

Mémoire présenté devant le CNAM pour l'obtention du Master

Droit Economie Gestion, mention Actuariat et l'admission à l'Institut des Actuares

le 22 novembre 2024

Par : Rakhsat EL HAJJ HASSAN

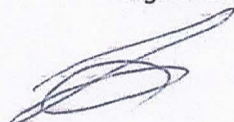
Titre: Construction d'une table de mortalité d'expérience prospective en France dans le contexte de suppression des tables de mortalité générées en assurance

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

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

Présidente du Jury :
M. Stéphane LOISEL

signatures



Entreprise :

Nom : Pas d'entreprise

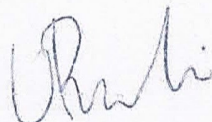
Membres présents du jury de
l'Institut des Actuares :

M. Jean BRUNET

Nom :

Signature :

M. Olivier RENAUDIN



Invité :

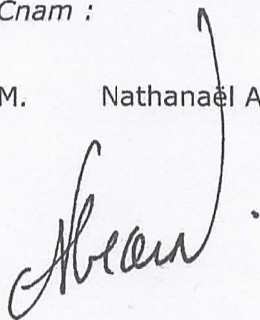
Nom :

M. Stève BAUMANN

Signature :

Membres présents du jury du
Cnam :

M. Nathanaël ABECERA



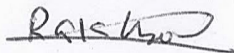
Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Secrétariat :

Signature du responsable entreprise

Bibliothèque :

Signature du candidat



Remerciements

En premier lieu, je tiens ici à exprimer la gratitude que j'éprouve envers Monsieur Pierrick Piette pour sa rigueur intellectuelle et ses remarques pertinentes pour une version préliminaire de ce mémoire.

Je tiens à présenter mes remerciements à Monsieur Stéphane Loisel, Directeur de la Charte Actuariat au Cnam, de m'avoir encouragé d'entamer la réalisation de ce mémoire.

J'aimerais également remercier les autres membres du jury Monsieur Nathanaël Abecera, Monsieur Jean Brunet et Monsieur Olivier Renaudin pour avoir pris le temps de lire mon travail et accepté de faire partie de ce jury.

Je suis aussi redevable de l'aide de nombreuses personnes, auxquelles je tiens à exprimer ici ma reconnaissance notamment Karim Barigou d'avoir pris le temps de répondre à mes questions concernant le package StanMoMo.

Un grand merci aux professeurs et au personnel du Cnam en particulier Gislaine Erny et Melinda Dupuis.

Je suis malheureusement certaine de ne pas citer toutes les personnes qui m'ont apporté leur soutien à un moment ou un autre durant ces années du Master d'actuariat au Cnam, mais j'espère qu'elles accepteront mes remerciements chaleureux et sincères.

Enfin, merci à tous mes proches surtout mes filles qui se reconnaîtront, pour leur présence, leur soutien, leur patience, leur affection et leurs encouragements tout au long du chemin de vie que présente un master.

A mes filles :
Carla, Clara et Laura.

Résumé et mots clés

La loi Industrie verte de 2023 complète d'autres dispositifs récents comme la loi Pacte de 2019 qui a instauré un nouveau Plan d'épargne retraite (PER). Elle rend obligatoire l'utilisation d'une table commune pour les hommes et les femmes au moment de la liquidation pour les contrats d'épargne retraite collectifs à adhésion obligatoire (PERO, Article 83 et L137-11-2). Ceci a un impact sur le métier d'assureur en termes de tarification et de provisionnement des rentes viagères. L'utilisation d'une table de mortalité unisexe pour le calcul de la rente s'applique le 25 octobre 2024 pour tous les nouveaux contrats, et à compter du 25 octobre 2024, pour les contrats souscrits antérieurement au 24 octobre 2023 et reconduits tacitement.

Dans le cadre de ce mémoire, nous avons construit des tables de mortalité d'expérience prospectives d'un organisme de protection sociale français en utilisant les méthodes de positionnement (SMR, BRASS et GLM-Poisson) par rapport à des tables de référence. Ces dernières ont été construites à partir de la base de données Human Mortality Database pour la population française. Différentes tables de référence réalisées pour les hommes, les femmes et tous sexes confondus et mettant en évidence l'impact de la pandémie Covid 19 sur la projection à long terme de la mortalité.

Aussi, ce mémoire a mis en évidence les impacts pour l'assureur de l'utilisation d'une table de mortalité unisexe pour les rentes viagères en termes de tarification et provisionnement. Cette analyse met en évidence l'intérêt pour l'assureur de construire une table de mortalité d'expérience prospective qui reflète la mortalité réelle de son portefeuille.

Mots clés : Rentes viagères, France, risque de longévité, tables de mortalité prospectives d'expérience, modèle stochastique, COVID-19, modèle bayésien, tarification, provisionnement, loi Industrie Verte, Human Mortality Database

Abstract

The Green Industry law complements other recent measures such as the Pacte law which introduced a new Retirement Savings Plan. It makes the use of a common table mortality for men and women at the time of liquidation for collective retirement savings contracts with compulsory membership (PERO, Article 83 and L137-11-2). This has an impact on the insurance business in terms of pricing and provisioning of life annuities. The use of a unisex mortality table for calculating the annuity applies from October 25, 2024 for all new contracts, and from October 25, 2024, for contracts taken out before October 24, 2023 and tacitly renewed.

As part of this study, we constructed prospective mortality tables from a French social protection organization using positioning methods (SMR, BRASS and GLM-Poisson) in relation to reference tables. These were constructed from the Human Mortality Database for the French population. Different reference tables produced for men, women and all sexes combined and highlighting the impact of the Covid 19 pandemic on the long-term trend in mortality.

Also, this study highlighted the impacts for the insurer of the use of a unisex mortality table for life annuities in terms of pricing and provisioning. This analysis highlights the interest for the insurer in constructing a mortality table based on prospective experience which reflects the actual mortality of its portfolio.

Keywords: Life annuities, France, longevity risk, prospective mortality tables based on experience, stochastic model, COVID-19, Bayesian model, pricing, provisioning, Green Industry law, Human Mortality Database

Table des matières

Remerciements	i
Résumé et mots clés	iii
Abstract	iv
Liste des figures	vii
Liste des tableaux	ix
INTRODUCTION	1
Chapitre 1 : CONTEXTE REGLEMENTAIRE	5
1.1. Les principes de réglementation des tables de mortalité en assurance	5
1.2. Les tables réglementaires existantes	5
1.2.1. Tables de mortalité	5
1.2.2. Tables TH 00-02 et TF 00-02	6
1.2.3. Tables TGH 05 – TGF 05	6
1.3. Les tables Insee	7
CHAPITRE 2 : CONSTRUCTION DES TABLES DE MORTALITE PROSPECTIVES DE REFERENCE A PARTIR DE LA BASE « HUMAN MORTALITY DATABASE »	8
2.1. Présentation de la base des données HMD	8
2.2. Analyse de la mortalité en France	9
2.3. Modèles Generalized Age-Period-Cohort	15
2.3.1. Modèle Lee-Carter	16
2.3.2. Modèle de Renshaw -Haberman	17
2.3.3. Modèle Age-Period-Cohort (APC)	17
2.3.4. Modèle de Cairns-Blake-Dowd	18
2.4. Résultats des modèles GAPC appliqués aux données de la population française	19
2.4.1. Mise en œuvre des modèles GAPC	19
2.4.2. Indicateurs AIC, BIC et nombre de paramètres	22
2.4.3. Analyse des résidus	24
2.5. Approche bayésienne dans la construction des tables de mortalité de référence	26
2.5.1. Différence entre approche fréquentiste et approche bayésienne	26
2.5.2. Mise en œuvre de l’approche bayésienne dans le cadre de la prédiction de la mortalité en utilisant « leave-future-out validation »	29
2.5.3. Performance prédictive des modèles	37
2.5.4. Modèles choisis en fonction de leurs performances	44
2.6. Résultats des projections pour les tables de référence « HMD »	45
2.6.1. Projections de la composante temporelle des modèles retenus : LC, RH et APC	45
2.6.2. Espérance de Vie médian 50-90	50
2.6.3. Evolution des quotients de mortalité	51
Chapitre 3 : Construction d’une table de mortalité d’expérience à partir d’un jeu de données confidentielles d’un organisme de protection sociale	57

3.1. Présentation et analyse du jeu de données	57
3.2. Construction de tables de mortalité d'expérience	59
3.2.1. Exposition au risque décès : censure et troncature	59
3.2.2. Calcul des taux bruts de mortalité	61
3.3. Synthèse des travaux antérieurs de construction des tables de mortalité	65
 dans le cadre des mémoires d'actuariat en France	65
Chapitre 4 : Construction des tables de mortalité d'expérience prospectives : les modèles de positionnement	67
4.1. Les tables de références externes	67
4.2. Trois méthodes de positionnement testées	68
4.3. Application et analyse des résultats des trois méthodes d'ajustement	70
4.3.1. Estimation du SMR	70
4.3.2. Estimation des paramètres du modèle logistique	71
4.3.3. Estimation des paramètres du modèle linéaire généralisé de Poisson	71
4.4. Validation des tables et choix du modèle	72
4.4.1. Validation visuelle des ajustements de la mortalité passée au niveau local	72
4.4.2. Résultats des tests de proximité	73
4.4.3. Résultats des tests de régularité et d'ajustement	77
4.4.4. Choix du modèle fondé sur les résultats des tests des différentes méthodes	80
4.4.5. Fermeture des tables : Méthode de DENUIT & GODERNIAUX	81
4.4.6. Cohérence des tendances d'évolution de la mortalité	83
4.4.7. Conclusion sur le choix du modèle	89
Chapitre 5 : Impact des tables de mortalité d'expérience non générées sur la tarification et le provisionnement des contrats d'épargne retraite d'une compagnie d'assurance vie	90
5.1. Impact sur la tarification	90
5.1.1 Exemple d'un contrat retraite	90
5.1.2. Comparaison des capitaux constitutifs de rente selon la table de mortalité retenue	93
5.2. Impact sur les provisionnements	95
5.2.1. Contrat d'épargne et provisionnement	95
5.2.2. La provision mathématique diffère donc selon la phase du contrat de rente viagère considéré	96
5.2.3. Illustration de l'impact d'un changement de table de mortalité sur le montant de provisionnement à travers des cas types	98
5.3. Conséquences financières pour l'assureur	102
5.3.1. Risque d'insuffisance des provisions	102
5.3.2. Réévaluation des engagements à long terme	102
5.4. Synthèse des impacts de suppression d'une table de mortalité générée	103
5.4.1. Effet sur les provisions mathématiques	103
5.4.2. Effet sur la tarification : incertitude accrue	104
Conclusion	105
Bibliographie	108
Annexes	i
Annexe A	i

Liste des figures

Figure 0-1 : Encours de l'assurance retraite en France. _____	2
Figure 0-2 : Evolution de l'espérance de vie à la naissance de 1740 à 2020. _____	3
Figure 2-1 : Evolution du taux de mortalité entre 1980 et 2000. _____	9
Figure 2-2 : Différence entre décès observés et décès attendus en France pour les années 2020 à 2022. _____	10
Figure 2-3 : Ecart entre les décès observés et attendus par mois de janvier 2020 à décembre 2022. _____	11
Figure 2-4 : Écart entre les décès observés et attendus par âge de 2020 à 2022. _____	12
Figure 2-5 : Evolution de la mortalité par âge en France, tous sexes confondus, de 2019 à 2021. _____	13
Figure 2-6 : Evolution de la mortalité par âge en France, de la population masculine, de 2019 à 2021. _____	14
Figure 2-7 : Evolution de la mortalité par âge en France, de la population féminine, de 2019 à 2021. _____	14
Figure 2-8 : Surmortalité des hommes par rapport aux femmes selon l'âge et la cause de décès. _____	15
Figure 2-9 : Paramètres du modèle LC pour tous sexes confondus. _____	20
Figure 2-10 : Paramètres du modèle LC pour la population masculine. _____	21
Figure 2-11 : Paramètres du modèle LC pour la population féminine. _____	22
Figure 2-12 : Résidus du modèle LC, pour les hommes, femmes, tous sexes confondus avant et y compris covid. _____	24
Figure 2-13 : Résidus du modèle RH, pour les hommes, femmes, tous sexes confondus avant et y compris covid. _____	25
Figure 2-14 : Résidus du modèle APC, pour les hommes, femmes, tous sexes confondus avant et y compris covid. _____	25
Figure 2-15 : Résidus du modèle CBD, pour les hommes, femmes, tous sexes confondus avant et y compris covid. _____	26
Figure 2-16 : Illustration de overfitting et underfitting. _____	30
Figure 2-17 : Illustration de déroulement d'un modèle de Stacking. _____	30
Figure 2-18 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	39
Figure 2-19 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	39
Figure 2-20 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	40
Figure 2-21 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	40
Figure 2-22 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	41
Figure 2-23 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA. _____	41
Figure 2-24 : Evolution de la composante temporelle du modèle Lee-Carter. _____	49
Figure 2-25 : Evolution de la composante temporelle du modèle Lee-Carter. _____	49
Figure 2-26 : Evolution de la composante temporelle du modèle Lee-Carter. _____	49
Figure 2-27 : Evolution de l'espérance de vie à 50 ans (tronquée à 90 ans) avant covid par sexe. _____	50
Figure 2-28 : Evolution de l'espérance de vie à 50 ans (tronquée à 90 ans) y compris covid par sexe. _____	51
Figure 2-29 : Projection des quotients de mortalité avant covid 19 – tous sexes confondus. _____	52
Figure 2-30 : Projection des quotients de mortalité avant covid 19 – Hommes. _____	52
Figure 2-31 : Projection des quotients de mortalité avant covid 19 – Femmes. _____	53
Figure 2-32 : Projection des quotients de mortalité y compris covid 19 – tous sexes confondus. _____	54
Figure 2-33 : Projection des quotients de mortalité y compris covid 19 – Hommes. _____	54
Figure 2-34 : Projection des quotients de mortalité y compris covid 19 – Femmes. _____	55
Figure 2-35 : Ecart entre réalisations et prévisions du nombre de décès par âge et selon le sexe des années 2020 et 2021. _____	56
Figure 3-1 : Répartition des adhérents selon leur âge et leur statut. _____	58
Figure 3-2 : Illustration des censures et troncatures _____	60
Figure 3-3 : Répartition par âge des expositions et des décès selon le sexe. _____	60
Figure 3-4 : Répartition des décès par âge et année d'observation. _____	61

<i>Figure 3-5 : Répartition des expositions par âge et année d'observation.</i>	61
<i>Figure 3-6 : Illustration du nombre des présents et des sorties sur la période d'observation.</i>	63
<i>Figure 3-7 : Quotients de mortalité par âge selon les deux méthodes : Koem et Kaplan-Meier.</i>	64
<i>Figure 3-8 : Quotients de mortalité à l'échelle logarithmique par âge et année d'observation.</i>	64
<i>Figure 4-1 : Comparaison des quotients de mortalité par âge pour l'année 2008 observés et ajustés selon les trois méthodes</i>	73
<i>Figure 4-2 : Comparaison des quotients de mortalité par âge pour l'année 2008 observés et ajustés selon les trois méthodes</i>	73
<i>Figure 4-3 : Taux de mortalité (échelle logarithmique) avant et après complétude de table de mortalité.</i>	82
<i>Figure 4-4 : Taux de mortalité (échelle logarithmique) avant et après complétude de table de mortalité.</i>	83
<i>Figure 4-5 : Espérance de vie par âge selon les années pour la population masculine.</i>	85
<i>Figure 4-6 : Evolution de l'espérance de vie résiduelle pour certains âges de la population féminine.</i>	86
<i>Figure 4-7 : Evolution de l'espérance de vie résiduelle des personnes âgées de 60 ans par sexe.</i>	87
<i>Figure 4-8 : Tables de mortalité prospectives.</i>	87
<i>Figure 4-9 : Tables de mortalité prospectives.</i>	88
<i>Figure 4-10 : Tables de mortalité prospectives.</i>	88
<i>Figure 4-11 : Tables de mortalité prospectives.</i>	88
<i>Figure 5-1 : Espérance de vie des 50-90 ans selon la table de mortalité.</i>	93
<i>Figure 5-2 : Schéma de déroulement de vie d'un contrat de rente viagère.</i>	98

Liste des tableaux

Tableau 2-1: Facteurs expliquant la hausse des décès en France depuis 2019. _____	11
Tableau 2-2 : Valeurs du nombre de paramètres, de l'AIC et du BIC obtenus pour les 4 modèles GAPC. _____	23
Tableau 2-3 : Comparaison entre les deux approches fréquentiste et bayésienne. _____	28
Tableau 2-4 : : Division de la base de données en trois sous-ensembles. _____	35
Tableau 2-5 : Répartition des données associées à chaque méthode pour la période avant covid 19. _____	36
Tableau 2-6 : Log score et CRPS selon les trois méthodes : BMA, Stacking et pseudo BMA. _____	43
Tableau 2-7 : Log score et CRPS selon les modèles individuels. _____	43
Tableau 2-8 : Ajustement d'un modèle ARIMA à la série temporelle des indices de mortalité Kt du modèle LC y compris Covid _____	46
Tableau 3-1 : Répartition du motif de sortie des assurés à la date de fin d'observation. _____	57
Tableau 3-2 : Répartition du statut des assurés à la date de début d'observation. _____	57
Tableau 3-3 : Répartition des assurés par sexe en fonction de leur statut à la date de début d'observation. ____	58
Tableau 3-4 : Répartition des assurés par sexe en fonction de leur motif de sortie à la date de fin d'observation. _____	59
Tableau 3-5 : Répartition des assurés par motif de sortie en fonction de leur statut à la date de début d'observation. _____	59
Tableau 3-6 : Critère de Cochran. _____	65
Tableau 4-1 : SMR selon la table de mortalité de référence retenue. _____	70
Tableau 4-2 : Valeurs des paramètres α et β du modèle de positionnement logistique. _____	71
Tableau 4-3 : Paramètres du modèle linéaire généralisé de Poisson pour les assurés. _____	72
Tableau 4-4 : Résultats des tests de SMR et Wilcoxon. _____	75
Tableau 4-5 : Valeurs d'intérêt selon les tables de référence retenues et les trois méthodes d'ajustement. ____	77
Tableau 4-6 : Résultats des tests des runs selon la méthode d'ajustement et la table de mortalité retenue. ____	79
Tableau 4-7 : Résultats des tests de signes selon la table de mortalité de référence et la méthode d'ajustement. _____	80
Tableau 4-8 : Affectation d'un poids à chaque méthode selon ses résultats aux différents tests. _____	81
Tableau 4-9 : Présentation des taux de survie par âge et année. _____	83
Tableau 4-10 : Illustration du sens de calcul des quotients de survie par âge et année . _____	84
Tableau 4-11 : Modèle retenu pour les projections de long terme des hommes, femmes et tous sexes confondus. _____	89
Tableau 5-1 : Prime unique d'une rente viagère de 100 € selon la table de mortalité retenue et l'âge de l'assuré à la souscription _____	92
Tableau 5-2 : Comparaison des capitaux constitutifs obtenus les différentes tables de mortalité _____	95
Tableau 5-3 : Calcul des ax en fonction de l'âge selon les tables réglementaires actuelles _____	100
Tableau 5-4 : Calcul du montant de l'arrérage r correspondant à un capital constitutif de 10000 € par assuré _____	100
Tableau 5-5 : Ecart de provisionnement par rapport aux tables réglementaires, en %. _____	101

INTRODUCTION

La loi industrie verte, publiée le 24 octobre 2023 au Journal Officiel, vise à accélérer la mobilisation de financements privés au service de la transition écologique. Certaines mesures impactent l'épargne retraite et l'épargne salariale. En effet, cette loi marque la fin de l'exception permettant aux contrats d'épargne retraite collectifs à adhésion obligatoire (PERO, Article 83 et L137-11-2) de recourir à des tables de mortalité genrées.

Les tables de mortalité, qui estiment l'espérance de vie d'une génération, et permettent aux assureurs de déterminer la durée de versement d'une rente viagère, ne tiennent pas compte du sexe de l'assuré depuis 2013 pour les contrats de retraite individuels. En effet, l'arrêt « Test-Achats » du 1er mars 2011 de la Cour de justice de l'Union européenne, était à l'origine de l'harmonisation de ces tables de mortalité pour les contrats individuels de retraite dès le 21 décembre 2012.

Cette disposition va désormais être étendue aux contrats d'épargne retraite collectifs, en vertu de l'article 35 de la loi Industrie verte, qui modifie l'article L111-7 du Code des assurances. En effet, l'interdiction d'utiliser des tables de mortalité différenciées par sexe s'applique :

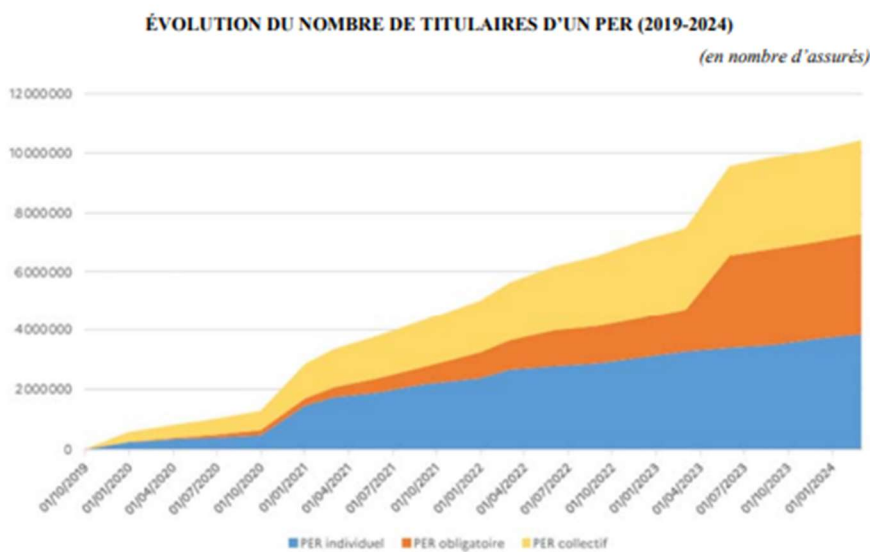
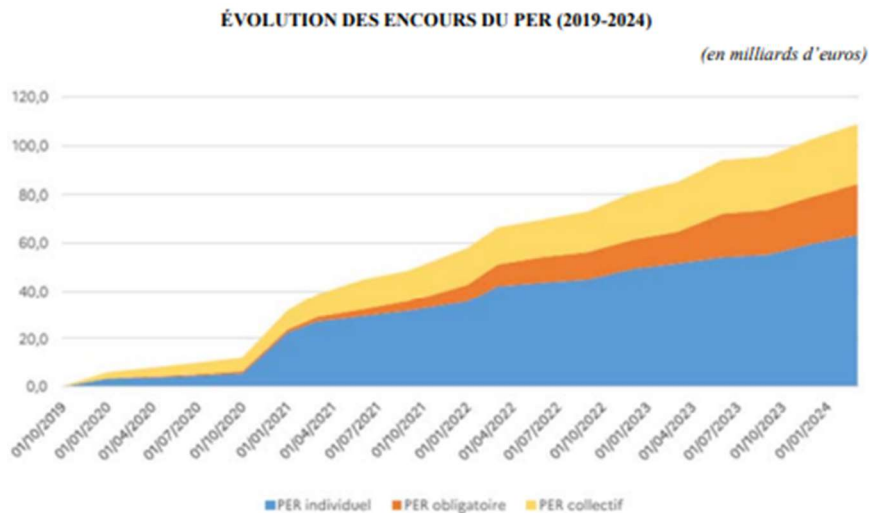
- Le 25 octobre 2024 au plus tard (un an après la parution au JO), pour tous les nouveaux contrats,
- A compter du 25 octobre 2024, pour les contrats souscrits antérieurement au 24 octobre 2023 et reconduits tacitement.

