,140%	0,089%

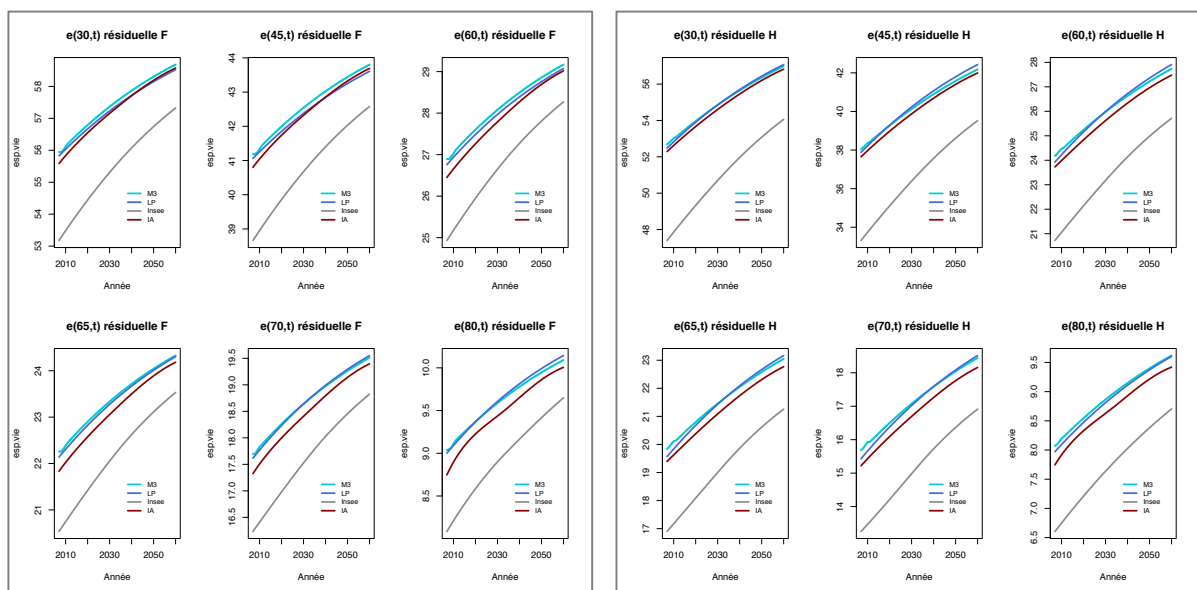
Tableau 7 - Coefficients directeurs des ajustements linéaires des SMR du modèle M3 hors portefeuille 6

Mais la suppression de l'année 2011 permet de construire une tendance endogène plus que comparable à celle de la population nationale, comme l'illustre le graphique des espérances de vie résiduelles à différents âges ci-dessous.