Cette obligation d'uniformisation des tables de mortalité soulève des questions : quelle table de mortalité adopter à l'avenir ? Comment la construire ? Quels sont les impacts directs sur le métier d'assureur en termes de tarification et de provisionnement des rentes viagères ?

Il convient de rappeler que la loi Pacte de 2019 a instauré un nouveau Plan d'épargne retraite (PER) afin de simplifier le dispositif et de démocratiser le marché de l'épargne retraite.

Le Plan d'Epargne Retraite (PER) d'entreprise obligatoire est un plan ouvert à tous les salariés d'une entreprise ou réservé à certaines catégories de salariés. Les salariés concernés ont l'obligation de souscrire. Ce plan succède aux contrats article 83. Ces derniers ne sont plus commercialisés depuis la mise en place du PER et de sa version obligatoire en entreprise : le PERO. Les deux formules ont en commun des avantages fiscaux et une même finalité : garantir aux salariés un revenu supplémentaire fixe à la retraite. Et, côté entreprise, motiver et fidéliser ses collaborateurs.

Le nouveau régime de retraite à prestations définies L137-11-2 est un régime de retraite à prestations définies à droits acquis. Il vient compléter une offre de produits déjà riche : les nouveaux plans d'épargne retraite (PER) issus de la loi Pacte, les Article 82, les LTI à vocation retraite.



Source : réponses écrites de la direction générale du Trésor.

Figure 0-1 : Encours de l'assurance retraite en France.

Source : GÉRARD F. ET DE COURSON C. (2024), Rapport d'information sur la fiscalité de l'épargne retraite par capitalisation-Assemblée Nationale.

Selon la direction générale du Trésor, l'encours des PER est passé de 66,4 milliards d'euros au 31 mars 2022 à 108,8 milliards d'euros au 31 mars 2024 (+ 63,9 %), réparti entre les PER individuels (63,3 milliards d'euros), les PERECO (24,7 milliards d'euros) et les PEROB (20,8 milliards d'euros).

Par ailleurs, le nombre de titulaires d'un PER s'établit à près de 10,4 millions de bénéficiaires au 31 mars 2024. Les PER individuels comptent 3,9 millions de titulaires, contre respectivement 3,2 millions de titulaires d'un PERECO et 3,4 millions de titulaires d'un PEROB.

▪ **L'espérance de vie, exprimée en années, a presque doublé au cours du vingtième siècle**

Au milieu du 18^{ème} siècle, l'espérance de vie ne dépassait pas 25 ans. Elle atteint 30 ans à la fin du siècle, puis fait un bond à 37 ans en 1810 en partie grâce à la vaccination contre la variole. La hausse se poursuit à un rythme lent pendant le 19^{ème} siècle, pour atteindre 45 ans en 1900. Pendant les guerres de Napoléon et la guerre de 1870, l'espérance de vie décline brutalement et repasse sous les 30 ans. A l'exception des deux guerres mondiales, la hausse de l'espérance de vie se poursuit, au cours du 20^{ème} siècle, grâce aux progrès médicaux.

Selon l'Ined, « l'espérance de vie à la naissance atteint 79,3 ans pour les hommes et 85,4 ans pour les femmes en 2021, contre 79,1 ans et 85,1 ans en 2020, soit un gain de 0,2 an pour les hommes et 0,3 an pour les femmes. C'est un gain important, mais l'espérance de vie avait reculé de 0,6 an pour les hommes et 0,5 an pour les femmes en 2020 par rapport à 2019 en raison de l'épidémie de Covid-19. La hausse de 2021 ne permet pas de rattraper le niveau de 2019 – 79,7 ans pour les hommes et 85,6 ans pour les femmes – l'épidémie de Covid-19 ayant continué et entraîné de nombreux décès en 2021 ».

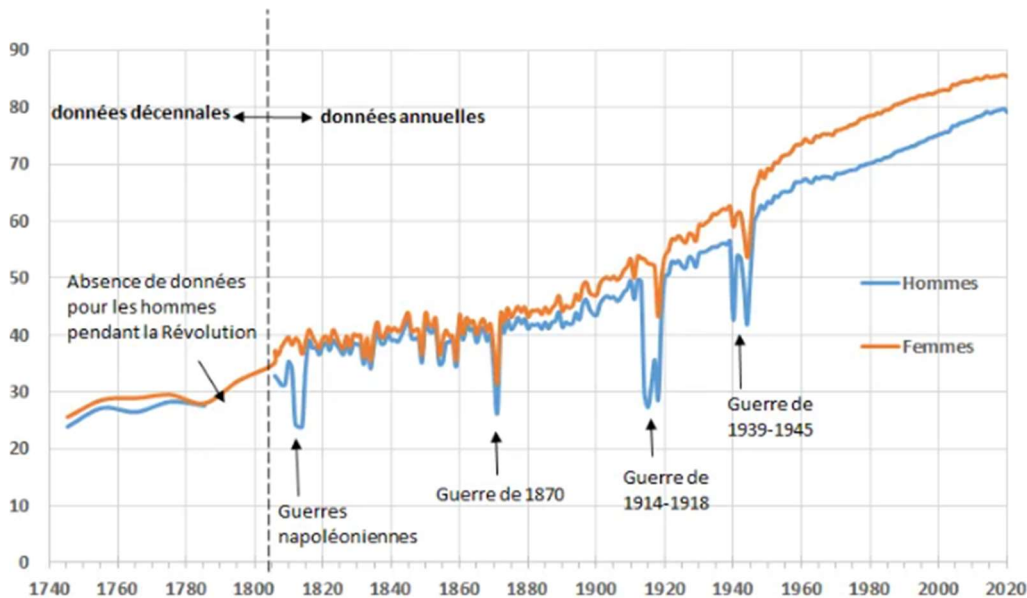


Figure 0-2 : Evolution de l'espérance de vie à la naissance de 1740 à 2020.

Source : www.ined.fr

L'espérance de vie à 60 ans continue de progresser en France ces dernières années. D'après les récentes études de l'INSEE, en vingt ans, elle a progressé de 2,3 ans pour les femmes et 3,2 ans pour les hommes (atteignant respectivement en 2018 27,6 ans et 23,2 ans). C'est dans ce contexte que les assureurs, notamment les organismes de retraite et d'assurance vie ont été amenés à étudier le risque de longévité inhérent à leur portefeuille et à le considérer dans leur cartographie. Ce risque provient de l'incertitude liée à l'allongement de la durée de vie des assurés et les concerne directement.

Comme il y a une décroissance du taux de mortalité au fil des années grâce notamment au progrès médical, et donc une augmentation de la longévité au cours du temps des assurés, l'utilisation par un assureur d'une table du moment pour évaluer ses risques et engagements entraînera une perte substantielle pour la société. C'est pourquoi, il est essentiel d'utiliser des tables de mortalité prospectives qui incorporent l'évolution présumée de la durée de vie.

Pour ces assureurs, les hypothèses de mortalité retenues sont structurantes pour l'établissement des tarifs et le calcul des provisions, dans les référentiels social, IFRS 17 et Solvabilité 2. Un assureur détenant principalement des rentes viagères immédiates ou différées verra ses engagements augmenter si la longévité observée est plus forte que celle estimée par les hypothèses.

L'objectif des tables prospectives est de prendre en compte les évolutions futures de la mortalité. Dans la littérature classique, la construction des tables de mortalité prospectives passe d'abord par l'ajustement des tendances passées, puis par l'extrapolation du futur.

La démarche adoptée comporte principalement deux étapes. Dans un premier temps les taux bruts, par âge et année d'observation sont estimés. Dans un second temps les taux bruts obtenus sont ajustés à un modèle permettant de corriger les irrégularités dues aux fluctuations d'échantillonnage et d'effectuer une projection des taux bruts futurs par extrapolation.

En général, les projections de taux de mortalité sont réalisées selon une procédure fréquentiste en deux étapes. D'abord, les estimations des paramètres ; puis dans un second temps, les paramètres sont projetés en utilisant des techniques de séries chronologiques.

Dans ce mémoire, nous considérons la modélisation de la mortalité dans un cadre bayésien en mettant en œuvre la méthodologie développée dans le cadre de l'article publié par Barigou K. et al. (2022) Comparée au cadre classique, l'approche bayésienne offre deux avantages notables. Premièrement, les étapes d'estimation et de prévision vont de pair, ce qui conduit à des estimations plus cohérentes. Deuxièmement, ceci rend mieux compte de différentes sources d'incertitude de manière naturelle et cohérente.

Ce mémoire a pour but de :

- Construire une table de mortalité prospective à partir des données nationales (données HMD) en tenant compte de l'impact de la pandémie Covid 19,
- Construire une table de mortalité d'expérience à partir de la base de données confidentielles d'un organisme de protection sociale en France,
- Positionner la table de mortalité d'expérience par rapport aux tables de mortalité HMD,
- Analyser l'impact pour l'assureur de l'utilisation d'une table de mortalité unisexe.

Pour construire des tables de mortalité prospectives d'expérience susceptibles d'être utilisées dans le cadre des contrats d'épargne retraite collectifs à adhésion obligatoire (PERO), notre démarche consiste à calibrer des tables de mortalité de la population française en utilisant les données de la base Human Mortality Database (HMD) pour laquelle on dispose d'un historique important. Ensuite, pour tenir compte de la mortalité spécifique des assurés bénéficiaires de rentes viagères, on procède par ajustement des taux de mortalité à partir de données confidentielles des assurés d'un organisme de protection sociale gérant de retraite. Plusieurs modèles peuvent être utilisés pour mettre en place des tables de mortalité prospectives sur la population générale que nous testerons pour opter pour le modèle le mieux adapté aux données parmi les modèles testés.

Le mémoire se divise en cinq chapitres. Le premier chapitre consistera à rappeler le contexte réglementaire dans lequel s'inscrit notre étude. Un deuxième chapitre portera sur la construction des tables de mortalité prospectives de référence à partir de la base « Human Mortality Database ». Dans le troisième chapitre, la description du portefeuille et la présentation de la table de mortalité d'expérience de ce portefeuille seront faites. Le quatrième chapitre présentera les différentes étapes de construction de tables de mortalité d'expérience prospectives. Un dernier chapitre sera consacré au chiffrage des impacts du changement de table de mortalité sur la tarification et le provisionnement d'un portefeuille fictif d'assurés.

Chapitre 1 : CONTEXTE REGLEMENTAIRE

1.1. Les principes de réglementation des tables de mortalité en assurance

Les tables de mortalité sont utilisées par les assureurs lors de la tarification des contrats ainsi que lors du calcul des provisions. Le choix des tables pour ces calculs est contraint par la réglementation.

L'article A.132-18 du Code des Assurances stipule que « Les tarifs pratiqués par les entreprises d'assurance sur la vie et de capitalisation et par les fonds de retraite professionnelle supplémentaire comprennent la rémunération de l'entreprise et sont établis d'après les éléments suivants :

1° Un taux d'intérêt technique fixé dans les conditions prévues à l'article [A. 132-1](#).

2° Une des tables suivantes :

a) Tables homologuées par arrêté du ministre de l'économie et des finances, établies par sexe, sur la base de populations d'assurés pour les contrats de rente viagère, et sur la base de données publiées par l'Institut national de la statistique et des études économiques pour les autres contrats ;

b) Tables établies ou non par sexe par l'entreprise d'assurance et certifiées par un actuaire indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire reconnues par l'autorité mentionnée à l'article [L. 310-12](#).

Les tables mentionnées au b sont établies d'après des données d'expérience de l'entreprise d'assurance, ou des données d'expérience démographiquement équivalentes.

Lorsque les tarifs sont établis d'après des tables mentionnées au a, et dès lors qu'est retenue une table unique pour tous les assurés, celle-ci correspond à la table appropriée conduisant au tarif le plus prudent.

Pour les contrats en cas de vie autres que les contrats de rente viagère, les tables mentionnées au a sont utilisées en corrigeant l'âge de l'assuré conformément aux décalages d'âge ci-annexés.

Pour les contrats de rentes viagères, en ce compris celles revêtant un caractère temporaire, et à l'exception des contrats relevant du chapitre III du titre IV du livre Ier, le tarif déterminé en utilisant les tables mentionnées au b ne peut être inférieur à celui qui résulterait de l'utilisation des tables appropriées mentionnées au a.

Pour les contrats collectifs en cas de décès résiliables annuellement, le tarif peut être établi d'après les tables mentionnées au a avec une méthode forfaitaire si celle-ci est justifiable ».

1.2. Les tables réglementaires existantes

Nous citons ici les tables les plus couramment utilisées (Planchet F., 2007).

1.2.1. Tables de mortalité

Une table de mortalité met en regard, pour tous les âges, le nombre de survivants d'âge x , L_x , dans un groupe fermé. Elle permet, ainsi, de déterminer pour chaque âge x : le nombre de décès, le taux annuel de mortalité, l'espérance de vie¹ et la probabilité de survivre ou de décéder sur une période donnée.

Les tables de mortalité constituent un élément central pour la tarification et le provisionnement d'une compagnie d'assurance. Il s'agit d'une construction statistique, destinée à suivre les probabilités de décès d'une population, en fonction de son âge et de son sexe, à un moment donné ou par génération.

Nous distinguons deux catégories de tables :

- La table de mortalité statique, périodique ou instantanée (ou « table de mortalité du moment ») qui caractérise la mortalité d'une population à un moment précis, toutes générations confondues. Selon l'Insee, « une table de mortalité annuelle suit le cheminement d'une génération fictive de 100 000 nouveau-nés à qui l'on fait subir aux divers âges les conditions de mortalité observées sur les diverses générations réelles, durant l'année étudiée. Pour éviter les aléas des tables annuelles et pour disposer d'une table détaillée par âge aussi précise que possible, on calcule également une table de mortalité couvrant une période de trois années ».
- La table dynamique ou prospective (ou « table de mortalité par génération ») intègre l'âge et le temps dans la probabilité de décès. Elle est établie à partir d'une génération réelle (et non fictive comme la précédente table), observant les niveaux de mortalité réels en fonction de l'année de naissance.
Ces tables permettent d'anticiper l'allongement de la durée de vie et de se prémunir contre le risque de survie des assurés. Il apparaît nécessaire de prendre en compte ce phénomène sur les contrats dont la durée d'engagement peut se révéler longue.

1.2.2. Tables TH 00-02 et TF 00-02

Les tables TH 00-02 et TF 00-02 – respectivement pour la population masculine et pour la population féminine- reflètent la mortalité moyenne de la population française dans sa globalité. Elles sont construites à partir de la table INSEE 2000-2002. Il s'agit des tables réglementaires pour les contrats d'assurance vie (sauf rentes viagères). Les tables TF 00-02 et TH 00-02 figurent à l'annexe 1 de l'article A 132-18 du Code des assurances. Elles sont consultables dans le fac-similé du JO n° 0301 du 29/12/2015. Elles sont entrées en vigueur en 2006 pour remplacer les tables TV et TD88-90 qui n'étaient pas différenciées par sexe.

1.2.3. Tables TGH 05 – TGF 05

A la différence d'une table de moment, une table de mortalité par génération s'appuie sur une génération réelle. Elle possède l'avantage d'être plus représentative de la réalité mais ne peut être construite qu'une fois la génération entièrement éteinte.

¹ L'espérance de vie à l'âge x est la durée moyenne restant à vivre après x , ou durée de vie moyenne au-delà de x , pour les survivants à l'âge x (Delwarde A. & Denuit M., 2006).

TGH 05 – TGF 05 sont des tables par génération. Elles comprennent en fait autant de tables de mortalité qu'il y a d'années de naissance comprises entre 1886 et 2005). Il s'agit des tables réglementaires pour les opérations de rentes viagères. Ces tables sont les premières différenciées par sexe pour la tarification des rentes.

Les tables par génération TGF05 et TGH05 ont été publiées pour la première fois en annexe à l'arrêté du 1er août 2006 portant homologation des tables de mortalité pour les rentes viagères et modifiant certaines dispositions du code des assurances en matière d'assurance sur la vie et de capitalisation. Ces tables sont venues remplacer la TPG93 en 2007.

À ces tables de mortalité, les assureurs peuvent par ailleurs utiliser des tables d'expérience établies à partir de leur portefeuille d'assurés et certifiées par un actuaire indépendant agréé à cet effet, conformément aux dispositions de l'article A335-1 du Code des Assurances. Cela permet de mieux représenter le risque de mortalité du portefeuille. Ces tables doivent être suivies annuellement pour s'assurer de leur validité car une table d'expérience est certifiée pour une durée de 5 ans maximum).

Conformément à la Charte de certification et de suivi des tables de mortalité, le certificateur a pour mission de :

- « Valider les données utilisées et leurs sources, qu'elles soient internes ou externes à l'entreprise,
- Vérifier les hypothèses de travail et les modalités utilisées pour construire les tables de mortalité ou les lois de maintien en incapacité de travail ou en invalidité,
- S'assurer que les principes de prudence communément admis ont été respectés, eu égard aux risques induits (en particulier stabilité des tables ou des lois de maintien),
- Définir précisément les conditions d'application et de validité des éléments certifiés, les statistiques ou tableaux de bord à préparer périodiquement par l'entreprise pour permettre le suivi des résultats d'expérience.»

1.3. Les tables Insee

Les tables INSEE mises en place chaque année par l'INSEE à partir de l'enregistrement des décès à l'état civil. Ces tables sont en général utilisées pour évaluer les engagements sociaux de type « indemnités de fin de carrière » ou « médailles du travail » des entreprises. Aussi, le Conseil d'Orientation des Retraites utilise ces tables dans le cadre de l'élaboration de son rapport annuel sur les évolutions et les perspectives des retraites en France. Dans ce cadre, il établit des préconisations en matière d'allongement de la durée de cotisation pour les régimes de retraite de base.

CHAPITRE 2 : CONSTRUCTION DES TABLES DE MORTALITE PROSPECTIVES DE REFERENCE A PARTIR DE LA BASE « HUMAN MORTALITY DATABASE »

2.1. Présentation de la base des données HMD

La Human Mortality Database (HMD) est une référence de place qui permet d'obtenir gratuitement des données de mortalité sur une dizaine de pays, avec une uniformité sur les outils démographiques utilisés et des standards de qualité.

L'étude de la mortalité en France peut être réalisée avec les données de la Human Mortality Database (HMD). Cette base de données est le fruit d'un projet international lancé en 2002 et qui, fin 2021, collecte et rend disponible des données sur les populations et les décès dans 41 pays dont la France. Pour la France, HMD s'appuie sur les données de l'Institut National d'Etudes Démographiques (INED) et de l'Institut National de la Statistique et des Etudes Economiques (INSEE).

Les données disponibles dans la base de données HMD pour chaque population sont :

- Les naissances par année,
- Les décès par âge et par année (y compris les triangles de Lexis),
- La population par âge et par année,
- L'exposition au risque par âge et par année (y compris les triangles de Lexis),
- Les taux de décès par âge et par année,
- Les tables de mortalité,
- L'espérance de vie à la naissance.

Les données sont différenciées par sexe (population totale, hommes et femmes) et sont disponibles par période ou par cohorte pour les âges 0 à 110+.

Dans le cadre de ce mémoire, j'utilise les données HMD de la France. L'historique des données s'étend de 1816 à 2021. Cependant, je n'utiliserai les données qu'à partir de 1982 afin d'éviter le biais des estimations des paramètres des modèles en incluant les données relatives aux périodes de guerres mondiales.

Les traitements effectués sur les données brutes sont explicités en détail dans les protocoles et la documentation de la HMD.

Fin 2021, les données par période pour la France sont disponibles de 1816 à 2021 avec une distinction entre la population civile et la population totale lors des guerres. Dans la suite de l'analyse, les données HMD utilisées seront celles de la population civile téléchargées au 31 décembre 2021.

2.2. Analyse de la mortalité en France

Avec l'arrivée des générations nombreuses du baby-boom à des âges de forte mortalité, le nombre de décès a tendance à augmenter plus vite sur les dix dernières années (+0,7 % par an en moyenne entre 2004 et 2014, puis +1,9 % entre 2014 et 2019 selon l'Insee). L'augmentation de la mortalité a cependant été sans commune mesure en 2020 du fait des pics de mortalité lors des deux premières vagues de la pandémie et les décès sont restés à un niveau élevé en 2021 et 2022.

▪ Evolution du taux de mortalité en France depuis 1982

L'évolution annuelle du taux de mortalité de la population française s'explique par deux facteurs qui évoluent en sens inverse : la mortalité à un âge donné, qui baisse tendanciellement de manière assez régulière depuis les années 1982 et la structure par âge de la population, dans laquelle le poids des personnes plus âgées croît. Depuis le début des années 2000, l'effet de vieillissement de la population l'emporte. Il est à noter que la mortalité globale en France est au niveau de celle de l'année 1986.

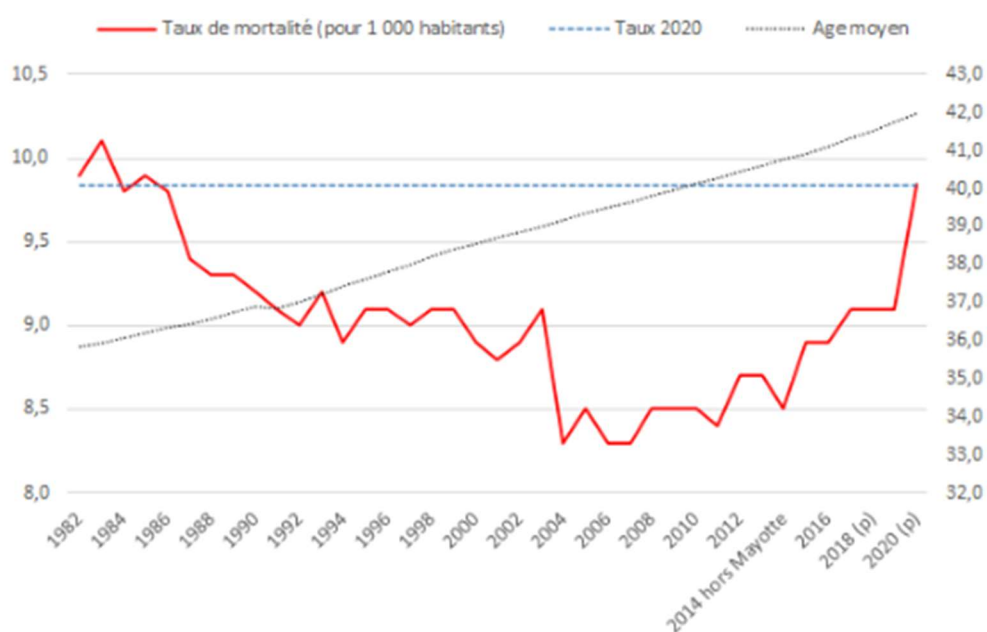


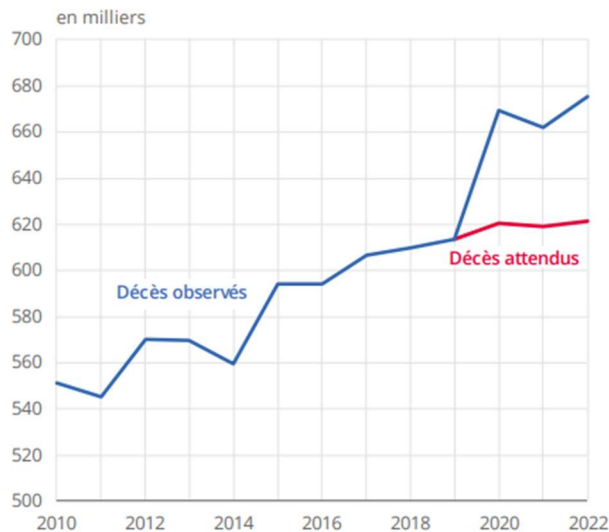
Figure 2-1 : Evolution du taux de mortalité entre 1980 et 2000.

Source : [Actu d'actuaires: Taux de mortalité en France depuis 1982 \(typepad.com\)](https://www.typepad.com/actu-dactuaires/taux-de-mortalite-en-france-depuis-1982)

▪ Une mortalité particulièrement élevée en France de 2020 à 2022

La pandémie Covid 19 explique partiellement l'écart de mortalité pour les années 2020 et 2021. En effet, les décès attribués au Covid-19 sont respectivement de 69 000 et 59 100 en 2020 et 2021. Aussi, en 2022, 38 300 décès ont été attribués au Covid. Selon l'INSEE, cela s'explique par une épidémie de

grippe particulièrement virulente et des épisodes de forte chaleur, mais admet que d'autres facteurs sont possibles, comme des reports d'opérations ou la baisse des dépistages d'autres maladies en 2020.



Note : données provisoires pour les décès attendus de 2020 à 2022 et observés en 2022.

Lecture : en 2022, 675 000 décès ont eu lieu. Le nombre de décès attendus est de 621 200, en prenant en compte l'augmentation et le vieillissement de la population, ainsi que la baisse des quotients de mortalité au rythme de la période 2010-2019.

Champ : France hors Mayotte jusqu'en 2013, France à partir de 2014.

Source : Insee, statistiques de l'état civil (extraction au 7 avril 2023) et estimations de population.

Figure 2-2 : Différence entre décès observés et décès attendus en France pour les années 2020 à 2022.

Si les quotients de mortalité par sexe et par âge avaient baissé entre 2020 et 2022 au même rythme que sur la période 2010-2019, 621 200 décès auraient eu lieu en 2022, soit 8 000 de plus qu'en 2019. En réalité, il y en a eu 67 500, soit 53 800 de plus qu'attendu.

Ce sont les décès dus à d'autres causes que la Covid-19 qui ont augmenté en 2022. Tout d'abord, l'année 2022 a compté deux épisodes de grippe. De plus, des épisodes de canicule ont occasionné davantage de décès en 2022 qu'en 2021, soit 200 personnes décédées en plus. Les températures élevées une grande partie de l'été ont pu aussi entraîner des décès en dehors des périodes de canicule.

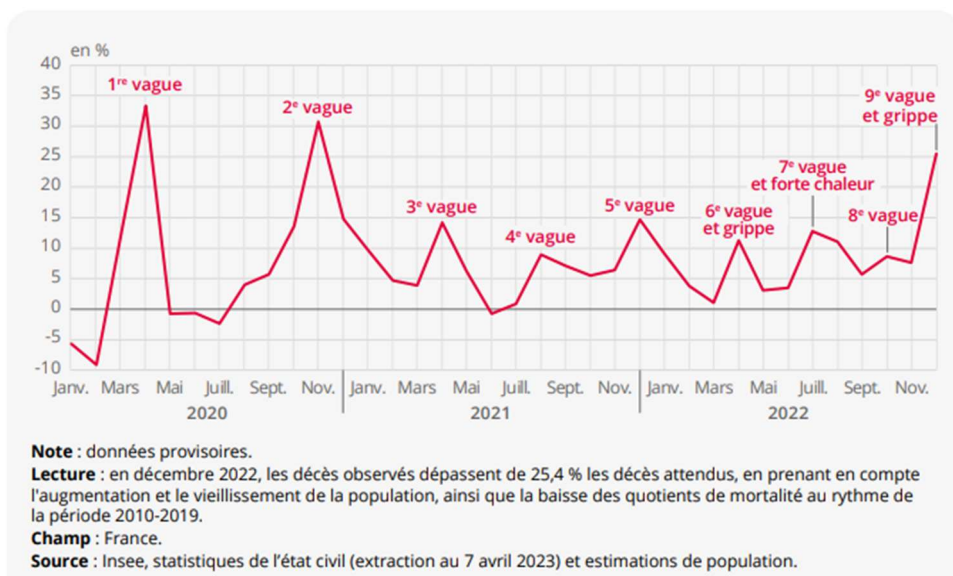


Figure 2-3 : Ecart entre les décès observés et attendus par mois de janvier 2020 à décembre 2022.

Tableau 2-1: Facteurs expliquant la hausse des décès en France depuis 2019.

	2020		Année	2021	2022
	Janvier-février	Mars-décembre		Année	Année
Décès observés en 2019	116 200	497 000	613 200	613 200	613 200
+Décès dus à l'augmentation et au vieillissement de la population	+2 400	+9 800	+12 200	+19 400	+28 600
+Décès dus à l'année bissextile en 2020	+1 900	///	+1 900	///	///
+Décès dus à l'évolution attendue des risques de décès	-3 200	-3 600	-6 800	-13 700	-20 600
=Décès attendus	117 300	503 200	620 500	618 900	621 200
+Décès dus à l'évolution inattendue des risques de décès	-8 600	+57 000	+48 400	+42 700	+53 800
=Décès observés	108 700	560 200	668 900	661 600	675 000

/// : absence de résultat due à la nature des choses.

Notes :

- données provisoires pour les décès attendus de 2020 à 2022 et observés en 2022 ;
- la répartition des décès attendus par mois est celle constatée en moyenne par sexe et âge entre 2010 et 2019.

Lecture : en 2022, 675 000 décès ont eu lieu en France, soit 53 800 de plus qu'attendus, en prenant en compte l'augmentation et le vieillissement de la population, ainsi que la baisse des quotients de mortalité au rythme de la période 2010-2019.

Champ : France.

Source : Insee, statistiques de l'état civil (extraction au 7 avril 2023) et estimations de population.

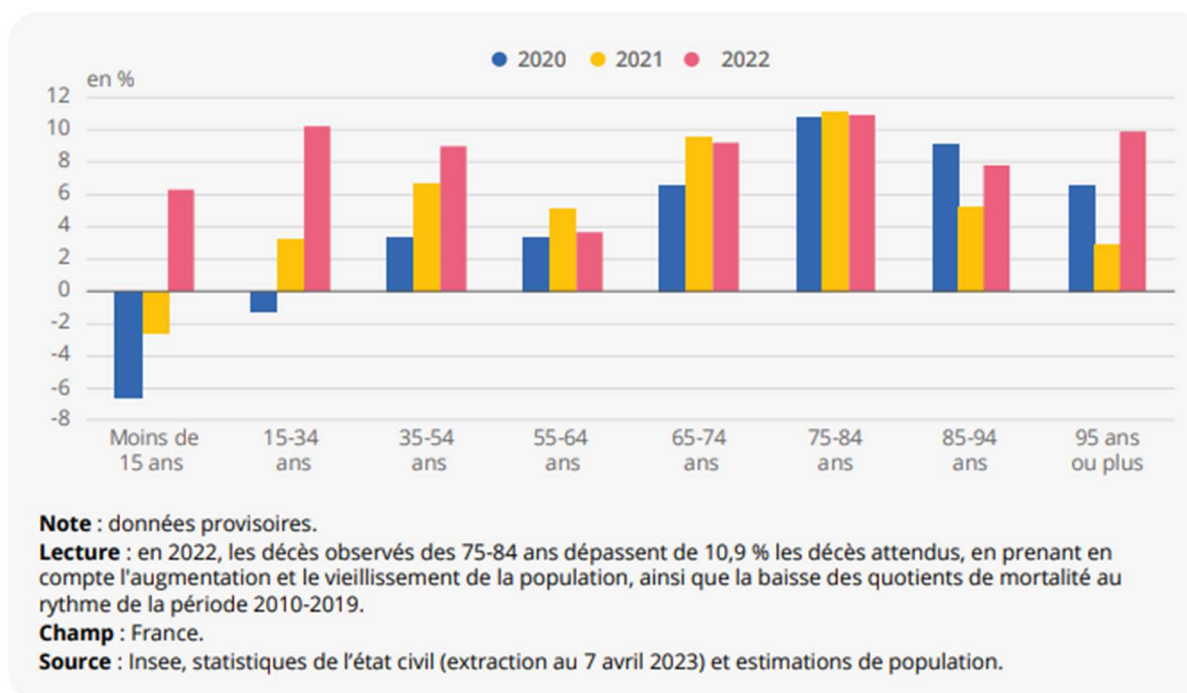


Figure 2-4 : Écart entre les décès observés et attendus par âge de 2020 à 2022.

En 2022, les décès observés sont supérieurs aux décès attendus pour toutes les classes d'âges, mais avec une surmortalité inégale qui varie de +4 % pour les 55-64 ans à +11 % pour les 75-84 ans. Par rapport à 2021, elle a augmenté avant 55 ans et après 85 ans.

Les figures suivants représentent en échelle logarithmique les taux de mortalité $\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}$ des hommes, des femmes et tous sexes confondus pour les années 2019, 2020 et 2021.

Où : $D_{x,t}$ représentent le nombre de décès l'année t et d'âge x et $E_{x,t}$ l'exposition au risque décès d'âge x l'année t.

Les données sont issues de la base « HMD ». Comme les taux croissent exponentiellement avec l'âge, on les représente sur une échelle logarithmique pour faciliter la comparaison aux différents âges.

Nous constatons qu'une mortalité infantile (de 0 à un an) élevée par rapport aux autres âges quel que soit le sexe de l'individu, suivie des faibles taux aux alentours de 10 à 12 ans. Aussi, on note une augmentation quasi linéaire du logarithme des taux de mortalité à partir de l'âge 40 ans.

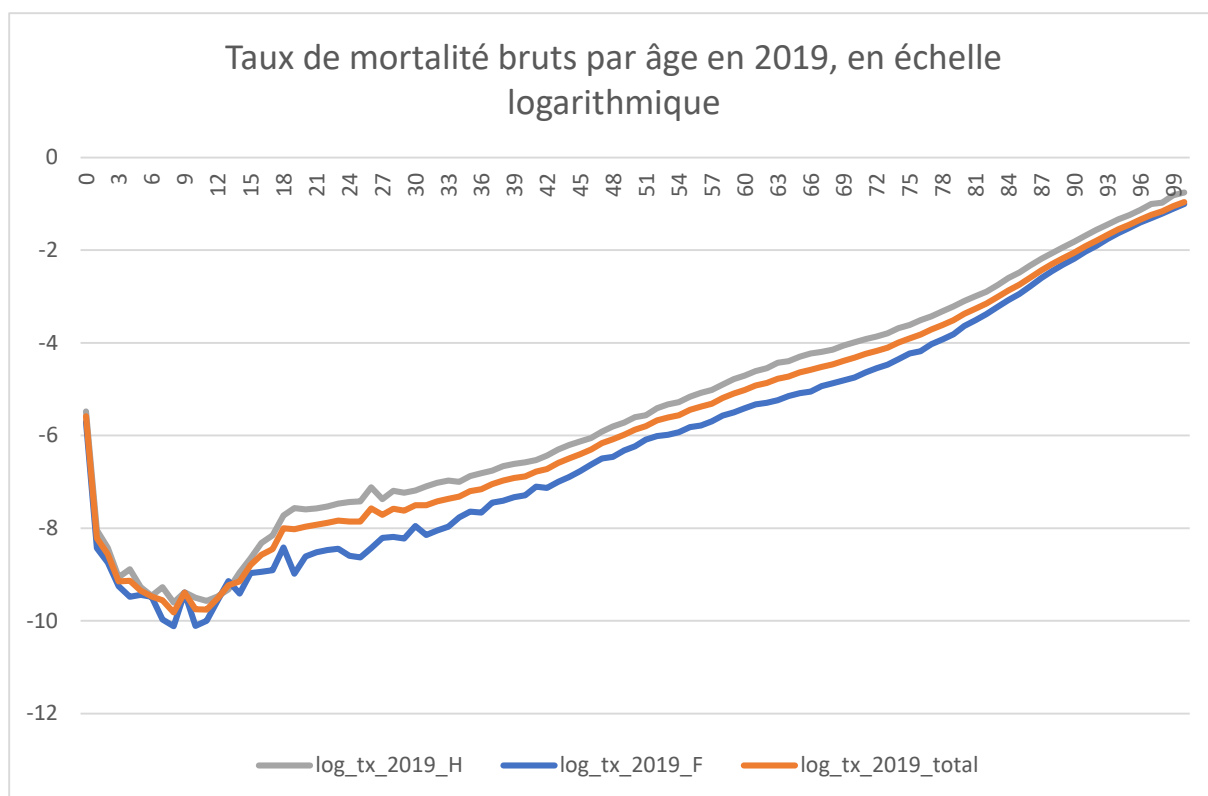


Figure 2-5 : Evolution de la mortalité par âge en France, tous sexes confondus, de 2019 à 2021.

Source : Figure construit à partir des données HMD.

Par ailleurs, on remarque que la mortalité des hommes est supérieure à celle des femmes presque à tous les âges et notamment à partir de 15 ans.

La courbe des logarithmes de tous sexes confondus, où on neutralise la variable sexe dans le calcul des taux de mortalité, se situe logiquement entre celle des hommes et des femmes.

On remarque aussi que les taux ne dépendent pas seulement de l'âge mais aussi de l'année calendaire t. et de l'année de naissance (la cohorte).

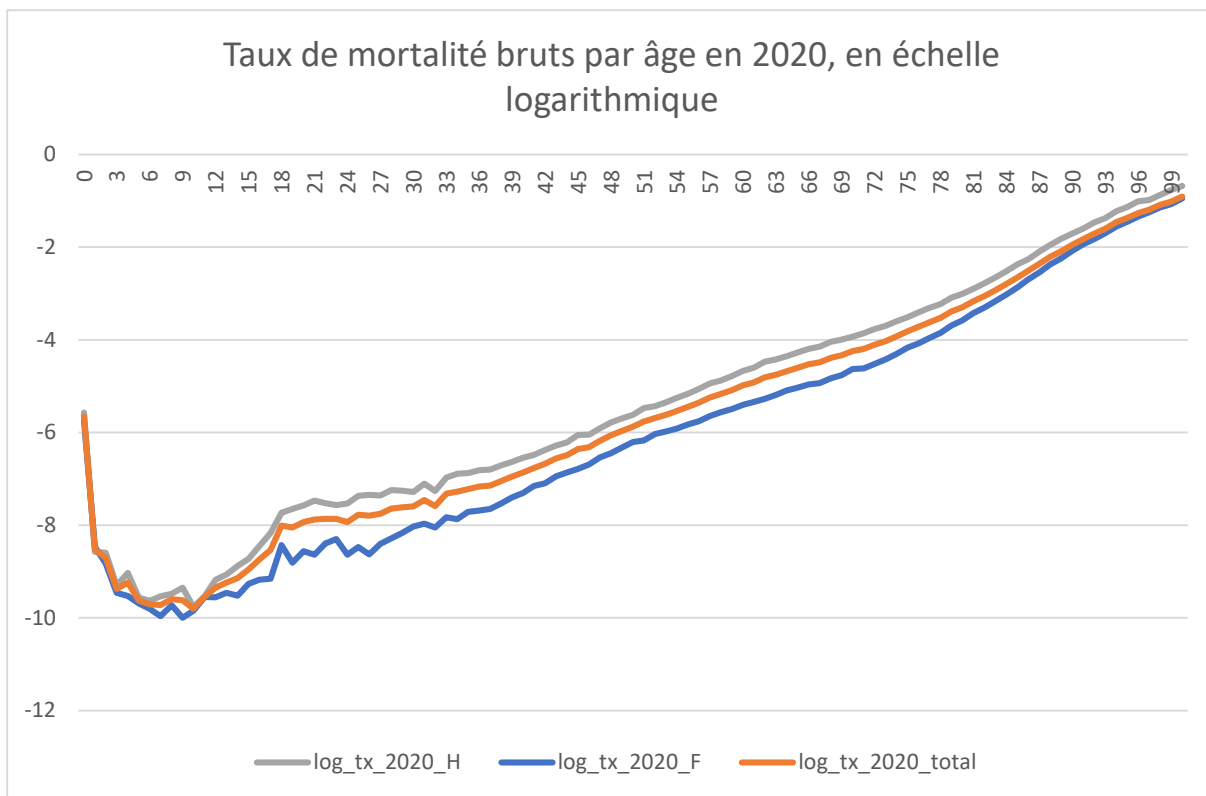


Figure 2-6 : Evolution de la mortalité par âge en France, de la population masculine, de 2019 à 2021.

Source : Figure construit à partir des données HMD.

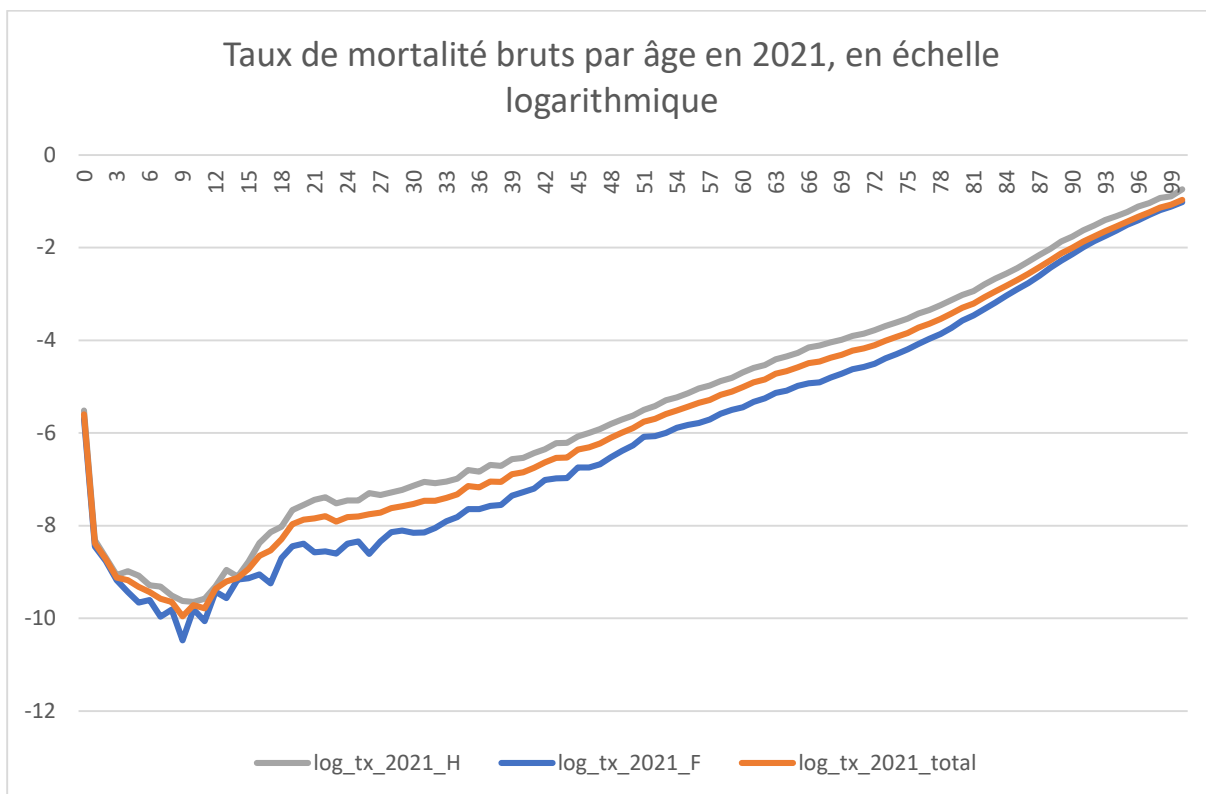


Figure 2-7 : Evolution de la mortalité par âge en France, de la population féminine, de 2019 à 2021.

Source : Figure construit à partir des données HMD.