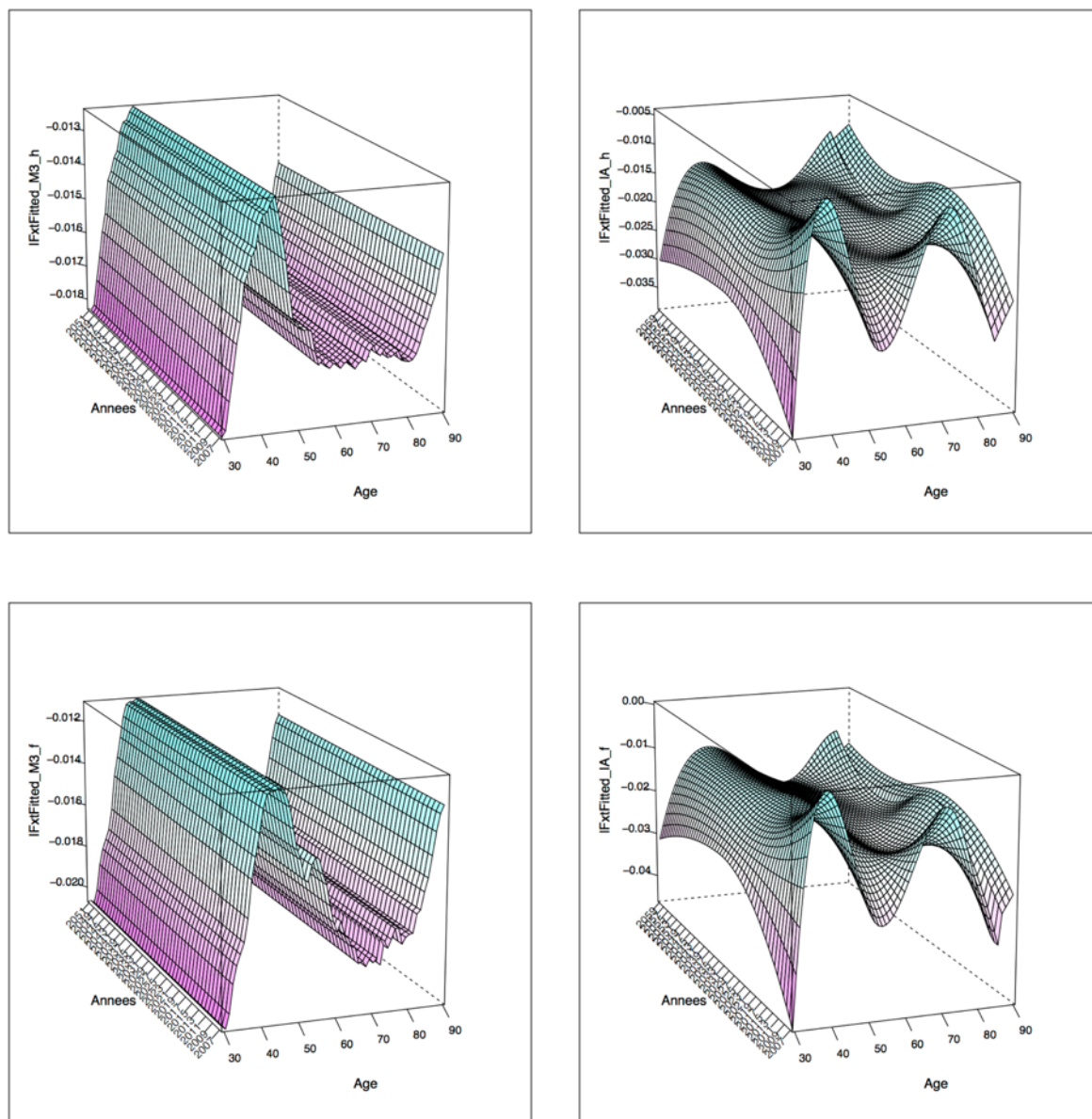
Graphique 32 – Comparaison de l'espérance de vie résiduelle à différents âges en fonction de l'année calendaire pour le modèle GLM Poisson (M3) ajusté sur 1996-2010 hors p6, la table Insee lissée et la table IA

La nouvelle tendance endogène est par ailleurs bien confirmée avec le modèle Log Poisson ajusté et lissé sur la même période 1996-2010.



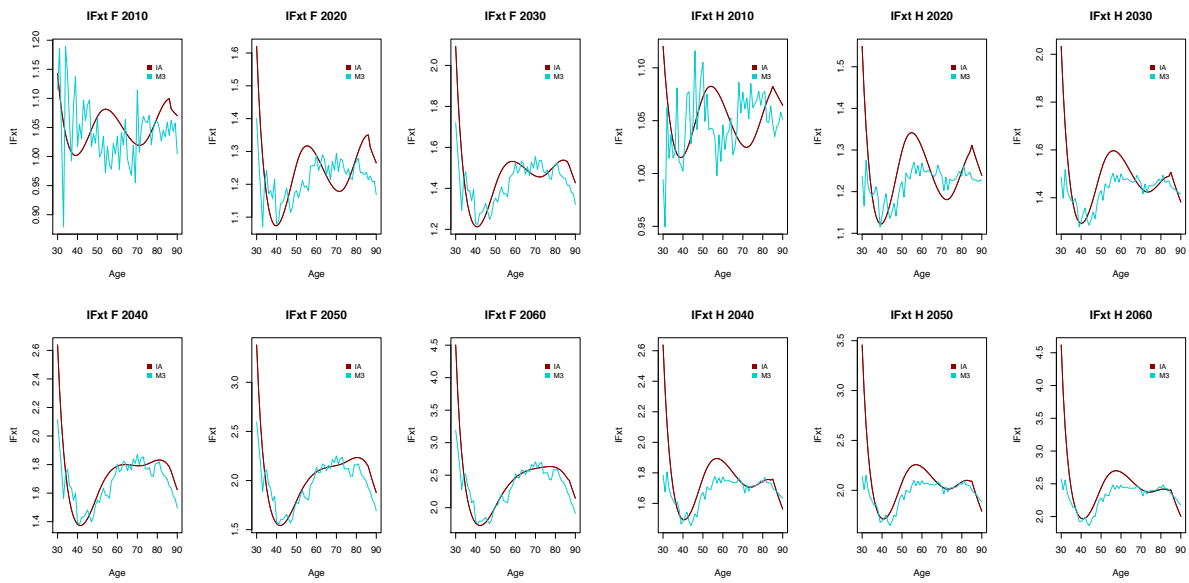
Graphique 33 - Comparaison de l'espérance de vie résiduelle à différents âges en fonction de l'année calendaire pour les modèles M3 et LP ajustés sur 1996-2010, la table Insee lissée et la table IA – femmes/hommes