- **La surmortalité masculine habituelle accrue avec la Covid-19**

La figure suivante nous montre une surmortalité des hommes par rapport aux femmes toutes causes de décès confondues. Elle est plus ou moins importante selon l'âge notamment aux âges de jeune adulte où les décès sont dus principalement aux suicides et accidents de la circulation, et entre 55 et 74 ans.

Aussi, on constate que pour une personne ayant entre 65 et 74 ans, le risque qu'elle meure de la Covid19 en 2020 et 2021 a été 2,4 fois plus élevé quand il s'agissait d'un homme plutôt qu'une femme ; en 2019 avant Covid19, le risque de mourir quelle qu'en soit la cause a été un peu plus du double pour un homme que pour une femme à ces âges.

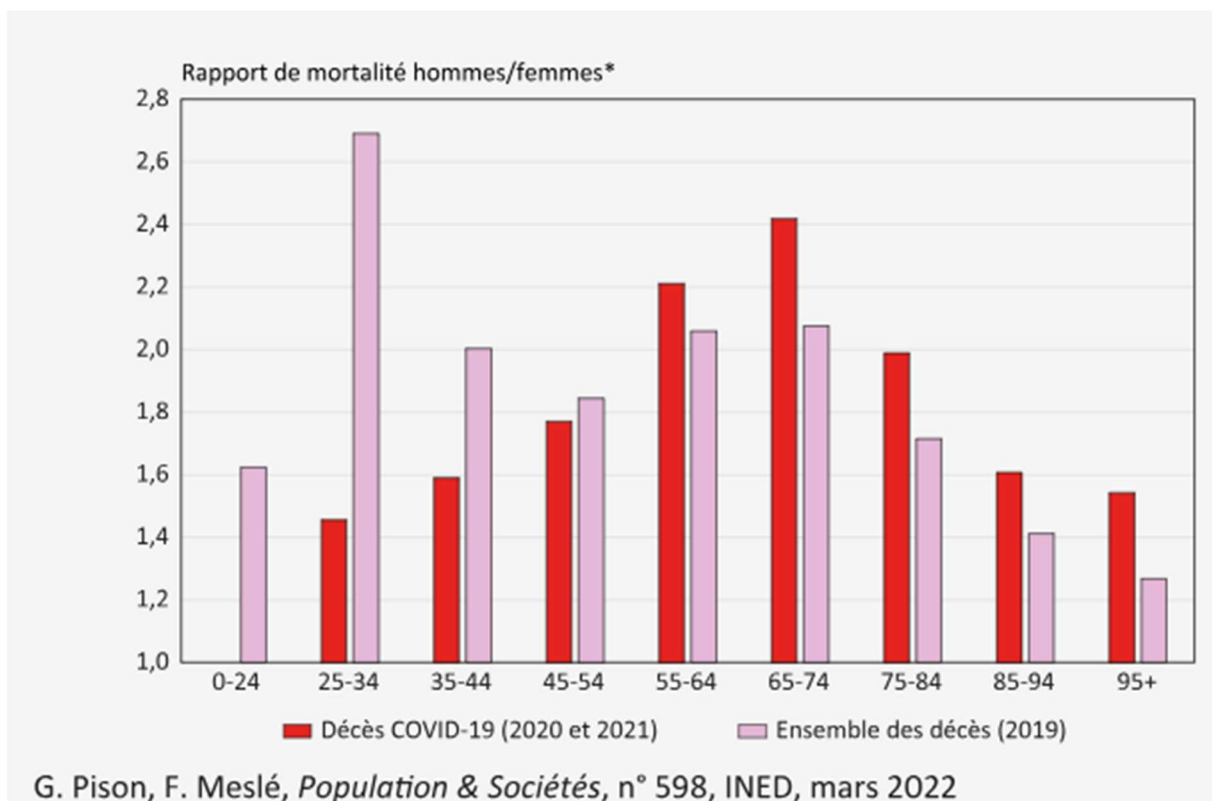


Figure 2-8 : Surmortalité des hommes par rapport aux femmes selon l'âge et la cause de décès.

2.3. Modèles Generalized Age-Period-Cohort

Les modèles Generalized Age-Period-Cohort (GAPC) sont une classe de modèles statistiques spécifiques utilisés pour analyser les variations des taux de mortalité en fonction de l'âge, de la période (l'année calendaire), et de la cohorte. Ces modèles visent à capter les effets combinés de ces trois dimensions dans les tendances de la mortalité.

Ici, par souci de concision, seront détaillés les résultats des trois premiers modèles : Lee-Carter (Lee, R. D. and Carter, L. R. (1992), Lee R. (2000), Planchet F., Lelieur Vincent (2007)) , Renshaw-Haberman (Renshaw, A. E. and Haberman, S. (2006)) et Modèle Age-Period-Cohort.

Pour calibrer les modèles évoqués précédemment, les paramètres sont estimés en maximisant la log-vraisemblance associée à l'aide de l'algorithme de Newton-Raphson. Tous les modèles sont calibrés en supposant que les décès suivent une loi de Poisson. La log vraisemblance associée s'écrit alors :

$$\log(\mathcal{L}(d_{x,t}, \hat{d}_{x,t})) = \sum_x \sum_t d_{x,t} \log(\hat{d}_{x,t}) - \hat{d}_{x,t} - \log(\hat{d}_{x,t}!) \quad (2-1)$$

Pour calibrer les modèles évoqués précédemment, on utilise le package R StMoMo (Villegas A. M. et al., 2018).

2.3.1. Modèle Lee-Carter

Le modèle de Lee-Carter (ci-après dénommé le modèle LC) est le modèle stochastique de mortalité le plus connu. Il a été introduit par Ronald Lee et Lawrence Carter en 1992. Le modèle LC décompose le logarithme du taux instantané de mortalité en deux composantes, dont l'une est propre à l'âge et l'autre tendancielle. A cela on ajoute en toute fin un bruit blanc. On écrit le modèle LC sous la forme suivante :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \varepsilon_{x,t} \quad (2-2)$$

Où :

$\mu_{x,t}$: est le taux de mortalité central des personnes ayant x ans au cours de l'année t. Il est calculé comme le rapport entre le nombre de personnes décédées et le nombre de personnes exposées au risque à la fois pour l'âge x et pour l'année t ;

α_x : décrit le comportement moyen du taux de mortalité central pour chaque âge. En outre, il garantit que la forme de la courbe de mortalité est conforme à l'expérience. En utilisant les contraintes applicables à l'estimation des paramètres, ce terme non paramétrique représente la moyenne arithmétique des $\mu_{x,t}$ sur toute la période observée ;

β_x : est le terme qui explique comment $\ln(\mu_{x,t})$ réagit à travers le temps. C'est un paramètre de sensibilité de la vitesse de réaction du taux de mortalité à k_t pour chaque âge. Le modèle permet d'obtenir des valeurs positives et négatives. Une valeur négative montre que le taux de mortalité pour un âge spécifique augmente avec le temps ;

k_t : représente l'indice temporel d'évolution de la mortalité dans l'espace-temps ;

$\varepsilon_{x,t}$: fait référence aux termes d'erreurs, à savoir les effets non saisis par le modèle.

Avec les contraintes suivantes pour résoudre les problèmes d'identifiabilité du modèle car si les vecteurs α , β , et κ sont solutions du problème de minimisation, alors, pour tout réel a et b, alors les estimateurs suivants $\tilde{\alpha}$, $\tilde{\beta}$ et $\tilde{\kappa}$ sont également solutions.

$$\begin{aligned}\tilde{\alpha}_x &= \alpha_x + b\beta_x \\ \tilde{\beta}_x &= \beta_x a \\ \tilde{k}_t &= a(k_t - b)\end{aligned}\tag{2-3}$$

Pour s'assurer de l'identifiabilité du modèle, il est nécessaire de rajouter des contraintes sur les paramètres k_t et β_x . Ces paramètres doivent être tels que : $\sum_t k_t = 0$ et $\sum_x \beta_x = 1$

L'avantage du modèle LC est sa simplicité d'utilisation ; néanmoins, il n'intègre pas l'effet de cohorte.

2.3.2. Modèle de Renshaw -Haberman

Renshaw et Haberman généralisent le modèle de Lee-Carter et incluent un effet cohorte.

La dynamique des taux de mortalité par âge et par année est donnée par :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \gamma_{t-x} + \varepsilon_{x,t}\tag{2-4}$$

γ_{t-x} : représente l'effet de cohorte qui est fonction de l'année de naissance $t - x$.

Avec les contraintes suivantes pour résoudre les problèmes d'identifiabilité du modèle :

$$\sum_t k_t = 0 \quad , \quad \sum_x \beta_x = 1 \quad \text{et} \quad \sum_{x,t} \gamma_{t-x} = 0\tag{2-5}$$

2.3.3. Modèle Age-Period-Cohort (APC)

Le modèle Age-Period-Cohort (ci-après dénommé le modèle APC), est une sous-structure du modèle RH. Il simplifie dans sa forme le modèle de RH en fixant la valeur du paramètre β : $\beta_x = 1$

Le modèle APC s'écrit sous la forme suivante :

$$\ln(\mu_{x,t}) = \alpha_x + k_t + \gamma_{t-x} + \varepsilon_{x,t}\tag{2-6}$$

Cette variante du modèle RH permet de résoudre certains problèmes d'instabilité du modèle de RH.

2.3.4. Modèle de Cairns-Blake-Dowd

Alors que plusieurs modèles se sont inspirés du modèle de référence de Lee-Carter, comme le modèle RH traité précédemment, d'autres chercheurs ont tenté de développer leur propre modèle. Le modèle de Cairns, Blake et Dowd (CBD) (Cairns et al. (2006a, 2006b, 2008), connu sous M5, en fait partie et considère les réalisations des probabilités de décès selon la structure suivante :

$$\text{logit}(q_{x,t}) = \log\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = k_t^1 + k_t^2 (x - \bar{x}) + \varepsilon_{x,t} \quad (2-7)$$

où :

\bar{x} est l'âge moyen sur la tranche d'âge de référence,

$q_{x,t}$ est la probabilité qu'un individu d'un âge x au temps t décèdera entre t et $t+1$, et elle se définit sous l'hypothèse d'uniformité des décès dans l'année par :

$$q_{x,t} = \frac{\mu_{x,t}}{1 + 0.5\mu_{x,t}} \quad (2-8)$$

k_t^1 et k_t^2 sont des paramètres évoluant dans le temps où :

Le paramètre k_t^1 représente le niveau général de la courbe de mortalité pour tous les âges et ressemble dans ce sens au Kappa du modèle de Lee-Carter. On en déduit qu'une réduction de ce paramètre implique une réduction globale de la mortalité, affectant les probabilités de décès de la même manière pour tous les âges.

Le paramètre k_t^2 exprime la pente de la courbe de mortalité. Un changement de k_t^2 reflète un changement dans la distribution de l'amélioration de la mortalité à travers les âges. Une diminution du paramètre signifie que la mortalité des âges plus avancés (au-dessus de l'âge moyen \bar{x}) s'améliore plus rapidement que pour les plus jeunes âges.

Contrairement au modèle de Lee-Carter, le modèle CBD (M5) n'a aucune contrainte sur ses paramètres.

Comme le modèle CBD (M5) n'a pas de contraintes, la méthode avec maximum likelihood et la méthode des moindres carrés ordinaires (OLS) donnent des paramètres similaires. La méthode OLS étant plus simple, on l'utilise pour l'estimation des paramètres qui consiste en une régression logistique du taux de mortalité par rapport à l'âge de l'individu. On considère que les coefficients de la régression sur l'âge sont stochastiques.

2.4. Résultats des modèles GAPC appliqués aux données de la population française

Plusieurs modèles peuvent être utilisés pour mettre en place des tables de mortalité prospectives sur la population générale, la référence étant le modèle de Lee-Carter. Dans le cas de la France, 4 modèles GAPC ont été comparés : le modèle de Lee-Carter, le modèle de Renshaw-Haberman, le modèle Age-Period-Cohort et le modèle de Cairns-Blake-Dowd.

Je détaillerai ici, uniquement les résultats du modèle LC et je mettrai en annexe les graphiques des autres modèles.

2.4.1. Mise en œuvre des modèles GAPC

Nous présenterons les résultats du modèle LC, appliqué à la population française avant et après covid pour les hommes, femmes et tous sexes confondus, de 1982 à 2019 ou 2021 et pour les âges allant de 50 à 90 ans.

A l'examen de la courbe des α_x , on observe qu'elle croît linéairement avec l'âge quel que soit le sexe et incluant ou pas les années marquées par la covid19 ; cela traduit l'augmentation du risque mortalité avec le vieillissement.

La courbe des κ_t quant à elle, ressemble à une courbe quasi-linéaire qu'on peut la rapprocher par une droite décroissante. Elle rend compte du déclin de la mortalité avec le temps. Un phénomène bien connu qui traduit une augmentation de l'espérance de vie et est imputable à de nombreux facteurs, parmi lesquels on pourrait citer les progrès de la médecine, des conditions de vie améliorées, une meilleure hygiène de vie, etc. Il faut toutefois souligner que le modèle ne cherche pas ici à justifier la raison de ce déclin de la mortalité. On se contentera dans le cadre de cette étude d'en extraire la tendance temporelle. On observe en particulier que jusqu'au début des années 1970, la tendance est assez chaotique, tout particulièrement chez les hommes, mais qu'ensuite elle devient beaucoup plus lisse. Chez les hommes, à nouveau de manière plus marquée, une rupture de pente correspondant à un déclin plus rapide de la mortalité s'observe également en 1970.

Enfin, la courbe des β_x met en avant le fait que ce déclin de la mortalité est inhomogène selon les âges. En effet, on peut clairement observer que ce recul est très marqué à partir de l'âge aux alentours des 70 ans mais augmente entre les âges 55 ans et 65 ans notamment chez les hommes ; cela s'explique par la mortalité précoce des hommes liée à la consommation de l'alcool et du tabac.

En effet, les rapports du Haut Conseil de Santé Publique ont mis en avant le niveau particulièrement élevé en France de la mortalité « prématurée » (décès avant 65 ans). Cette spécificité de la situation française apparaît très nettement lorsque l'on compare les risques de décès des hommes entre pays de niveau de développement comparable. Une « surmortalité prématurée » s'observe également, mais dans une moindre mesure, pour les femmes. Au contraire, après 65 ans, la situation française apparaît favorable par rapport aux autres pays.

Les causes de décès responsables de cette surmortalité sont très spécifiques : suicide, alcoolisme, accidents de la circulation, autres types d'accident, cancers des voies aérodigestives supérieures et cancers du poumon. Ci-dessous figurent les graphiques des paramètres du modèle LC pour les hommes, femmes, tous sexes confondus, avant et y compris Covid

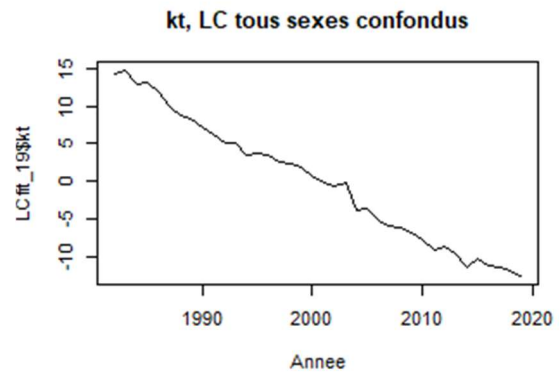
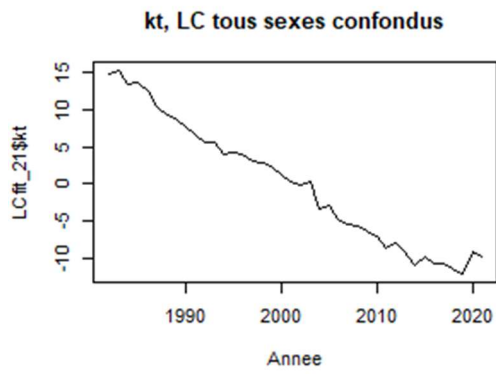
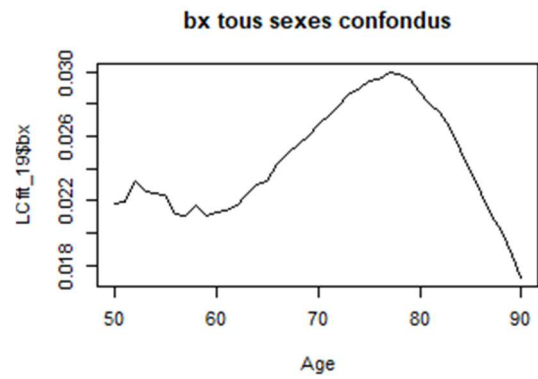
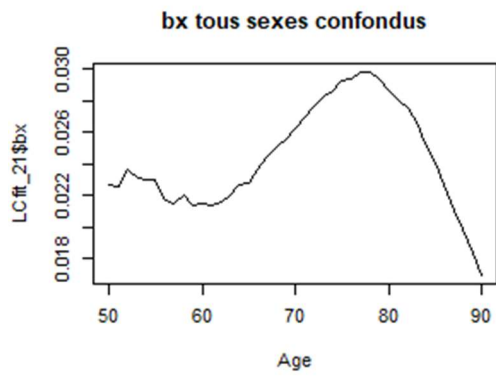
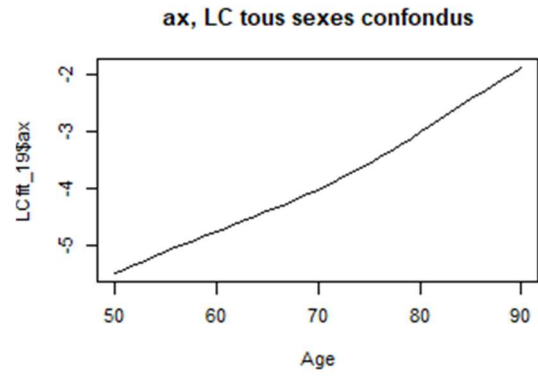
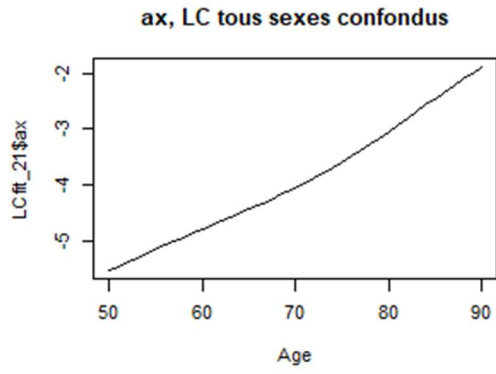


Figure 2-9 : Paramètres du modèle LC pour tous sexes confondus.

(a) y compris Covid -tous sexes confondus

(b) avant Covid- tous sexes confondus

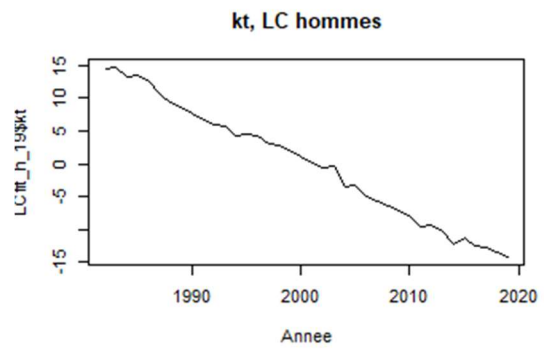
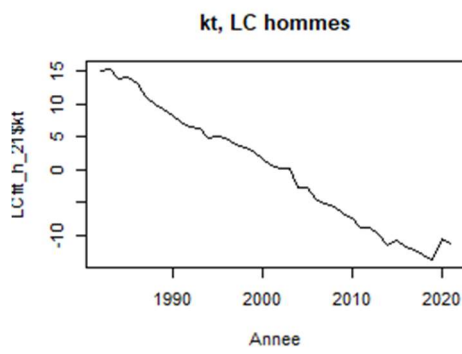
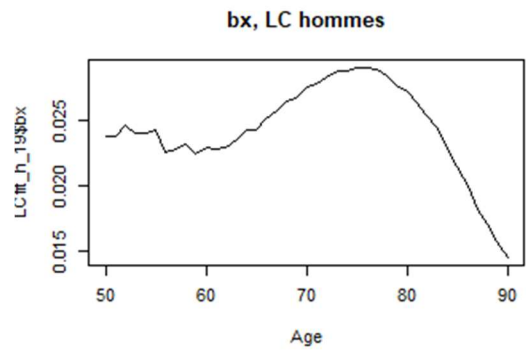
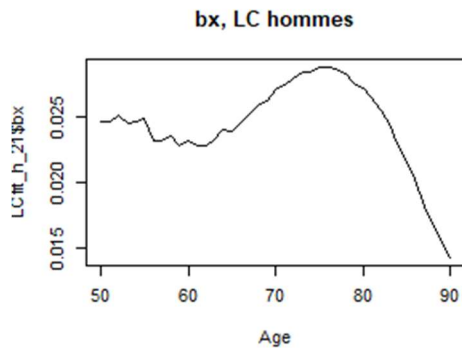
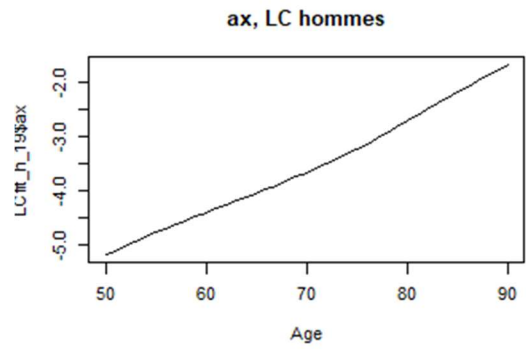
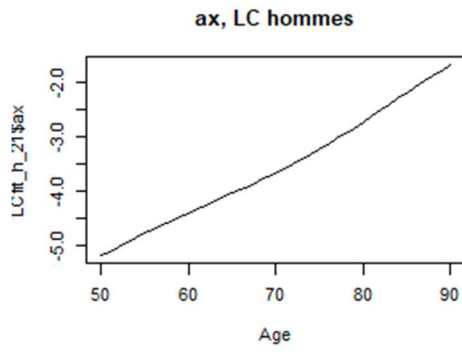


Figure 2-10 : Paramètres du modèle LC pour la population masculine.