Les surfaces des coefficients d'amélioration (IF) de la table ajustée avec le modèle M3 et de la table IA, malgré une apparence différente, sont, rapportées à échelle d'ordonnée équivalente, plutôt semblables.

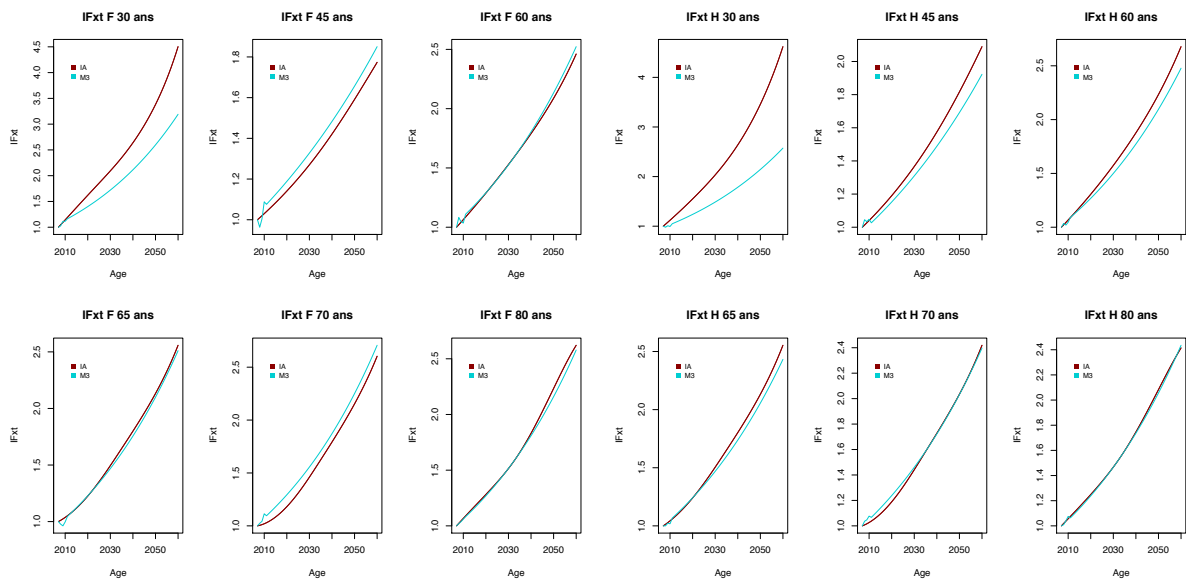


*Surface 6 – Facteurs d'amélioration des taux de mortalité Femmes / Hommes
Tables IA (droite) et Tables modèle M3 (gauche)*

La représentation de ces mêmes facteurs d'amélioration en fonction de l'âge et en fonction de l'année d'observation atteste de cette évolution comparable.



Graphique 34 – Comparaison des facteurs d'amélioration en fonction de l'âge pour différentes années d'observation pour le modèle GLM Poisson (M3) ajusté sur 1996-2010 hors p6, et la table IA



Graphique 35 – Comparaison des facteurs d'amélioration en fonction de l'année d'observation pour différents âges pour le modèle GLM Poisson (M3) ajusté sur 1996-2010 hors p6, et la table IA

Il est à noter que dans le cas où nous aurions conservé le portefeuille 6 pour l'évaluation endogène de la tendance, il aurait fallu supprimer les années 2010 et 2011 pour retrouver une tendance comparable à celle de la population nationale.