(c) Y compris Covid -hommes

(d) avant Covid-hommes

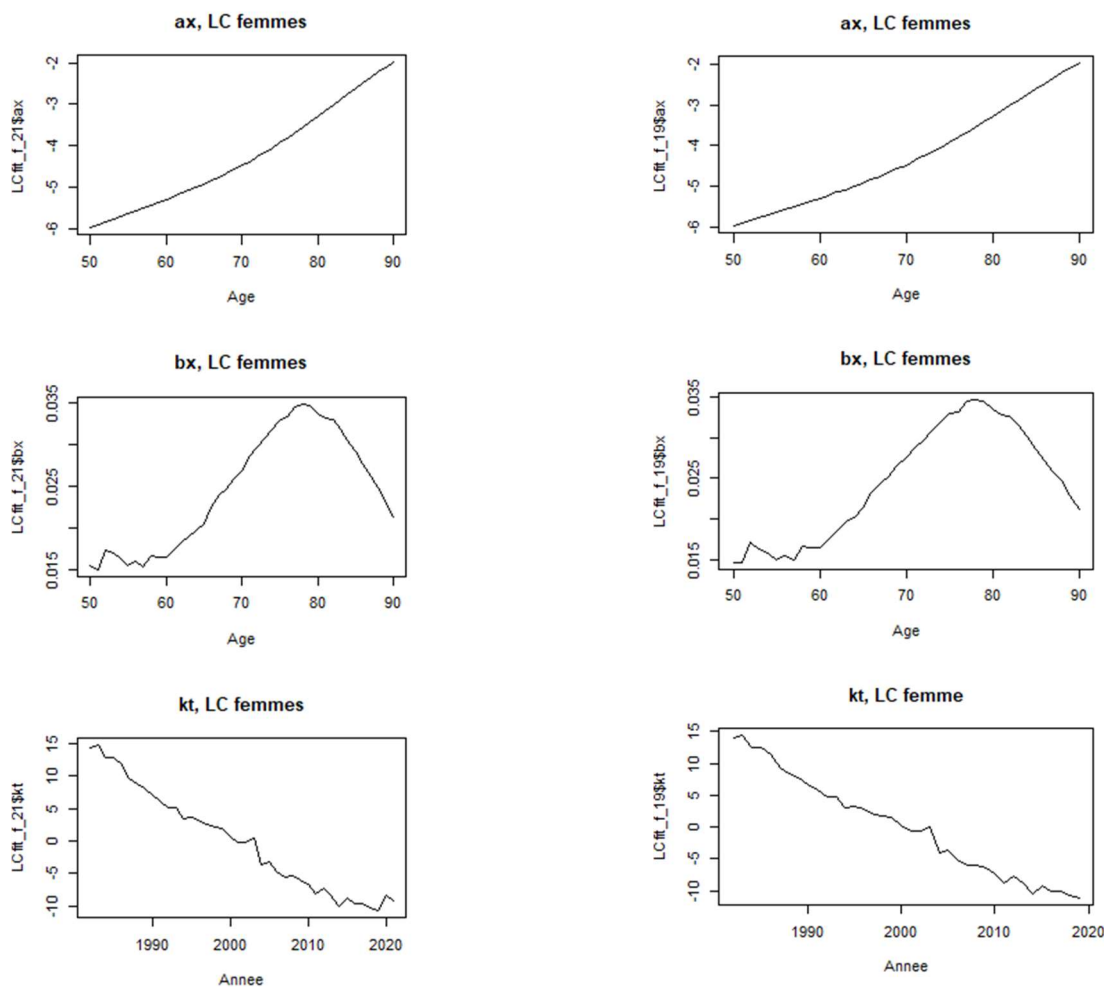


Figure 2-11 : Paramètres du modèle LC pour la population féminine.

(e) Y compris Covid - Femmes

(f) Avant Covid - Femmes

2.4.2. Indicateurs AIC, BIC et nombre de paramètres

Pour comparer les différents modèles, on examine d'abord les critères de vraisemblance pénalisée : l'AIC (Critère d'Information d'Akaike) et le BIC (Critère d'Information Bayésien). Ces indicateurs associent au critère du maximum de vraisemblance, une pénalisation par le nombre de paramètres du modèle. Ainsi, ils permettent de trouver un compromis entre complexité et qualité du modèle.

Le tableau ci-dessous fournit les valeurs du nombre de paramètres, de l'AIC et du BIC obtenus pour les 4 modèles GAPC étudiés sur la période 1982-2021 (y compris covid) et 1982-2019 (avant covid) pour la tranche d'âge 50-90 ans.

Tableau 2-2 : Valeurs du nombre de paramètres, de l'AIC et du BIC obtenus pour les 4 modèles GAPC.

Modèle	Type de population	AIC	BIC	Nombre de paramètres
LC	Total_yc_covid	26 641	27 289	122
	H	25 368	26 015	122
	F	20 728	21 375	122
	Total_avant_covid	24 478	25 109	120
	H	23 579	24 209	120
	F	19 179	19 809	120
RH	Total_yc_covid	20 441	21 482	202
	H	19 230	20 271	202
	F	18 344	19 385	202
	Total_avant_covid	19 250	20 260	198
	H	18 200	19 210	198
	F	17 343	18 354	198
APC	Total_yc_covid	27 317	28 171	161
	H	22 740	23 594	161
	F	22 991	23 844	161
	Total_hors_covid	25 494	26 318	157
	H	21 231	22 055	157
	F	21 509	22 333	157
CBD	Total_avant_covid	296 630	297 062	79
	H	127 462	127 894	79
	F	200 186	200 618	79
	Total_avant_covid	278 136	278 543	75
	H	118 660	119 067	75
	F	187 286	187 693	75

Le modèle de Renshaw-Haberman semble être le modèle le plus performant pour les hommes, alors que le modèle LC est mieux adapté pour les femmes. Malgré le nombre important de paramètres utilisés, ces modèles ont le plus faible BIC, ce qui témoigne d'un bon compromis entre complexité et qualité du modèle.

Par ailleurs, pour évaluer la qualité des modèles, il est nécessaire d'effectuer une analyse des résidus associés. Si les résidus présentent une structure régulière, cela indique que le modèle est incapable de décrire correctement les caractéristiques des données. Le package StMoMo permet d'obtenir les résidus de déviance normalisés $r_{x,t}$ définis par :

$$r_{x,t} = \text{sign}(d_{x,t} - \hat{d}_{x,t}) \sqrt{\frac{\text{dev}(x,t)}{\hat{\phi}}} \quad (2-9)$$

$$\text{où} \begin{cases} \text{dev}(x,t) = 2(d_{xt} \log(\frac{d_{xt}}{\hat{d}_{xt}}) - (d_{xt} - \hat{d}_{xt})) \\ \hat{\phi} = \frac{1}{N-k} \sum_x \sum_t \text{dev}(x,t) \end{cases}$$

avec N le nombre d'observations et k le nombre de paramètres. Les graphiques ci-dessous permettent de visualiser les résidus obtenus pour les différents modèles testés.

2.4.3. Analyse des résidus

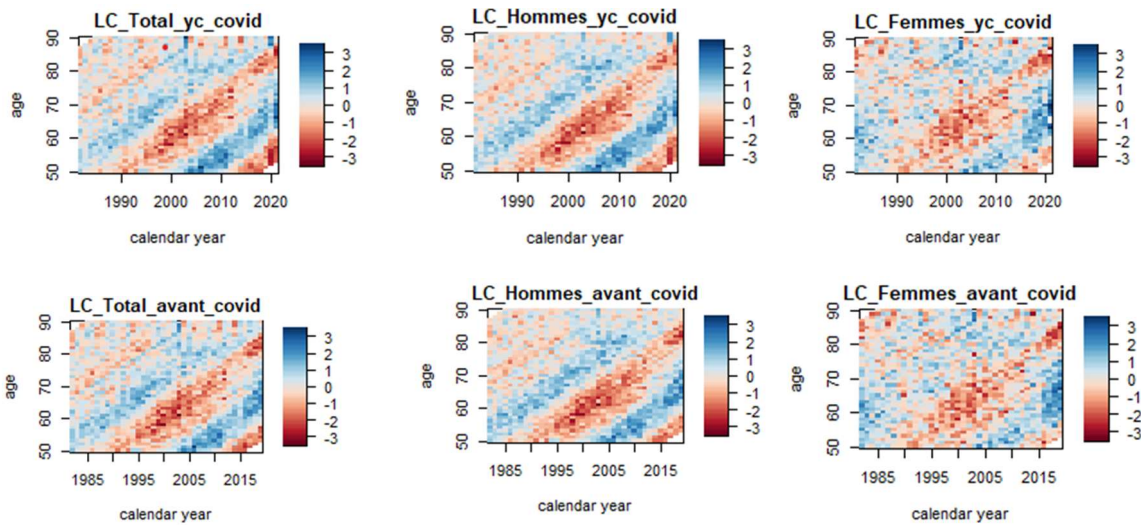


Figure 2-12 : Résidus du modèle LC, pour les hommes, femmes, tous sexes confondus avant et y compris covid.

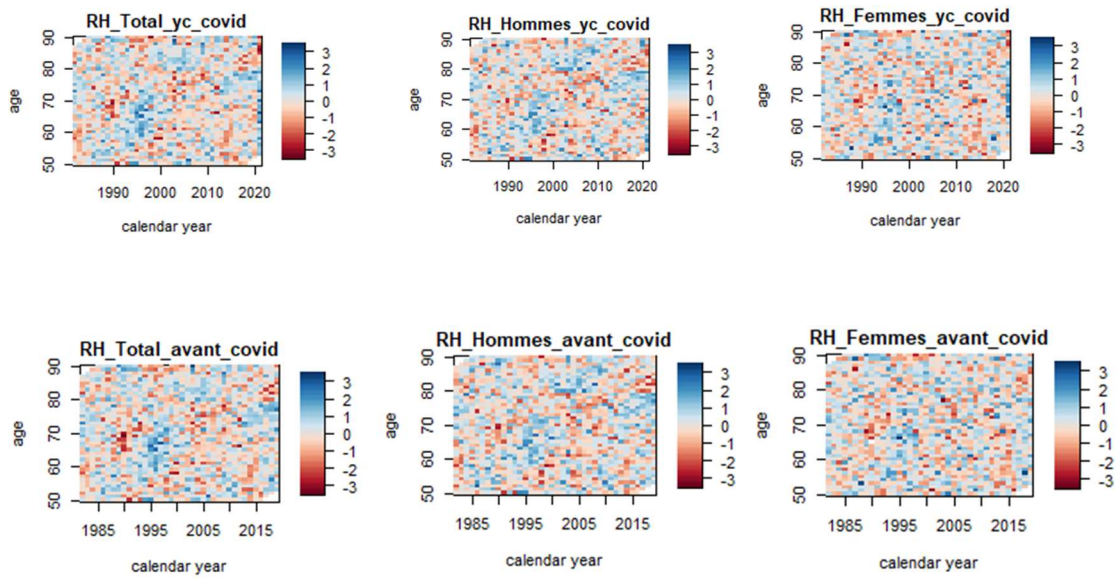


Figure 2-13 : Résidus du modèle RH, pour les hommes, femmes, tous sexes confondus avant et y compris covid.

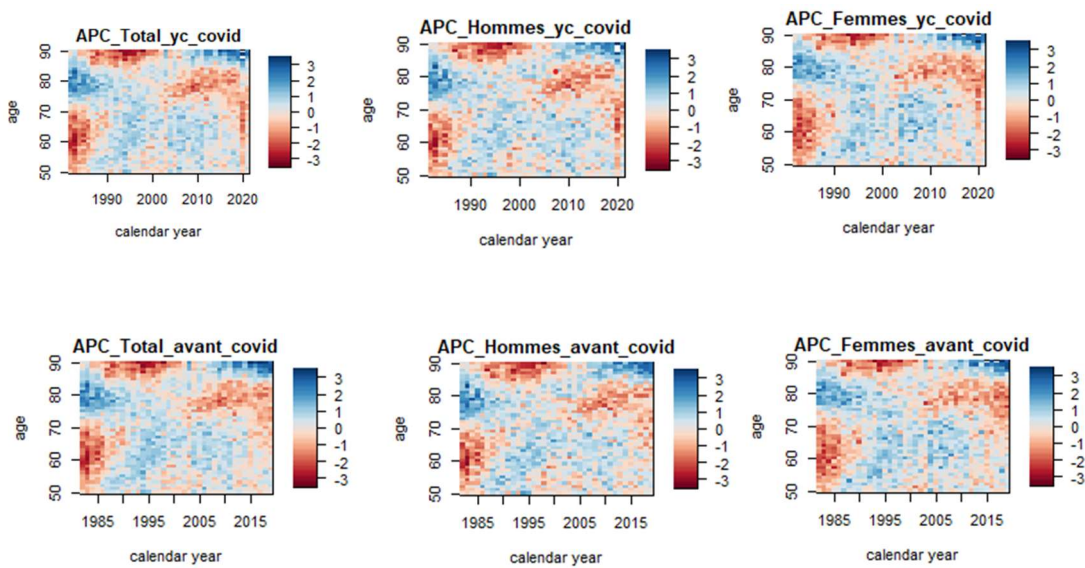
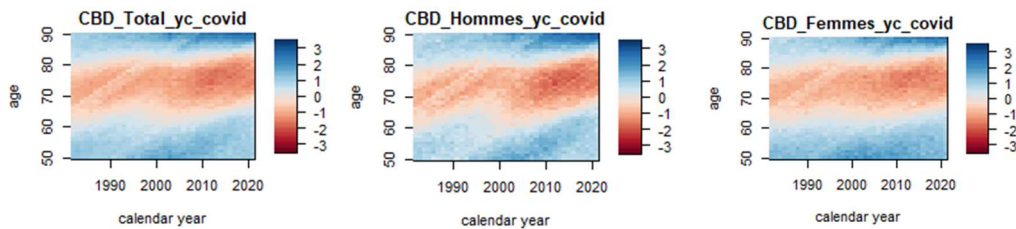


Figure 2-14 : Résidus du modèle APC, pour les hommes, femmes, tous sexes confondus avant et y compris covid.



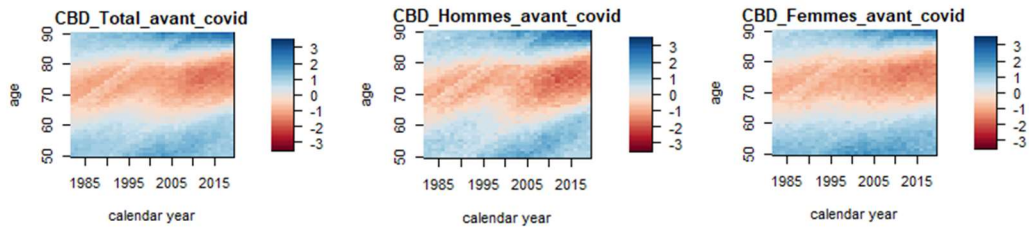


Figure 2-15 : Résidus du modèle CBD, pour les hommes, femmes, tous sexes confondus avant et y compris covid.

Les différentes figures montrent que les résidus des modèles de LC, APC et CBD possèdent une structure non-aléatoire, contrairement au modèle de Renshaw-Haberman pour lequel les résidus seraient approximativement aléatoires. Néanmoins, le modèle RH ne converge pas dans le cas de la population féminine. Les modèles LC et CBD, qui n'intègrent pas d'effet cohorte, présentent des structures diagonales très marquées, indiquant l'incapacité de ces modèles à décrire les effets cohorte caractéristiques de la population française. Le modèle APC présente lui des clusters de résidus que l'on peut expliquer par la non-prise en compte des améliorations de mortalité qui varient avec l'âge.

A l'issue de ces analyses des graphiques des résidus et les indicateurs AIC et BIC, nous ne sommes pas en mesure d'opter pour un modèle qui soit convenable pour les hommes, femmes et tous sexes confondus avant covid 19 ou y compris Covid.

D'où l'utilité d'explorer d'autres approches en s'inspirant de l'article publié par Barigou K. et al. (2022) : "Baysien model averaging for mortality forecasting using leave-future-out-validation" avec le package R STANMOMO.

2.5. Approche bayésienne dans la construction des tables de mortalité de référence

2.5.1. Différence entre approche fréquentiste et approche bayésienne

L'approche fréquentiste repose sur la théorie des probabilités où la probabilité d'un événement est basée sur la fréquence à laquelle cet événement se produit dans un grand nombre d'essais ou d'échantillons. Dans cette approche, les paramètres à estimer sont considérés comme fixes et les données sont utilisées pour les estimer.

Autrement dit, l'approche fréquentiste repose sur une vision déterministe des paramètres des modèles probabilistes qui sont les objets que l'inférence statistique cherche à estimer. Les estimations par le maximum de vraisemblance font partie des outils fondamentaux de la statistique fréquentiste, tout comme la théorie des tests statistiques avec le concept d'intervalle de confiance qui lui est associé.

La différence fondamentale entre l'approche fréquentiste et l'approche bayésienne est donc que cette dernière considère les paramètres non pas comme fixes (i.e. pour lesquels il existe une vraie valeur), mais plutôt comme des variables aléatoires.

Cette manière de considérer les paramètres comme des variables aléatoires induit une distribution marginale $\pi(\theta)$. Cette distribution est appelée a priori (ou parfois prior en anglais).

La loi a priori est construite à partir des informations disponibles sur le paramètre θ avant de collecter des données. L'information (dite a priori) provient d'avis d'expert, de résultats d'expériences précédentes dont les résultats sont supposés similaires.

Notons y les observations dont nous disposons. De la même manière que dans un modèle fréquentiste, une modélisation bayésienne paramétrique consiste d'abord à proposer un modèle probabiliste, qu'on appelle « modèle d'échantillonnage » pour ces observations :

$$Y_i \stackrel{iid}{\sim} f(y|\theta) \tag{2-10}$$

Dans la modélisation bayésienne, par rapport à la modélisation fréquentiste, on ajoute une loi de probabilité (définie sur l'espace Θ des paramètres), appelée distribution a priori :

$$\begin{aligned} \theta &\sim \pi(\theta) \\ Y_i|\theta &\stackrel{iid}{\sim} f(y|\theta) \end{aligned} \tag{2-11}$$

On va donc traiter θ comme une variable aléatoire, mais qui n'est jamais observée.

A partir du modèle bayésien, on obtient une loi de probabilité sur le paramètre : la loi a posteriori cette loi résume l'information provenant des données et de l'information a priori.

L'objet d'une telle modélisation bayésienne est la distribution des paramètres a posteriori, c'est-à-dire la loi de θ conditionnellement aux observations : $p(\theta|Y = y)$, appelée distribution a posteriori. Elle se calcule à partir du modèle d'échantillonnage $f(y|\theta)$ — à partir duquel on obtient la vraisemblance $f(y|\theta)$ pour toutes les observations — et de la loi a priori $\pi(\theta)$ par le théorème de Bayes :

$$p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{f(\mathbf{y})} \tag{2-12}$$

$$\text{où } f(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta) d\theta \text{ est la loi marginale de } \mathbf{Y}. \tag{2-13}$$

Tableau 2-3 : Comparaison entre les deux approches fréquentiste et bayésienne.

Aspect	Approche fréquentiste	Approche bayésienne
Paramètres	Considérés comme fixes et inconnus, à estimer à partir des données.	Considérés comme des variables aléatoires avec des distributions de probabilité.
Incertitude	Traite l'incertitude via des estimations ponctuelles et des intervalles de confiance.	Intègre explicitement l'incertitude via les distributions a priori et a posteriori.
Données	Se fonde uniquement sur les données observées pour l'estimation des paramètres.	Combine les données observées avec des informations a priori pour produire une estimation a posteriori.
Intervalles	Intervalles de confiance pour estimer la précision des paramètres.	Intervalles de crédibilité ² basés sur la distribution a posteriori.
Flexibilité	Moins flexible, difficile d'intégrer des informations extérieures.	Plus flexible, permet l'intégration de connaissances a priori et la modélisation de données rares ou incertaines.
Complexité	Relativement simple à mettre en œuvre avec des outils classiques comme le maximum de vraisemblance.	Plus complexe, nécessite souvent des méthodes de simulation comme Markov Chain Monte Carlo (MCMC) pour l'estimation.

Un point essentiel de l'approche bayésienne est donc de donner une distribution aux paramètres. Dans l'inférence bayésienne, on part d'une distribution a priori, et l'information contenue dans les observations est utilisée pour obtenir la distribution a posteriori. La distribution a priori apporte de la flexibilité par rapport à un modèle fréquentiste, en permettant d'incorporer dans le modèle de la connaissance externe. Cela peut par exemple permettre de résoudre des problèmes d'identifiabilité parfois rencontrés par une approche purement fréquentiste lorsque l'information apportée par les observations ne suffit pas pour estimer tous les paramètres d'intérêt. C'est donc un grand avantage de l'approche bayésienne. Mais d'un autre côté, le choix de cette distribution a priori des paramètres introduit une subjectivité intrinsèque dans l'analyse, qui peut être décriée.

Comme il l'a dit justement Gilbert Saporta "Être ou ne pas être bayésien, là n'est plus la question : il s'agit d'utiliser à bon escient les outils adaptés quand cela est nécessaire."

² lien avec les intervalles de confiance : un intervalle de confiance de niveau $(1 - \alpha)$ est un intervalle aléatoire $[a(X), b(X)]$ tel que, pour tout θ , $P_{\theta}([a(X), b(X)] \ni \theta) = 1 - \alpha$

$(1 - \alpha)$ % des intervalles de confiance contiennent la vraie valeur du paramètre .

un intervalle de crédibilité de niveau $(1 - \alpha)$ est un intervalle $[l(X), u(X)]$ tel que $P(\theta \in [l(x), u(x)] | X) = 1 - \alpha$

Ayant observé X , le paramètre appartient à l'intervalle avec une probabilité $(1 - \alpha)$.

2.5.2. Mise en œuvre de l'approche bayésienne dans le cadre de la prédiction de la mortalité en utilisant « leave-future-out validation »

Dans une approche fréquentiste, nous estimons d'abord l'indice temporel k_t puis on l'extrapole dans le futur à l'aide d'un processus ARIMA. Cette procédure en deux temps peut être la source d'une certaine incohérence (Delwarde et Denuit, 2006). Pour éviter cet écueil, il est possible d'intégrer ces deux étapes dans un modèle bayésien. Ce dernier considère tous les paramètres inconnus comme des variables aléatoires dont la distribution conditionnelle provient de l'information connue : les décès et les expositions.

Nous nous inspirons largement, dans ce chapitre, de l'article cité ci-dessous ; il s'agit d'une mise en application de la méthode développée dans cet article pour choisir le modèle le plus performant permettant de produire des projections pertinentes des quotients de mortalité à long terme pour la population française.