CONCLUSION PARTIE 3

L'analyse menée au cours de cette troisième partie permet d'attester à l'appui de deux modèles endogènes distincts la forte hétérogénéité que peut engendrer un portefeuille de données (le portefeuille 6) ainsi que l'importante sensibilité que peut induire une année de période d'ajustement (l'année 2011) dans l'évaluation de la tendance du portefeuille agrégé.

Une fois mis de côté la problématique du portefeuille 6, qui pourrait être approfondie par le biais de méthodes alternatives à la régression de Cox pour supprimer l'hétérogénéité, chacun des deux modèles montre que pour un ajustement sur la période 1996-2011, la tendance propre du portefeuille « assurés » est très inférieure à celle du portefeuille de référence de la population nationale. Avec la suppression de l'année 2011, population « assurés » et population nationale se rejoignent sur une même tendance.

Faut-il y voir une rupture de tendance en 2011 ? Le portefeuille « assurés » est-il en avance de phase par rapport à l'année 2015 de la population nationale qui coïncide avec la 1^{ère} année de constatation de baisse de la longévité³¹, qui pourrait être précurseur d'une nouvelle tendance. Ou bien s'agit-il d'un simple effet d'échantillonnage sur le portefeuille étudié, à l'image du portefeuille 6 ?

Il pourrait être intéressant de mettre en application des méthodes de détection de rupture de tendance, comme l'utilisation d'un processus de Poisson (El Karoui, 2015), ou à défaut de reprendre cette étude à l'appui des données complémentaires d'assurés de 2012 à 2015.

³¹ Baisse de 3,5 mois pour les hommes et les femmes de l'espérance de vie à la naissance, notamment lié à l'épidémie de grippe de 2015 – source INSEE

CONCLUSION

L'objectif poursuivi était de pouvoir challenger la pertinence de la tendance des tables de références assurantielles quant à une approche au plus juste ou *best estimate* du calcul des provisions. L'étude a été rendue possible par la mise à disposition d'une large base de données regroupant les informations de décès des assurés de 15 organismes de place. Afin de préserver une profondeur d'historique aussi grande que possible, essentielle pour une évaluation endogène de la tendance de la mortalité, la méthode de Cox a été appliquée pour gommer l'hétérogénéité intrinsèque au cumul des différentes sous-populations. Un portefeuille agrégé a ainsi été constitué en ajustant l'exposition de chaque sous-population à l'aide des coefficients résultant de la régression de Cox, appliquée pour le critère cumulé « Organisme + Plage d'observation ».

L'extrapolation de la mortalité du portefeuille total agrégé fait état d'une vitesse d'espérance de vie bien moins importante que celle constatée sur la population nationale, quel que soit le modèle utilisé (GLM Poisson ou Log-Poisson), et ajusté sur la période 1996-2011. Il serait cependant à ce stade prématuré de conclure que le provisionnement pratiqué par les assureurs aujourd'hui pour les garanties en cas de vie, est particulièrement prudent.

En effet, les études de sensibilité qui ont été pratiquées en troisième partie reflètent des biais dans l'estimation de cette tendance : un premier lié aux données ou à l'agrégation des sous-populations, et un second issu de la période d'observation utilisée pour l'ajustement des modèles. En effet, la suppression d'un portefeuille précis (le portefeuille 6) et de l'année d'observation 2011 a pour conséquence de rendre quasiment parfaitement équivalentes les tendances de la population d'assurés étudiée et de la population nationale !

La sensibilité observée à la **suppression du portefeuille 6** met en évidence une limite à l'exploitation du modèle de Cox dans l'atténuation de l'hétérogénéité. Si la régression appliquée a sans doute permis de prendre en compte les inégalités liées aux différentes périodes d'observation des organismes sources, elle n'a pas intégré les spécificités sous-jacentes à la typologie des garanties. Il serait ainsi intéressant d'envisager une agrégation des portefeuilles ajustée à l'aide d'un critère cumulé de type : « Organisme + Plage d'Observation + Garantie ». Cependant, ce raisonnement n'a finalement pas été ici appliqué dans la mesure où une partie des sous-populations ainsi construites aurait présenté un nombre insuffisant de données pour rendre pertinente l'application de la régression de Cox.