De manière synthétique, cet article rappelle les risques de s'appuyer sur un modèle unique dans le cadre des projections des taux de mortalité notamment une mauvaise spécification du modèle, l'incertitude des paramètres et le surajustement. Pour y remédier, les auteurs étudient la modélisation de la mortalité dans un modèle bayésien binomial négatif. Ce cadre permet de tenir compte de la surdispersion et de l'incertitude concernant les estimations des paramètres. Ils considèrent que les techniques de moyenne des modèles sont une réponse aux erreurs de spécification du modèle. Dans cet article, ils proposent deux méthodes (Stacking et pseudo-BMA) basées sur « leave-future-out_validation » en laissant de côté les M dernières années de données et évaluant les performances de prédiction sur ces dernières M années). Les résultats de la mise en application de ces modèles sont comparées à ceux issus du modèle bayésien moyen (BMA).

Le Stacking fait partie des modèles ensemblistes. Ces derniers sont basées sur l'hypothèse que la combinaison de plusieurs modèles peut souvent produire un modèle beaucoup plus puissant. Autrement dit, la méthode ensembliste est un paradigme d'apprentissage automatique dans lequel plusieurs modèles de base sont formés pour résoudre le même problème et combinés pour obtenir de meilleurs résultats. L'hypothèse principale est que lorsque les modèles de base sont correctement combinés, nous pouvons obtenir des modèles plus précis et/ou robustes.

Pour pouvoir "résoudre" un problème, nous voulons que notre modèle ait suffisamment de degrés de liberté pour résoudre la complexité sous-jacente des données avec lesquelles nous travaillons, mais nous voulons aussi qu'il n'ait pas trop de degrés de liberté pour éviter une grande variance et être plus robuste. C'est le compromis bien connu biais-variance.

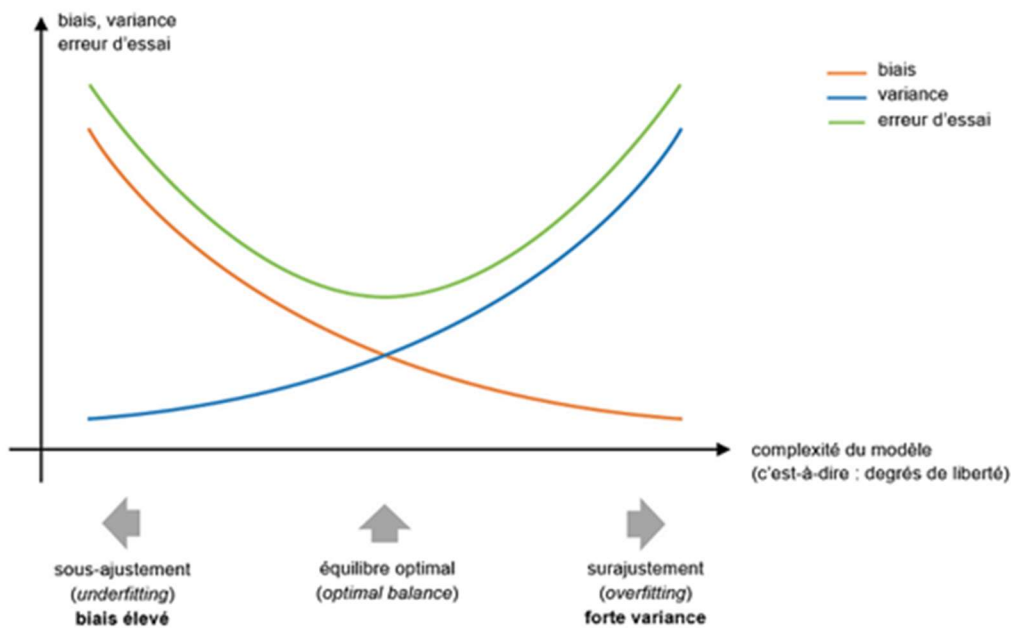


Figure 2-16 : Illustration de overfitting et underfitting.

Afin de mettre en place un modèle ensembliste, nous devons d’abord sélectionner nos modèles de base à agréger. Notre choix de ces modèles de base doit être cohérent avec la façon dont nous agrégeons ces modèles. Si on choisit des modèles de base à faible biais mais à forte variance, ce doit être avec une méthode d’agrégation qui tend à réduire la variance alors que si l’on choisit des modèles de base à faible variance mais à fort biais, ce doit être avec une méthode d’agrégation qui tend à réduire le biais. Par exemple, le stacking apprend, en parallèle et combine les modèles de base en un méta-modèle pour produire une prédiction basée sur les prédictions des différents modèles de base.

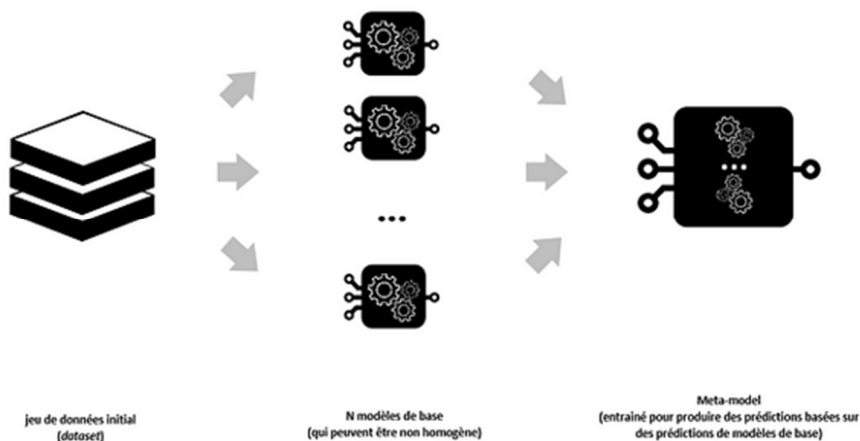


Figure 2-17 : Illustration de déroulement d’un modèle de Stacking.

D’une certaine façon, il s’agit de prédire quels sont les meilleurs modèles de base et de les pondérer. Cette démarche a l’avantage de pouvoir agréger des modèles très différents et d’améliorer sensiblement la qualité de la prédiction finale

- **Modèles bayésiens moyens : BMA, Stacking et Pseudo-BMA**

Chaque modèle parmi les modèles de base testés ci-dessus (LC, RH, APC, BCD) a ses points forts (capturer des tendances spécifiques, modéliser les cohortes, etc.) et ses faiblesses (simplification excessive, surajustement, etc.), ce qui peut conduire à des erreurs de prédiction lorsqu'ils sont utilisés individuellement. De plus, la mortalité est influencée par des facteurs complexes qui ne peuvent pas être capturés de manière optimale par un seul modèle.

Les méthodes BMA, stacking, et pseudo-BMA offrent un cadre de combinaison de modèles qui permet d'améliorer les performances des projections de tables de mortalité par rapport à l'utilisation de chaque modèle (LC, RH, APC, BCD, M6) individuellement. Ces approches permettent de tirer parti des forces de chaque modèle tout en atténuant leurs faiblesses respectives.

Trois types de modèles seront abordés : les modèles bayésiens moyens (BMA), le stacking bayésien et le pseudo-BMA. Il s'agit de combiner plusieurs modèles de prévision de mortalité (comme LC, RH, APC, BCD et M6) afin d'obtenir des projections plus robustes et précises. Ces méthodes visent à exploiter les forces de chaque modèle tout en réduisant leurs faiblesses individuelles. Elles offrent des avantages complémentaires pour la projection de la mortalité, en particulier dans des contextes où plusieurs modèles de prévision sont en concurrence.

- **Bayesian Model Averaging (BMA)**

Au lieu de choisir un modèle unique, le BMA intègre les prédictions de plusieurs modèles, ce qui réduit l'incertitude liée à la sélection du modèle. Il utilise une moyenne pondérée des prédictions de tous les modèles de prédiction de la mortalité. Le BMA repose sur le théorème de Bayes. Chaque modèle M_k est pondéré par sa probabilité a posteriori .

$$P(M_k/data) = \frac{P(data/M_k) \times P(M_k)}{\sum_k P(data/M_k) \times P(M_k)} \quad (2-14)$$

$$p(M_k / y) = \frac{p(y/M_k) \times p(M_k)}{\sum_{k=1}^K p(y/M_k) \times p(M_k)} \quad (2-15)$$

Où :

- $P(M_k / y)$ est la probabilité a posteriori du modèle M_k
- $P(y / M_k)$ est la vraisemblance marginale

$$p(y/M_k) = \int_{\theta} p(y/\theta_k, M_k) \times p(\theta_k/M_k) d\theta_k \quad (2-16)$$

- $P(M_k)$ est la probabilité à priori du modèle M_k

Chaque modèle est pondéré par sa probabilité a posteriori pour obtenir une moyenne pondérée des prédictions :

$$\hat{y} = \sum_k (P(M_k / \text{data}) \hat{y}_k) \quad (2-17)$$

Où \hat{y}_k est la prédiction du modèle M_k

Ici, dans le cadre de projections des taux de mortalité , le BMA attribue des poids à chaque modèle parmi les modèles (LC, RH, APC, BCD, M6) en fonction de sa performance historique pour les prédictions de mortalité passées.

Le poids affecté à chaque modèle est :

$$w_k = \frac{P(y/M_k)}{\sum_{k=1}^K P(y/M_k)P(M_k)} P(M_k) \quad (2-18)$$

Le résultat final est une projection de mortalité qui intègre l'incertitude sur le choix du meilleur modèle, en offrant une combinaison optimisée des différents modèles.

Le BMA réduit le risque de sélectionner un mauvais modèle et produit des prédictions plus robustes en exploitant les performances relatives des différents modèles. Néanmoins, il peut être coûteux en termes de calculs, car il nécessite une évaluation complète des modèles candidats et des calculs de vraisemblance.

Par ailleurs, BMA est très sensible aux choix a priori et tend à sélectionner un seul modèle de manière asymptotique. De plus, comme le Bayes Information Criterion (BIC), le BMA mesure dans quelle mesure le modèle s'adapte au passé, mais pas dans quelle mesure le modèle prédit l'avenir.

Nous proposons deux approches alternatives de moyenne de modèles, appelées Stacking et Pseudo-BMA, basés sur le Leave-Future-Out validation.

Dans ces deux approches, on divise la base de données en trois parties : une partie pour l'entraînement, une deuxième pour validation et la troisième pour la prédiction.

➤ Stacking Bayésien

Le stacking est une méthode ensembliste adaptée à la problématique des projections des taux de mortalité comme l'ont montré les auteurs de l'article précité.

Après avoir ajusté les modèles (LC, RH, APC, BCD et M6) aux données des N premières années, on peut obtenir une distribution empirique des futurs taux de mortalité $\mu_{x,t}$ pour les années N+1 à N+M basé sur des échantillons MCMC. Combiné avec les expositions des M dernières années, on peut alors obtenir une distribution empirique des décès futurs pour chaque modèle M_k parmi les modèles (LC, RH, APC, BCD, M6) :

Les données $y_{1:N}$ sont constituées des décès $d_{x,t}$ et des expositions $e_{x,t}$ pour tout âge x et pour les années allant de 1 à N ; ces données forment la base d'entraînement

Et la base de validation est constituée des décès $d_{x,t}$ et des expositions $e_{x,t}$ pour tout âge x et pour les années allant de N+1 à N+M ;

La distribution empirique pour les futurs décès pour chaque modèle peut être obtenue par :

$D_{x,t} \sim \text{Poisson}(\mu_{x,t}^k \cdot e_{x,t})$ où $\mu_{x,t}^k$ sont les taux de mortalité prévus selon les modèles M_k et $e_{x,t}$, pour $t = t_{N+1}, \dots, t_{n+M}$, sont les expositions de la base de validation.

La distribution a posteriori est obtenue via MCMN (Markov Chain Monte Carlo) :

$$p(d_{x,j} | y_{1:N}, M_k) \approx \frac{1}{S} \sum_{s=1}^S p(d_{x,j} | y_{1:N}, \theta^{(s)}, M_k) \quad (2-19)$$

Une bonne approche de moyenne devrait regrouper les modèles de telle sorte que le modèle résultant maximise la probabilité du nombre de décès observé sur la partie de la base de données réservée à la validation. C'est l'idée clé de « stacking » de distributions prédictives.

On maximise la log scoring rule sur la base de validation :

$$\max_{w \in S_1^K} \sum_{x=x_1}^{x_n} \sum_{j=t_{N+1}}^{t_{n+M}} \log \sum_{k=1}^K w_k P(d_{x,j} | y_{1:N}, M_k) \quad (2-20)$$

Où $P(d_{x,j}|y_{1:N}, M_k)$ est la densité prédictive du modèle M_k sur la base d'entraînement.

Contrairement au BMA, qui utilise des probabilités bayésiennes pour pondérer les modèles, le stacking apprend activement les poids qui maximisent les performances des prédictions combinées.

➤ Pseudo-BMA Bayésien

Le pseudo-BMA est une version simplifiée du BMA, qui utilise des scores prédictifs pour ajuster les pondérations des modèles au lieu des probabilités bayésiennes complètes. Les pondérations des modèles sont ajustées en fonction de leur performance prédictive passée, souvent mesurée par un critère comme le log-score³ ou le CRPS (Continuous Ranked Probability Score)⁴.

Le poids de chaque modèle est défini par :

$$w_k = \frac{\exp(\text{elpd}^k)}{\sum_{k=1}^K \exp(\text{elpd}^k)} \quad (2-21)$$

Où :

$$\text{elpd}^k = \sum_{x=x_1}^{x_n} \sum_{j=t_{N+1}}^{t_{N+M}} \log \left(P(d_{x_j} | y_{1:N}, M_k) \right) \quad (2-22)$$

Les poids w_k pour chaque modèle M_k sont obtenus en normalisant les scores prédictifs. Un modèle qui obtient un meilleur score prédictif sur les données passées recevra un poids plus élevé.

Le pseudo-BMA peut être utilisé pour projeter la mortalité en ajustant les pondérations des modèles en fonction de leurs performances historiques sur des périodes passées.

³ Le log-score : mesure la qualité des prédictions de probabilités.

⁴ Le CRPS : évalue la qualité d'une distribution prédictive continue.

Le pseudo-BMA est plus simple à mettre en œuvre que le BMA complet, tout en maintenant une grande partie des bénéfices de l'approche bayésienne (incorporation de l'incertitude, combinaison de modèles).

- .Choix des périodes d'ajustement, de validation et de prédiction

Nous appliquons les trois approches BMA, Stacking et Pseudo-BMA présentées dans la section précédente aux données de mortalité de la France extraites de la base « HMD » pour les hommes, les femmes et tous sexes confondus. Ces données distinguent deux périodes : il s'agit de la période avant covid 19 concerne la période 1982-2019 et de celle y compris covid en intégrant les données des années 2020 et 2021 de la période précédente , soit l'historique 1982-2021. Ce choix a été fait pour comparer les résultats des projections selon ces deux périodes afin d'illustrer l'impact de la crise sanitaire covid 19 sur les tables de mortalité prospectives.

Les données choisies concernent les décès et les expositions correspondantes des hommes, des femmes et tous sexes confondus pour les âges de 50 à 90 ans.

Pour évaluer les performances de prédiction, nous divisons les données en trois sous-ensembles :

- Un sous-ensemble d'entraînement : utilisé pour entraîner le modèle. Le modèle ajuste ses paramètres sur ces données afin de minimiser l'erreur ou maximiser une fonction d'objectif (comme la vraisemblance pour un modèle bayésien ou la précision pour un modèle classique).
- Un sous-ensemble de validation pour évaluer la performance pendant l'entraînement. Il aide à ajuster les configurations du modèle sans le biaiser. Ainsi, en s'appuyant sur l'erreur de validation, on peut choisir les paramètres qui permettent au modèle d'être le plus performant sans sur-apprentissage (overfitting). Cela aide à éviter que le modèle ne soit trop ajusté aux données d'entraînement et ainsi perdre sa capacité à bien généraliser sur de nouvelles données.
- Un sous-ensemble de test : utilisé à la toute fin du processus, après avoir choisi les paramètres et finalisé le modèle. Il permet de mesurer l'erreur finale du modèle sur des données qu'il n'a jamais vues ni lors de l'entraînement ni lors de la validation. Cet sous-ensemble donne une évaluation plus fiable des performances du modèle sur des données nouvelles, simulant ainsi la façon dont le modèle fonctionnerait en production ou sur des situations réelles.

En résumé, cette division permet de développer, ajuster et évaluer un modèle de manière équilibrée et rigoureuse, garantissant ainsi une bonne généralisation des résultats.

Tableau 2-4 : : Division de la base de données en trois sous-ensembles.

Sous-ensemble d'entraînement	Sous-ensemble validation	Sous-ensemble test
------------------------------	--------------------------	--------------------

Pour évaluer les performances de prédiction, nous divisons les données en deux parties : les données des années 1982-2009 (ou 1982-2011 dans le cas y compris covid) sont utilisées pour la sélection des pondérations et les 10 dernières années 2010-2019 (ou 2012-2021) sont utilisées pour comparer les prévisions pondérées. Pour le calcul des poids de Stacking et pseudo-BMA, les données sont ensuite divisées en deux parties : les données des années 1982-1999 (ou 1982-2001 dans le cas y compris

covid) sont utilisées comme ensemble d'entraînement tandis que les 10 années restantes sont utilisées pour la validation.

La répartition des données entre les sous-ensembles d'entraînement, de validation et de test est résumée dans les tableaux suivants.

Tableau 2-5 : Répartition des données associées à chaque méthode pour la période avant covid 19.

Avant covid 19	1982-1999	2000-2009	2010-2019
BMA	Ajustement		Prédiction
Stacking	Ajustement	Validation	Prédiction
Pseudo-BMA	Ajustement	Validation	Prédiction

Tableau 2-5 : Répartition des données associées à chaque méthode pour la période y compris covid 19.

Y compris covid 19	1982-2001	2002-2011	2012-2021
BMA	Ajustement		Prédiction
Stacking	Ajustement	Validation	Prédiction
Pseudo-BMA	Ajustement	Validation	Prédiction

Nous voulons mesurer la capacité de la méthode de sélection à choisir le modèle le plus adapté. Nous faisons cela en inspectant la valeur des poids renvoyés par chaque méthode. Nous évaluons le pouvoir prédictif de chaque modèle en examinant dans quelle mesure les taux de mortalité prévus $\mu_{x,t}$ chevauchent les taux de mortalité de l'ensemble de test.

- **Poids associés à chaque modèle**

Le tableau suivant fournit les poids obtenus via la BMA standard (vraisemblance marginale), le Stacking et pseudo-BMA pour la France, en distinguant les deux périodes avant et y compris covid 19 dans les cas des hommes, femmes et tous sexes confondus.

On constate que les approches BMA et pseudo-BMA ont tendance à sélectionner un seul modèle. Cela était attendu compte tenu de la taille de l'ensemble de données (voir Yao et al. (2018) et les références qui y figurent).

D'un côté, l'approche par Stacking sélectionne le modèle APC dans tous les cas de figure que ce soient pour les hommes, femmes, tous sexes confondus avant et y compris Covid 19 ; néanmoins avec un poids plus important pour les femmes.

D'un autre côté, le modèle RH est choisi avec des poids importants pour les hommes et tous sexes confondus alors que le modèle LC est privilégié pour les femmes, dans les deux cas avant et y compris Covid 19. En revanche, on note que les modèles CBD et M6 n'ont pas été sélectionnés dans le cadre de stacking dans quasiment tous les cas de figure.

Globalement, on observe le BMA et le pseudo BMA sélectionnent presque toujours le même modèle. Nous notons également que l'approche standard BMA privilégie le modèle RH; ceci est en accord avec

la littérature fréquentiste dans lequel le modèle RH avec effet de cohorte est souvent été identifié comme le meilleur modèle candidat lorsque la sélection du modèle est basée sur le critère BIC ou AIC, voir Renshaw et Haberman (2011).

Tableau 2-5 : Poids des modèles selon les différents cas de figure.

Avant covid : période 1982-2019 et y c covid : période 1982-2021				
Poids des modèles	Sexe	BMA	Stacking	Pseudo-BMA
LC yc covid	H	0	0	0
	F	0	0,372	0
	Total	0	0	0
avant covid	H	0	0,019	0
	F	0	0,879	1
	Total	0	0	0
RH yc covid	H	1	0,701	1
	F	1	0,086	0
	Total	1	0,777	1
avant covid	H	0	0,822	1
	F	1	0,001	0
	Total	1	0,635	0,999
APC yc covid	H	0	0,299	0
	F	0	0,541	1
	Total	0	0,223	0
avant covid	H	0	0,155	0
	F	0	0,12	0
	Total	0	0,365	0,001
CBD yc covid	H	0	0	0
	F	0	0	0
	Total	0	0	0
avant covid	H	0	0.004	0
	F	0	0	0
	Total	0	0	0
M6 yc covid	H	0	0	0
	F	0	0,001	0
	Total	0	0	0
avant covid	H	0	0	0
	F	0	0	0
	Total	0	0	0

2.5.3. Performance prédictive des modèles

Dans cette section, nous examinerons la précision des prévisions à travers l'analyse de la mesure « erreur absolue moyenne (MAE) » mais aussi de deux coefficients Log score et CRPS.

- **Back-testing sur les observations : Erreur absolue moyenne (MAE)**

L'erreur absolue moyenne (MAE) est une moyenne de l'écart entre les prévisions et les valeurs réelles. Le MAE est calculé en additionnant les différences absolues entre les valeurs prédites et les valeurs réelles et en divisant par le nombre d'observations.

Nous prenons les données de la population française de mortalité issues de HMD pour les années civiles de 1982 à 2019 (ou 2021 dans le cas y compris covid) et les personnes âgées de 50 à 90 ans, pour les hommes, femmes et tous sexes confondus. Nous distinguons deux cas de figure :

Avant Covid, l'erreur quadratique moyenne est définie pour $x=50, \dots, 90$ ans, par :

$$MAE_x = \frac{1}{10} \sum_{t=2010}^{2019} |D_{x,t} - E_{x,t} \hat{\mu}_{x,t}| \quad (2-23)$$

Dans le cas, y compris covid :

$$MAE_x = \frac{1}{10} \sum_{t=2012}^{2021} |D_{x,t} - E_{x,t} \hat{\mu}_{x,t}| \quad (2-24)$$

Où : $\mu_{x,t}$ est le taux de mortalité a posteriori prévu,

$D_{x,t}$: décès d'âge x l'année t,

$E_{x,t}$: expositions d'âge x l'année t.