La richesse des données demeure assurément l'ingrédient majeur pour la réalisation d'études poussées sur la mortalité. Il est alors impossible de ne pas imaginer ce monde parfait du point de vue des actuaires et des chercheurs où les données des assurés seraient regroupées dans une base commune de place...

Cette problématique de données existe de manière plus opérationnelle dans le domaine de la couverture du risque de longévité. Bien que le choix a été fait de ne pas détailler ce sujet, l'étude ici menée dans le traitement de l'hétérogénéité d'un groupement de portefeuilles pourrait avoir son intérêt dans un tel cadre. En effet, aujourd'hui, la couverture de la tendance de la longévité, à travers des swaps indiciels de longévité notamment, ne bénéficie qu'aux portefeuilles de rentes suffisamment conséquents en montant³², au grand regret des régimes de rentes plus « confidentiels ». Regrouper ces derniers, sous forme d'une sorte de « pool » de rentes, serait une solution possible : les paramètres de la régression de Cox pourraient ainsi être exploités pour tarifier et répartir la prime du swap parmi les différents régimes concernés. Ces pratiques de mutualisation commencent à s'entrevoir en réassurance, et une récente transaction innovante au Canada de rachat de rentes par un assureur auprès de deux régimes différents³³, confirme cette perspective nouvelle.

Par ailleurs, la forte rupture observée dans les vitesses d'espérance de vie lors de **la suppression de l'année 2011** remet en cause l'intégration de cette année au sein de la période d'ajustement des modèles. Plusieurs arguments peuvent être avancés pour expliquer une telle sensibilité. Tout d'abord, un biais pourrait persister dans l'application de la régression de Cox : l'ajustement de l'agrégation avec un critère plus précis autour du type de garantie, comme explicité précédemment, serait une première piste. Le principe de l'existence d'une volatilité intrinsèque de la mortalité pourrait également être envisagé et ceci d'autant plus que selon l'INSEE la courbe de l'espérance de vie devrait désormais évoluer en dents de scie tout en restant orientée à la hausse. La population des assurés étudiée pourrait donc au final être en avance de phase et présenter les premiers signes de cette volatilité intrinsèque. Enfin, une rupture avérée de tendance au sein de la population des assurés pourrait en dernière hypothèse être considérée comme une piste envisageable.

³² Axa est aujourd'hui le seul assureur en France à bénéficier d'un swap de longévité avec un portefeuille de rentes de 750M€ pour 22 000 assurés auprès de Hannover Re.

³³ Montage réalisé en janvier 2016, facilité par Mercer qui a plus récemment créé une Bourse d'achat de rente faisant le lien entre assureurs et promoteurs de régimes de retraite à prestations définies au Canada.

En définitive, la pertinence de l'utilisation de la tendance de la mortalité nationale comme référence pour celle des assurés, n'a pas été ici remise en cause. Cependant, et compte tenu des réserves émises, il serait particulièrement intéressant de mettre à jour cette étude à l'appui de données, présentant non seulement un historique plus récent, mais regroupant également une palette plus large d'organismes de place, dans l'idée de tendre progressivement vers la construction de la base parfaite. Il nous faudra probablement faire preuve de patience pour envisager une estimation plus juste. Mais cela n'est pas cause perdue à en croire un célèbre dicton Wolof, dans la mesure où cette patience pourrait nous mener à une autre forme de longévité... « Au bout de la patience il y a le ciel ». Celle de l'esprit !

BIBLIOGRAPHIE

- [1] Barrieu P. et al. [2010] "Understanding, Modelling and Managing Risk Longevity: Key Issues and Main Challenge", *Scandinavian Actuarial Journal*, vol 2012, Issue3.
- [2] Blanpin N., Chardon O. [2010] "Projections de population 2007-2060 pour la France métropolitaine : méthode et principaux résultats", *Direction des Statistiques Démographiques et sociales - N°F1008*.
- [3] Bongaarts J. [2005] "Long-range trends in adult mortality: models and projection methods", *Demography*, 42(1), 23-49.
- [4] Box G., Jenkins G.M., Reinsel G.C. [1994] "Time series analysis: forecasting and control (3rd ed.)", Prentice Hall, Englewood, NJ.
- [5] Brass W. [1971] "On the scale of mortality", In W. Brass (ed.), editor, *Biological aspects of demography*. London : Taylor & Francis.
- [6] Brouhns N., Denuit M. [2001] "Risque de longévité et rentes viagères. Tables de mortalité prospectives pour la population belge", *Discussion Paper*, Institut de Statistique, Université catholique de Louvain, Louvain-la- Neuve, Belgique.
- [7] Brouhns N., Denuit M., Vermunt J.K. [2002] "A Poisson log-bilinear regression approach to the construction of projected lifetables", *Insurance: Mathematics and Economics*, vol. 31, 373-393.
- [8] Cairns A.J.G., Blake D., Dowd K., Coughlan G., Epstein D., and Khallaf-Allah M. [2008] "The plausibility of mortality density forecasts: an analysis of six stochastic mortality models", *Working paper*, Heriot-Watt University, and Pensions Institute Discussion Paper PI-0801.
- [9] Currie I. D., Durban M., and Eilers P. H. C. [2006] "Generalized linear array models with applications to multidimensional smoothing". *Journal of the Royal Statistical Society*, 68(Part 2), 259–280.
- [10] Currie I.D., Durban M. and Eilers P.H.C. [2004] "Smoothing and forecasting mortality rates", *Statistical Modelling*, 4 : 279-298.
- [11] Debonneuil E. [2014] "Modèle paramétrique de mortalité en fonction de l'âge pour des applications sur un portefeuille de retraite", *mémoire d'actuariat*.
- [12] Delwarde A., Kachkhdze D., Olie L., and Denuit M. [2004] "Modèles linéaires et additifs généralisés maximum de vraisemblance local et méthodes relationnelles en assurance sur la vie", *Bulletin Français d'Actuariat*, 6(12), 77–102.
- [13] Denuit M. and Goderniaux A. C. [2005] "Closing and projecting life tables using log-linear models", *Bulletin of the Swiss Association of Actuarial Science*, (1), 29–48.
- [14] El Karoui N., Loisel S., Salhi Y. [2015] "Minimax Optimality in Robust Detection of a Disorder Time in Poisson Rate", *prépublication*.
- [15] Gschlössl S., Schoenmaekers P., and Denuit M. [2011] "Risk classification in life insurance: methodology and case study". *European Actuarial Journal*, 1(1), 23–41.
- [16] Juillard M., Planchet F., Thérond P.E. [2008] "Perturbations extrêmes sur la dérive de mortalité anticipée", *Assurances et gestion des risques*, Vol. 76 (3).

- [17] Lee R.D., Carter L. [1992] "Modelling and forecasting the time series of US mortality", *Journal of the American Statistical Association*, vol. 87, 659–671.
- [18] Lelieur V. [2005] "Construction de tables de mortalité prospectives : le cas des petites populations", *Mémoire d'actuariat ISFA*.
- [19] Oeppen J. and Vaupel J. W. [2002] "Broken limits to life expectancy". *Science*, 296(5570), 1029–1031.
- [20] Planchet F. and 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*, 7(14), 118–146.
- [21] Planchet F., Lelieur V., [2010] "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*, 118-146.
- [22] Planchet, F. and Théron, P. [2011] "Modélisation statistique des phénomènes de durée - Applications actuarielles". *Assurance Audit Actuariat. Economica Paris*.
- [23] Renshaw A.E., Haberman S. [2006] "A cohort-based extension to the Lee- Carter model for mortality reduction factors", *Insurance: Mathematics and Economics*, 38: 556-570.
- [24] Tomas J. et Planchet F. [2011] "Construction et validation des références de mortalité de place", *Institut des Actuaire*, II 1291-11 v1.4.
- [25] Tomas J. et Planchet F. [2011] "Méthodes de positionnement : aspects méthodologiques", *Institut des Actuaire*, II 1291-12 v1.7.
- [26] Tomas J. et Planchet F. [2013] "Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance", *Insurance: Mathematics and Economics*, 52(3), 573–589.
- [27] Tomas J. et Planchet F. [2013] "Prospective mortality tables : taking heterogeneity into account", *Cahiers de Recherche de l'ISFA*, 2013(5), 1–33.