Les graphiques suivants montrent le MAE par âge pour la France, hommes, femmes, tous sexes confondus avant et y compris Covid selon les trois méthodes : BMA, stacking et pseudo BMA.

Les niveaux de MAE pour l'approche Stacking se situent clairement en dessus des niveaux de MAE de BMA et de pseudo BMA pour les femmes aux âges compris entre 50 et 75 ans ; puis ils repassent en dessous entre 76 et 90 ans . Pour les hommes et tous sexes confondus, les performances des trois méthodes sont proches pour les 50 à 80 ans mais le stacking a de meilleurs MAE après 80 ans.

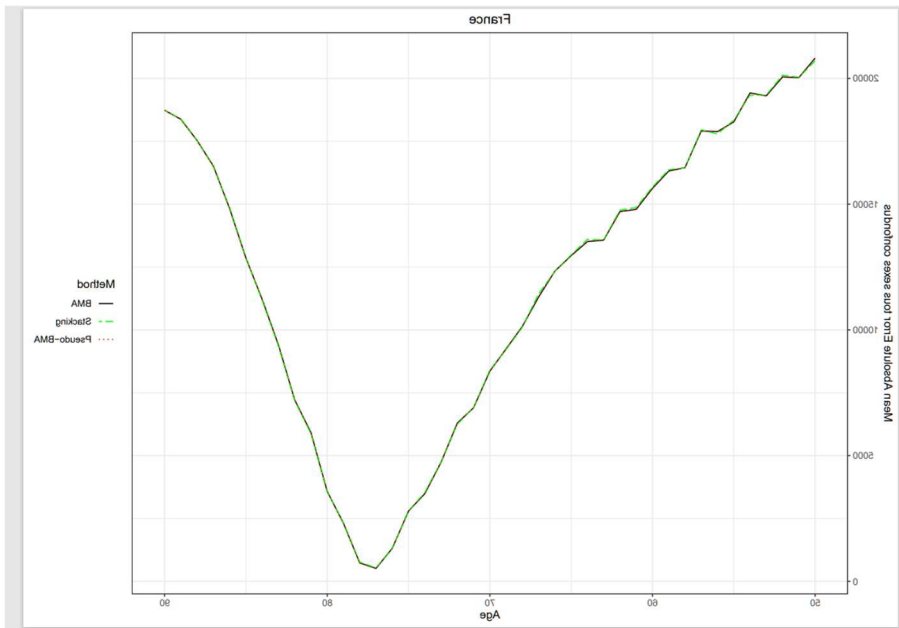


Figure 2-18 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(a) Tous sexes confondus , y compris covid

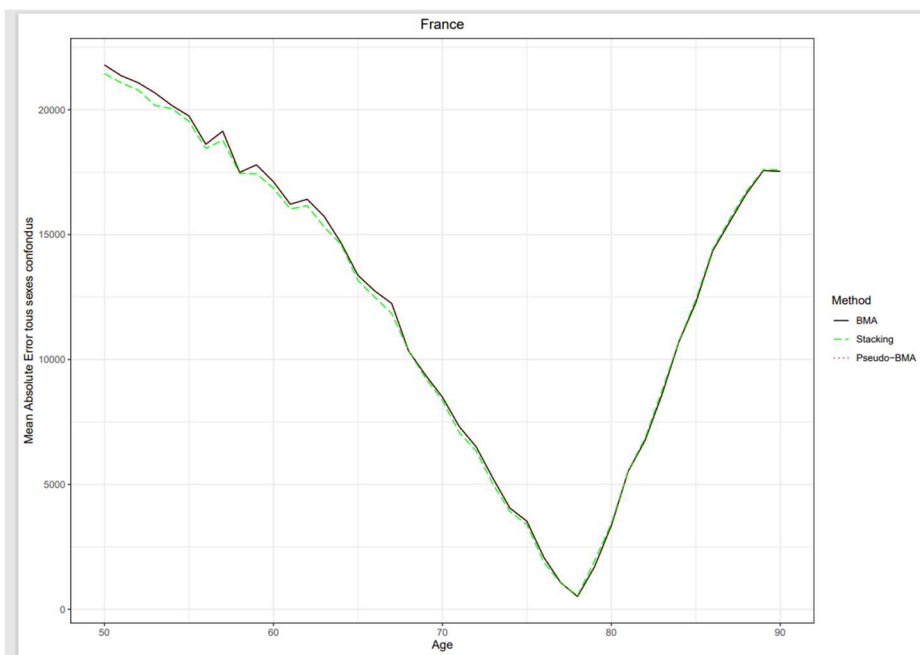


Figure 2-19 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(b) Tous sexes confondus, avant covid 19

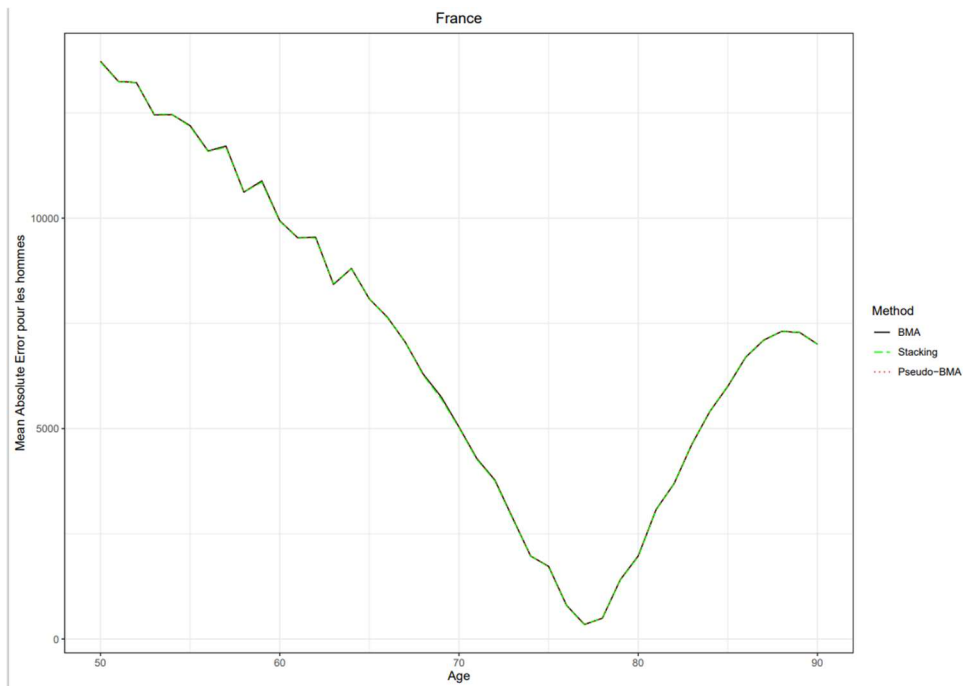


Figure 2-20 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(c) Hommes, y compris covid

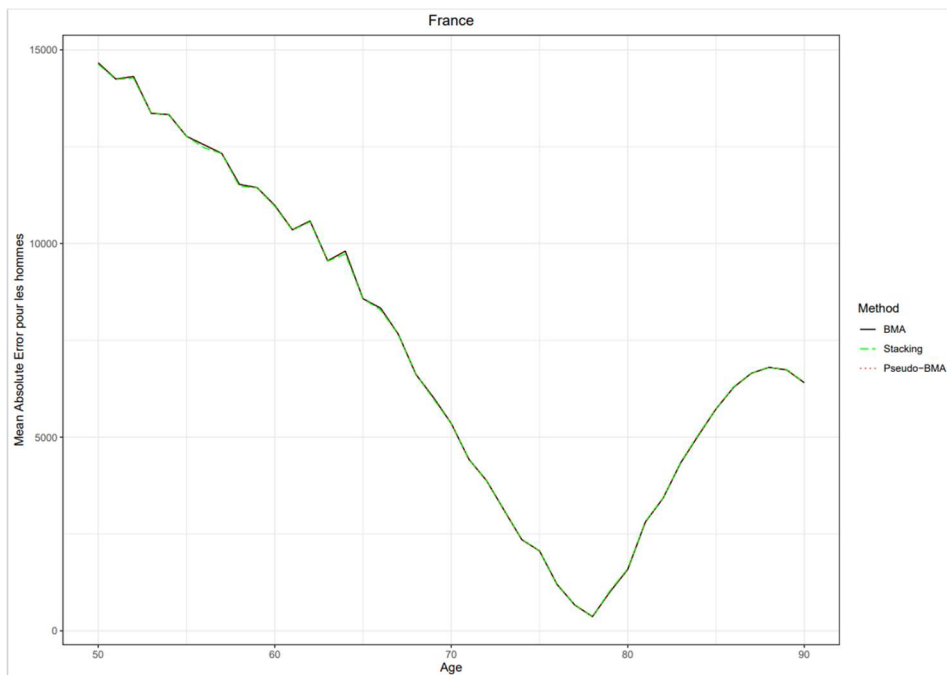


Figure 2-21 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(d) Hommes, avant covid

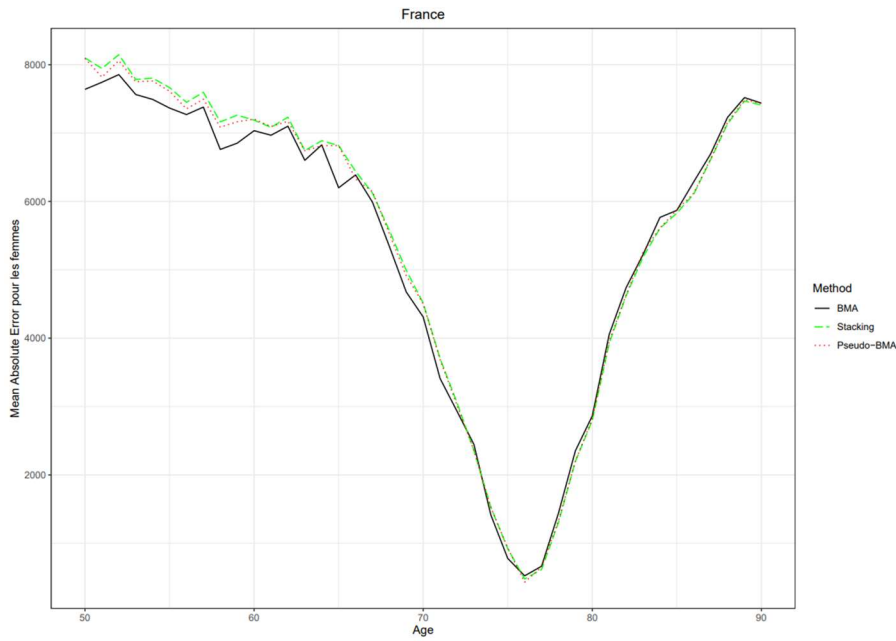


Figure 2-22 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(e) Femmes, y compris covid

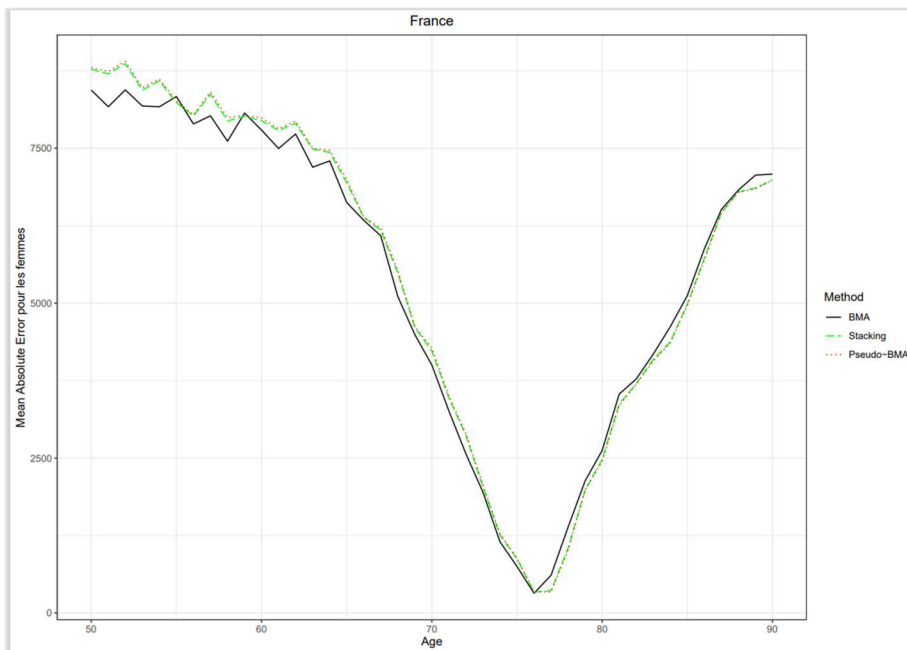


Figure 2-23 : MAE par âge selon les trois méthodes : BMA, Stacking et Pseudo-BMA.

(f) Femmes, avant covid

- **Log score et CRPS pour la France**

Le LogScore, aussi appelé log-vraisemblance, évalue la performance d'un modèle probabiliste en mesurant la qualité des probabilités attribuées aux événements observés.

Le CRPS (The continuous ranked probability score) est une mesure utilisée pour évaluer les modèles qui produisent des distributions de probabilité prédites. Il compare la distribution prédite à la valeur observée de manière continue.

Comme nous utilisons l'inférence bayésienne, nous avons une distribution de probabilité $F_{x,t}$ autour de chaque taux de mortalité $\mu_{x,t}$. La précision de cette famille de distributions de prévision $F = (F_{x,t})$ est mesurée via deux règles de notation.

Le score logarithmique est défini, pour le cas avant covid, comme :

$$\text{LogS}(F, y) = -\frac{1}{40} \sum_{x=50}^{90} \frac{1}{10} \sum_{t=2010}^{2019} \log \left[f_{x,t} \left(\frac{d_{x,t}}{e_{x,t}} \right) \right], \quad (2-25)$$

Où $f_{x,t}$ est la fonction de densité probabiliste de $F_{x,t}$.

Le Log Score est plus faible (donc meilleur) lorsque le modèle attribue une probabilité élevée à l'événement qui se produit effectivement.

Le CRPS est défini par :

$$\text{CRPS}(F, y) = \frac{1}{40} \sum_{x=50}^{90} \frac{1}{10} \sum_{t=2010}^{2019} \int_{\mathbb{R}} (F_{x,t}(\mu) - \mathbf{1}\{d_{x,t}/e_{x,t} \leq \mu\})^2 d\mu. \quad (2-26)$$

Un CRPS de 0 indique une prédiction parfaite.

L'utilisation combinée de ces deux métriques peut fournir une évaluation complète de la performance des modèles prédictifs de mortalité, en capturant à la fois la qualité globale de la distribution des prédictions et la précision des probabilités spécifiques.

Dans l'ensemble, cet exercice de validation montre que, sur la période avant covid 19, le Stacking a tendance à surpasser le pseudo-BMA et BMA standard en termes de capacité de prévision à 10 ans pour les hommes, femmes et tous sexes confondus. On remarque que quand intègre les deux années de covid 19, 2020 et 2021, la situation n'est pas évidente : le stacking demeure meilleur pour tous sexes confondus mais les règles de notation donnent les meilleures performances au standard BMA pour les hommes et pseudo-BMA pour les femmes.

En résumé, cette section montre qu'un modèle qui a fourni de bons prévisions pour les 10 dernières années a de bonne chance de bien performer pour les 10 années suivantes.

D'un autre côté, un modèle qui s'ajuste bien aux données de mortalité n'a aucune raison a priori d'être bon dans la prévision de mortalité futures. Nous recommandons donc un stacking basé sur le principe de validation Leave-Futur-Out plutôt que l'ajustement (standard BMA) à des fins de prévision.

Tableau 2-6 : Log score et CRPS selon les trois méthodes : BMA, Stacking et pseudo BMA.

Model Weights	Sexe	Log score	CRPS
BMA yc covid	H	-4,1011	0,0022
	F	-2,9499	0,0021
	Total	-3,2453	0,0021
avant covid jusqu'à 2019	H	-4,7809	0,0010
	F	-4,4431	0,0012
	Total	-4,6623	0,0011
Stacking yc covid	H	-3,9346	0,0023
	F	-4,7474	0,0016
	Total	-3,6533	0,0021
avant covid jusqu'à 2019	H	-4,9556	0,0009
	F	-5,6016	0,0006
	Total	-5,1877	0,0011
Pseudo- BMA yc covid	H	-4,1011	0,0022
	F	-5,5083	0,0017
	Total	-3,2453	0,0021
avant covid jusqu'à 2019	H	-4,7809	0,0010
	F	-5,5304	0,0006
	Total	-4,6628	0,0011

Pour les modèles individuels, Le modèle APC est meilleur dans tous les cas de figure : hommes, femmes, tous sexes confondus, avant et y compris covid.

Tableau 2-7 : Log score et CRPS selon les modèles individuels.

Model Weights	Sexe	Log score	CRPS
LC : yc covid	H	-4,4916	0,0013
	F	-4,7253	0,0012
	Total	-4,8967	0,0012
avant covid jusqu'à 2019	H	-4,8974	0,0010
	F	-5,5304	0,0006
	Total	-5,4923	0,0006
RH yc covid	H	-4,1011	0,0022
	F	-2,9499	0,0021
	Total	-3,2453	0,0021

avant covid jusqu'à 2019	H	-4,7809	0,0010
	F	-4,4431	0,0012
	Total	-4,6623	0,0011
APC yc covid	H	-4,7988	0,0025
	F	-5,5083	0,0017
	Total	-5,2435	0,0019
avant covid jusqu'à 2019	H	-5,4819	0,0014
	F	-5,8462	0,0010
	Total	-5,7547	0,0011
CBD yc covid	H	-4,2552	0,0046
	F	-4,5446	0,0039
	Total	-4,4720	0,0039
avant covid jusqu'à 2019	H	-4,2662	0,0040
	F	-4,5846	0,0036
	Total	-4,5242	0,0035
M6 yc covid	H	-4,5159	0,0020
	F	-4,2419	0,0024
	Total	-4,4720	0,0035
avant covid jusqu'à 2019	H	-3,9737	0,0029
	F	-3,5709	0,0033
	Total	-4,0702	0,0027

2.5.4. Modèles choisis en fonction de leurs performances

Compte tenu des analyses précédentes concernant les deux coefficients log score et CRPS, les modèles présentant les meilleures performances parmi les modèles ensemblistes et qui seront retenus pour réaliser les projections à long terme sont :

1. Avant covid, Stacking sélectionné pour hommes, femmes et tous sexes confondus selon les combinaisons suivantes :
 - Pour les hommes : $0.019*LC + 0.822*RH + 0.155*APC$
 - Pour les femmes : $0.879*LC + 0.001 *RH + 0.120*APC$
 - Pour tous sexes confondus : $0.635*RH + 0.365 *APC$
2. Y compris covid, les modèles les plus performants retenus sont :

- BMA pour hommes, soit le modèle RH
- pseudo-BMA pour femmes, soit le modèle APC
- et Stacking pour tous sexes confondus : $0.777 \cdot RH + 0.223 \cdot APC$

2.6. Résultats des projections pour les tables de référence « HMD »

2.6.1. Projections de la composante temporelle des modèles retenus : LC, RH et APC

L'intérêt de mettre en place un modèle capable de capturer l'évolution des taux de mortalité historiques est par la suite de pouvoir prédire du mieux possible les taux de mortalité futurs. Cela consiste à modéliser les séries temporelles de k_t grâce à un modèle ARIMA et d'utiliser les futurs \hat{k}_t estimés avec les coefficients α_x et β_x pour chaque âge.

▪ Une modélisation des κ_t via le modèle ARIMA(p,d,q)

Généralement, l'extrapolation des taux de mortalité futurs est réalisée grâce à la modélisation de la série des k_t sous forme ARIMA (AutoRegressive-Integrated-Moving Average).

Les modèles ARIMA sont composés de :

- Processus autorégressifs qui supposent que chaque valeur peut être prédite par la somme pondérée d'un ensemble de valeurs antérieures,
- Processus d'intégration qui supposent que chaque valeur présente une différence constante avec la valeur antérieure,
- Processus de moyenne mobile qui supposent que chaque valeur est fonction des erreurs entachant les points précédents,
- Un terme aléatoire d'erreur.

Le paramètre p correspond au nombre de termes autorégressifs, le paramètre d au nombre de différence et le paramètre q au nombre de moyennes mobiles.

Les modèles ARIMA peuvent être définies comme suit : k_t suit un processus ARIMA(p,d,q) si :

$$\left(1 - \sum_{i=1}^p a_i L^i\right) (1 - L^d) \kappa_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t \quad (2-27)$$

Où :

L est l'opérateur retard,

les ε_t sont des termes d'erreurs indépendantes et identiquement distribués (i.i.d.) qui suivent une loi normale centrée,

p, d et q désignent respectivement les paramètres autorégressifs, d'intégration et moyenne mobile,

Pour la modélisation des k_t , il convient de suivre les étapes de la méthode de Box et Jenkins comme suit :

- Identifier le modèle,
- Estimer les paramètres,
- Valider le modèle,
- Réaliser les prévisions

Ici, la fonction "auto.arima" du package *forecast* est utilisée pour effectuer une modélisation automatique de séries chronologiques à l'aide de la méthode ARIMA (AutoRegressive Integrated Moving Average).

En effet, la fonction "auto.arima" utilise des techniques de sélection de modèle automatiques pour déterminer le meilleur modèle ARIMA pour une série chronologique donnée. Elle explore différents ordres de modèles ARIMA, tels que les ordres d'autorégression (p), d'intégration (d) et de moyenne mobile (q), et utilise des critères d'information tels que le critère d'information d'Akaike (AIC) ou le critère d'information bayésien (BIC) pour évaluer la qualité de chaque modèle. Cela permet de trouver rapidement et efficacement un modèle ARIMA approprié, sans avoir besoin d'effectuer manuellement des tests et des évaluations pour différents ordres de modèles.

On va détailler ici uniquement la méthode de modélisation dans le cas d'un modèle Lee-Carter pour les hommes et les femmes y compris Covid. Je présenterai en annexe les résultats obtenus dans les différents cas de figure des RH et APC. Aussi, les résultats du modèle LC avant covid sera présenté en annexe pour des soucis de concision.

Dans le cadre du modèle Lee-Carter y compris covid, le meilleur modèle sélectionné à l'aide de la fonction auto.arima est de type ARIMA(1,1,0) pour les hommes, femmes et tous sexes confondus.