ANNEXES

A – RECAPITULATIF DES RETRAITEMENTS REALISES SUR LA BASE DE DONNEES

Les retraitements réalisés ont donné lieu à la suppression de 490 755 lignes réparties comme suit selon les organismes :

	Lignes initiales	Lignes post retraitement	Lignes supprimées
organisme01	763 713	747 632	16 081
organisme02	6 302	5 547	755
organisme03	89 914	89 747	167
organisme04	197 097	191 284	5 813
organisme05	26 177	25 974	203
organisme06	971 090	832 298	138 792
organisme07	261 765	261 732	33
organisme08	215 651	215 651	-
organisme09	1 336 967	1 271 017	65 950
organisme10	64 687	64 687	-
organisme11	798 522	754 380	44 142
organisme12	1 073 394	863 040	210 354
organisme13	1 445 871	1 438 920	6 951
organisme14	735 463	735 045	418
organisme15	215 152	214 056	1 096
Total	8 201 765	7 711 010	490 755

Les retraitements réalisés sont les suivants :

- Suppression des lignes pour lesquelles l'âge d'entrée est négatif ou nul
 - 5585 lignes
- Suppression des lignes pour lesquelles l'âge d'entrée est supérieur à 120 ans
 - 15 lignes
- Suppression des lignes pour lesquelles l'âge de sortie est négatif ou nul
 - 307 lignes
- Suppression des lignes pour lesquelles l'âge de sortie est inférieur ou égal à l'âge d'entrée
 - 373094 lignes
- Suppression des lignes pour lesquelles le sexe n'est pas renseigné
 - 111754 lignes

B – RESULTATS DETAILLES DE LA REGRESSION DE COX FEMMES / HOMMES

Le portefeuille de référence utilisé pour la régression : organisme 09 pour la période d'observation 01/01/1998 à 31/12/2011.

Femmes	coef	exp(coef)	exp(-coef)	lower 95	upper 95	se(coef)	z	Pr(> z)
organisme01-01/01/1952	0,88107	2,41348	0,4143	2,0478	2,8445	0,08383	10,510	< 2e-16 ***
organisme01-01/01/1967	-0,84013	0,43165	2,3167	0,3925	0,4747	0,04850	-17,323	< 2e-16 ***
organisme01-01/01/1998	-0,87576	0,41655	2,4007	0,3265	0,5315	0,12433	-7,044	1,87e-12 ***
organisme01-01/01/2009	0,20925	1,23275	0,8112	1,1517	1,3195	0,03470	6,031	1,63e-09 ***
organisme02-01/01/2004	-0,42005	0,65701	1,5220	0,5321	0,8112	0,10758	-3,905	9,44e-05 ***
organisme02-01/01/2007	0,37156	1,45000	0,6897	0,9011	2,3332	0,24270	1,531	0,125776
organisme02-01/01/2008	0,40736	1,50284	0,6654	1,1216	2,0137	0,14931	2,728	0,006366 **
organisme03-01/01/2000	0,31000	1,36342	0,7334	1,2477	1,4899	0,04527	6,847	7,53e-12 ***
organisme04-01/01/1999	0,16292	1,17694	0,8497	1,0392	1,3330	0,06353	2,565	0,010329 *
organisme05-01/01/2006	0,84425	2,32624	0,4299	1,8257	2,9640	0,12361	6,830	8,50e-12 ***
organisme06-01/01/2001	-0,07013	0,93227	1,0726	0,9136	0,9513	0,01032	-6,793	1,10e-11 ***
organisme07-01/01/1997	0,09463	1,09925	0,9097	1,0394	1,1626	0,02857	3,312	0,000925 ***
organisme07-01/01/2004	0,16121	1,17494	0,8511	1,1076	1,2464	0,03012	5,352	8,68e-08 ***
organisme08-01/01/1990	0,02740	1,02778	0,9730	0,9936	1,0632	0,01727	1,586	0,112689
organisme08-01/01/2005	-0,05671	0,94487	1,0583	0,8673	1,0294	0,04372	-1,297	0,194562
organisme09-12/26/1979	-0,42726	0,65230	1,5330	0,2929	1,4527	0,40851	-1,046	0,295617
organisme09-12/01/1989	0,04266	1,04358	0,9582	0,4687	2,3237	0,40843	0,104	0,916819
organisme09-06/01/1994	0,00851	1,00855	0,9915	0,3784	2,6882	0,50019	0,017	0,986421
organisme10-12/31/1998	0,01310	1,01319	0,9870	0,9763	1,0515	0,01892	0,692	0,488667
organisme11-01/01/2009	0,42392	1,52795	0,6545	1,4335	1,6287	0,03257	13,016	< 2e-16 ***
organisme12-01/01/2000	0,20855	1,23189	0,8118	1,1588	1,3096	0,03120	6,685	2,31e-11 ***
organisme13-01/01/1989	-0,26238	0,76922	1,3000	0,7283	0,8125	0,02791	-9,401	< 2e-16 ***
organisme13-01/01/2008	-0,77870	0,45900	2,1786	0,3899	0,5404	0,08326	-9,353	< 2e-16 ***
organisme14-07/01/2000	-0,24267	0,78453	1,2747	0,7620	0,8077	0,01484	-16,358	< 2e-16 ***
organisme15-01/01/2000	0,39675	1,48698	0,6725	1,4412	1,5342	0,01595	24,874	< 2e-16 ***

Hommes	coef	exp(coef)	exp(-coef)	lower 95	upper 95	se(coef)	z	Pr(> z)
organisme01-01/01/1952	1,43414	4,19605	0,2383	4,03	4,369	0,02060	69,622	<2e-16***
organisme01-01/01/1967	-0,66124	0,51621	1,9372	0,4862	0,5481	0,03055	-21,648	<2e-16***
organisme01-01/01/1998	-1,06926	0,34326	2,9132	0,2847	0,4139	0,09553	-11,193	<2e-16***
organisme01-01/01/2009	0,33191	1,39363	0,7175	1,312	1,48	0,03069	10,816	<2e-16***
organisme02-01/01/2004	-0,64108	0,52672	1,8985	0,4629	0,5994	0,06593	-9,724	<2e-16***
organisme02-01/01/2007	-0,29498	0,74455	1,3431	0,5945	0,9325	0,11486	-2,568	0,010224*
organisme02-01/01/2008	-6,49461	0,00151	661,5634	8,559E-30	2,67E+23	30,83524	-0,211	0,833182
organisme03-01/01/2000	0,32912	1,38975	0,7196	1,313	1,471	0,02902	11,340	<2e-16***
organisme04-01/01/1999	0,29787	1,34699	0,7424	1,284	1,413	0,02451	12,152	<2e-16***
organisme05-01/01/2006	0,49673	1,64333	0,6085	1,342	2,012	0,10338	4,805	1,55e-06***
organisme06-01/01/2001	-0,09709	0,90747	1,102	0,8944	0,9207	0,00739	-13,144	<2e-16***
organisme07-01/01/1997	0,07716	1,08021	0,9257	1,063	1,097	0,00802	9,615	<2e-16***
organisme07-01/01/2004	0,14217	1,15277	0,8675	1,087	1,222	0,02977	4,776	1,79e-06***
organisme08-01/01/1990	0,03791	1,03864	0,9628	1,024	1,054	0,00729	5,201	1,98e-07***
organisme08-01/01/2005	-0,13377	0,87479	1,1431	0,8118	0,9427	0,03816	-3,506	0,000455***
organisme09-12/26/1979	-0,33903	0,71247	1,4036	0,498	1,019	0,18276	-1,855	0,063586
organisme09-12/01/1989	0,06422	1,06632	0,9378	0,6627	1,716	0,24265	0,265	0,791287
organisme09-06/01/1994	-0,96800	0,37984	2,6327	0,1225	1,178	0,57740	-1,676	0,093644
organisme10-12/31/1998	0,02883	1,02925	0,9716	1,002	1,057	0,01346	2,143	0,032138*
organisme11-01/01/2009	0,43575	1,54612	0,6468	1,481	1,614	0,02199	19,815	<2e-16***
organisme12-01/01/2000	0,09934	1,10444	0,9054	1,07	1,14	0,01625	6,113	9,81e-10***
organisme13-01/01/1989	0,05841	1,06015	0,9433	1,007	1,116	0,02634	2,218	0,026559*
organisme13-01/01/2008	-0,54414	0,58034	1,7231	0,5318	0,6333	0,04454	-12,218	<2e-16***
organisme14-07/01/2000	-0,31489	0,72987	1,3701	0,715	0,7451	0,01053	-29,906	<2e-16***
organisme15-01/01/2000	0,24683	1,27996	0,7813	1,247	1,314	0,01333	18,519	<2e-16***

Le portefeuille « organisme 2 – 01/01/2008 » comporte 776 lignes, ce qui est très faible et explique les résultats obtenus dans la régression de Cox.