Tableau 2-8 : Ajustement d'un modèle ARIMA à la série temporelle des indices de mortalité K_t du modèle LC y compris Covid

Kt y compris covid ARIMA(1,1,0)	Hommes		Femmes		Tous sexes confondus	
	ar1	drift	ar1	drift	ar1	drift
Coefficients	-0,255	-0,6791	-0.3630	-0.6077	-0,3125	-0,6392
s.e	0,1549	0,1166	0.1487	0.1164	0,152	0,1197
sigma^2	0,8707		1,021		1,002	
log likelihood	-51,65		-54,79		-54,4	
AIC	109,29		115,58		114,8	
AICc	109,98		116,27		115,49	
BIC	114,28		120,57		119,79	

Que ce soient pour la population masculine, la population féminine ou bien tous sexes confondus pour l'historique y compris Covid19, le modèle ARIMA (1,1,0) a été sélectionné. Les nombres entre parenthèses représentent les ordres (p, d, q) de la partie ARIMA. Cela signifie qu'il y a un terme d'autorégression d'ordre 1 (AR(1)), un terme de différenciation (d = 1) et aucun terme de moyenne mobile (MA = 0).

Le paramètre d est le nombre de différenciations effectuées pour stationnariser la série.

- "Coefficients": Cette section affiche les estimations des coefficients du modèle ARIMA. Ici, on a deux coefficients pour chaque modèle.
- s.e.": Cette section affiche les erreurs standard des coefficients estimés.
- "sigma^2": C'est l'estimation de la variance du bruit (ou de l'erreur) dans le modèle ARIMA. Dans cet exemple, sigma^2 est estimé à 0.8707 dans le cas des hommes et à 1.021 dans le cas des femmes..
- "log likelihood": Il s'agit de la valeur de la log-vraisemblance maximisée du modèle ARIMA. Dans cet exemple, la log-vraisemblance maximisée est -51.65 pour les hommes et -54.79 pour les femmes.
- "AIC", "AICc" et "BIC": Ce sont des critères d'information utilisés pour comparer différents modèles. AIC (Akaike Information Criterion), AICc (AIC corrigé) et BIC (Bayesian Information Criterion) sont souvent utilisés pour sélectionner le meilleur modèle. Un modèle avec des valeurs de critères plus faibles est généralement préféré car il indique un meilleur ajustement aux données avec une complexité minimale.

Dans un ARIMA(1,1,0), l'équation générale est :

$$\Delta k_t = \phi_1 \cdot \Delta k_{t-1} + \mu + \epsilon_t \quad (2-28)$$

Où :

$$\Delta k_t = k_t - k_{t-1} \quad (2-29)$$

est la différence entre la valeur de k_t et $k_{(t-1)}$ à deux instants consécutifs.

ϕ_1 est le paramètre autorégressif d'ordre 1.

μ est la dérive ("drift") qui correspond à la tendance linéaire de la série.

Et ϵ_t est un terme d'erreur ou bruit blanc.

L'équation représentant le modèle ARIMA trouvé pour les hommes dans le cas y compris covid à partir des résultats donnés ($\phi_1 = -0.2550$ et $\mu = -0.6791$) est :

$$\Delta k_t = -0.2550 \cdot \Delta k_{t-1} - 0.6791 + \epsilon_t \quad (2-30)$$

Ou en termes de k_t et k_{t-1} :

$$k_t - k_{t-1} = -0.2550 \cdot (k_{t-1} - k_{t-2}) - 0.6791 + \epsilon_t \quad (2-291)$$

Ce qui se simplifie en :

$$k_t = 0.7450 \cdot k_{t-1} + 0.2550 \cdot k_{t-2} - 0.6791 + \epsilon_t \quad (2-32)$$

- Interprétation :
 - k_t dépend de la valeur de k_{t-1} , avec un coefficient de 0.7450, et de k_{t-2} , avec un coefficient de 0.2550.
 - Le drift ($\mu=-0.6791$) représente une tendance linéaire négative dans la série, c'est-à-dire que, indépendamment des valeurs passées, il y a une tendance à la baisse de k_t
 - ϵ_t est un terme d'erreur ou bruit blanc qui introduit une variabilité imprévisible à chaque étape.

Avec un ARIMA(1,1,0) ayant $ar_1 = -0.2550$ et $drift=-0.6791$, l'évolution de la composante temporelle k_t des hommes est principalement influencée par la tendance linéaire négative (drift), ainsi que par une combinaison pondérée des deux valeurs précédentes k_{t-1} , et k_{t-2} .

De la même manière, l'évolution de la composante temporelle pour les femmes avec un ARIMA(1,1,0) ayant $ar_1 = -0.3630$ et $drift=-0.6077$ s'écrit :

$$k_t = 0.673 \cdot k_{t-1} + 0.3630 \cdot k_{t-2} - 0.6077 + \epsilon_t \quad (2-33)$$

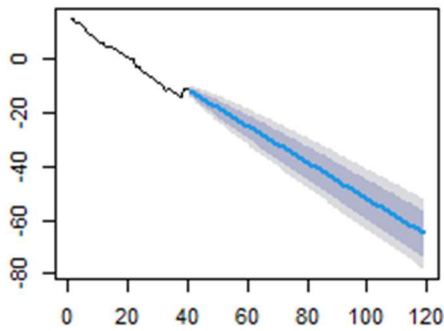
De la même façon, pour tous sexes confondus, l'évolution de la composante temporelle s'écrit :

$$k_t = 0.6875 \cdot k_{t-1} + 0.3125 \cdot k_{t-2} - 0.6392 + \epsilon_t \quad (2-304)$$

- Projection des taux de mortalité dans le futur

Ayant à présent déterminé les modèles ARIMA suivis par la composante k_t pour les deux populations, la prochaine étape est de faire la projection de cette composante afin de calculer les taux de mortalité dans le futur. Nous effectuons la projection de la composante jusqu'à 2100, Ci-après nous présentons l'allure de la composante. Comme évoqué précédemment, les graphiques des prévisions du paramètre des modèles RH et APC se trouvent en annexe.

Forecasts from ARIMA(1,1,0) with drift



Forecasts from ARIMA(1,1,0) with drift

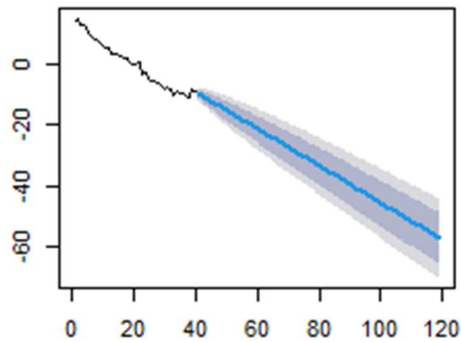
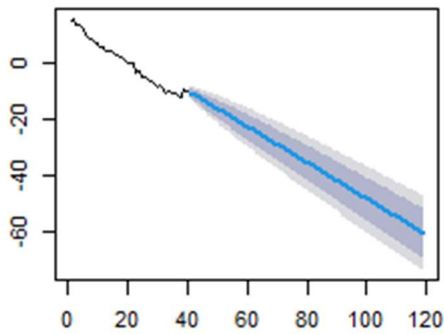


Figure 2-24 : Evolution de la composante temporelle du modèle Lee-Carter.

(a) Hommes- y compris covid

(b) Femmes – y compris covid

Forecasts from ARIMA(1,1,0) with drift



Forecasts from ARIMA(1,1,0) with drift

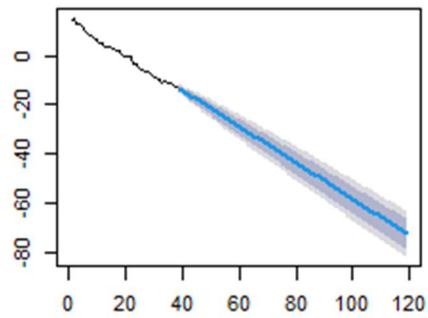
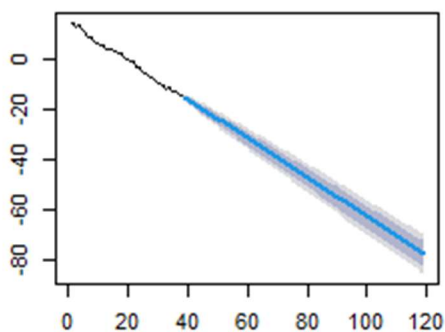


Figure 2-25 : Evolution de la composante temporelle du modèle Lee-Carter.

(c) Tous sexes confondus – y compris covid

(d) Tous sexes confondus – avant covid

Forecasts from ARIMA(1,1,0) with drift



Forecasts from ARIMA(1,1,0) with drift

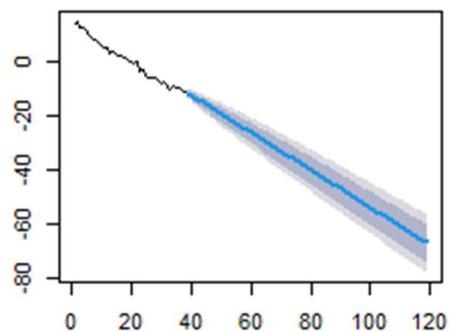


Figure 2-26 : Evolution de la composante temporelle du modèle Lee-Carter.

(e) Hommes- avant covid

(f) Femmes – avant covid

On constate que l'intervalle de confiance est plus large dans le cas des projections y compris covid et notamment pour les femmes.

Il suffit maintenant d'appliquer les coefficients estimés par le modèle de Lee & Carter en remplaçant les k_t avec les valeurs projetées pour avoir les taux de mortalité futurs de la population française.

2.6.2. Espérance de Vie médian 50-90

Comme expliqué dans l'article de référence déjà cité, étant donné que l'espérance de vie à la naissance nécessiterait une tranche d'âge complète, nous nous concentrons sur la tranche d'âge de 50 à 90 ans. Pour cette raison, nous calculons plutôt une probabilité de survie d'une personne d'âge $x = 50$ pour toute année t d'une période de 40 ans comme suit :

$${}_{40}P_{50,t} = \prod_{i=0}^{39} p_{50+i,t} = \prod_{i=0}^{39} \exp(-\mu_{50+i,t}). \quad (2-35)$$

Cela correspond à la probabilité qu'une personne de 50 ans vive plus de 40 années supplémentaires compte tenu des conditions de mortalité de l'année t .

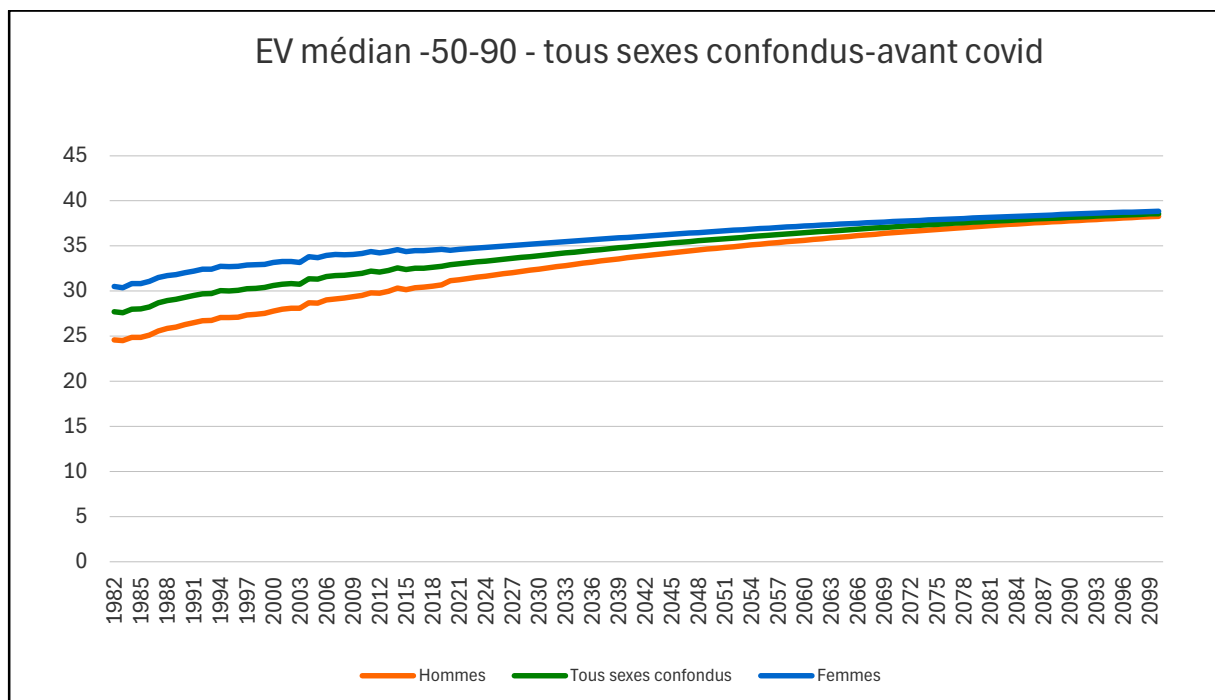


Figure 2-27: Evolution de l'espérance de vie à 50 ans (tronquée à 90 ans) avant covid par sexe.

On note une convergence entre les espérances de vie à 50 ans tronquées à 90 ans entre les hommes et les femmes. Cette convergence est davantage marquée en ne tenant pas compte des années covid

19. Aussi, on note une baisse marquée de cet indicateur en 2020 et 2021 notamment pour les hommes. Cela aurait un impact sur le rapprochement entre les espérances de vie à long terme entre la population masculine et la population féminine.

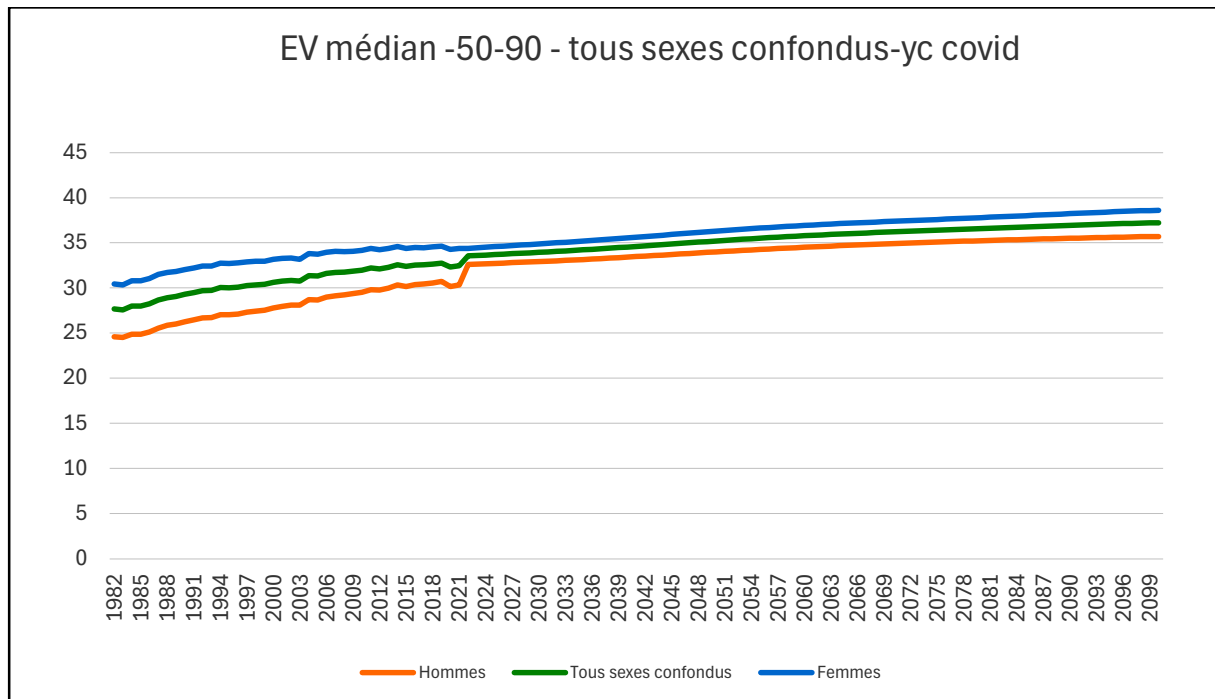


Figure 2-28 : Evolution de l'espérance de vie à 50 ans (tronquée à 90 ans) y compris covid par sexe.

2.6.3. Evolution des quotients de mortalité

- **Projections « avant covid » à partir d'un historique de 1982 à 2019**

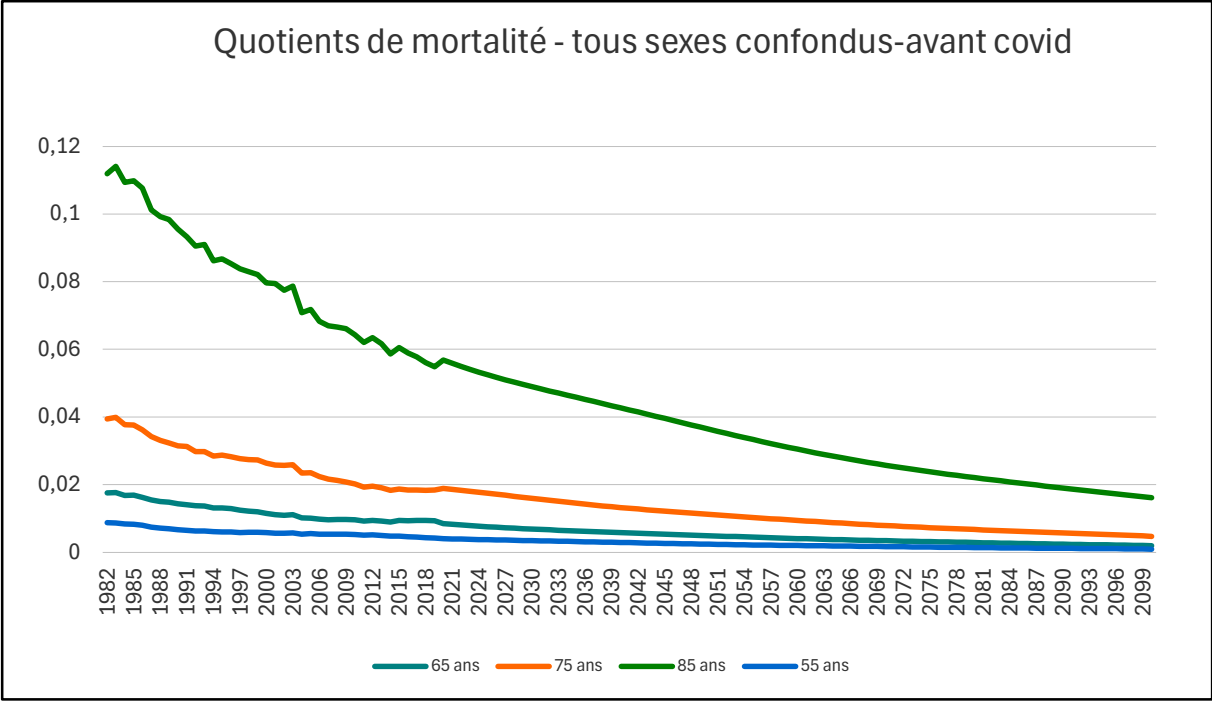


Figure 2-29 : Projection des quotients de mortalité avant covid 19 – tous sexes confondus.

Avant covid 19 et en se basant sur l'historique 1982-2019 pour réaliser les projections des quotients de mortalité, on constate une légère baisse des quotients de mortalité en 2020 des personnes âgées tous sexes confondus ou bien masculins et ayant 85 ans liée à l'augmentation des espérances de vie des dernières années. Cette baisse est moins prononcée chez les femmes.

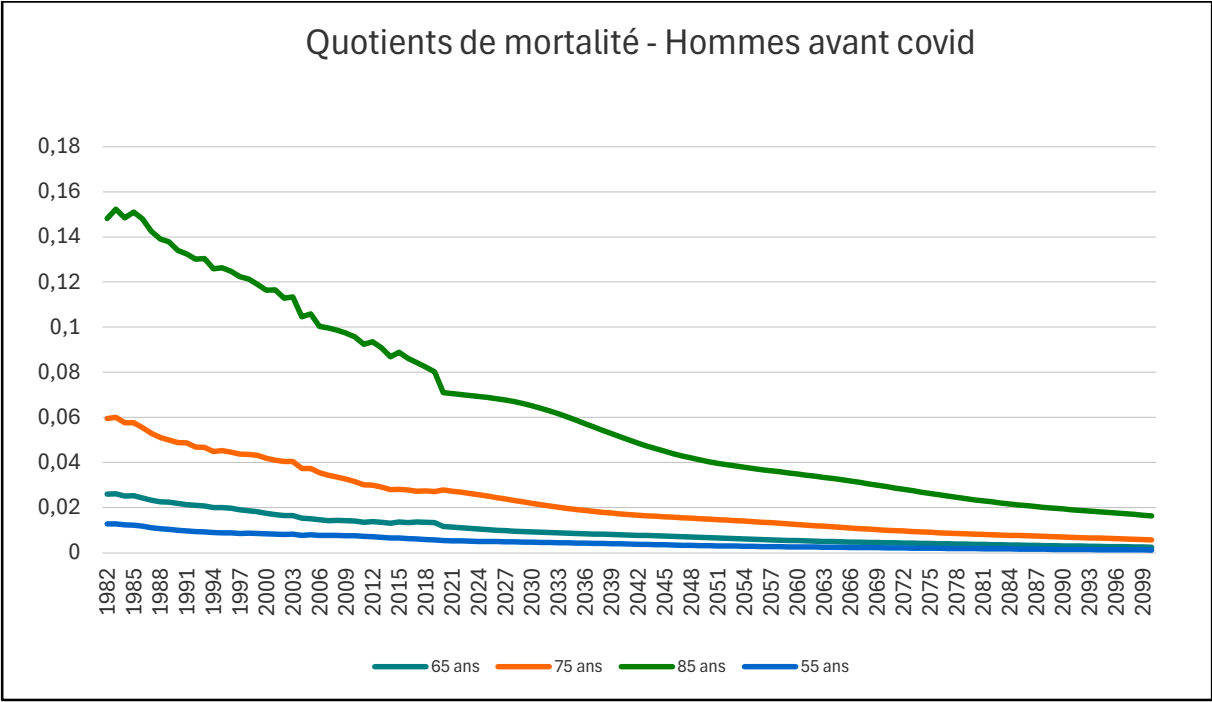


Figure 2-30 : Projection des quotients de mortalité avant covid 19 –Hommes.

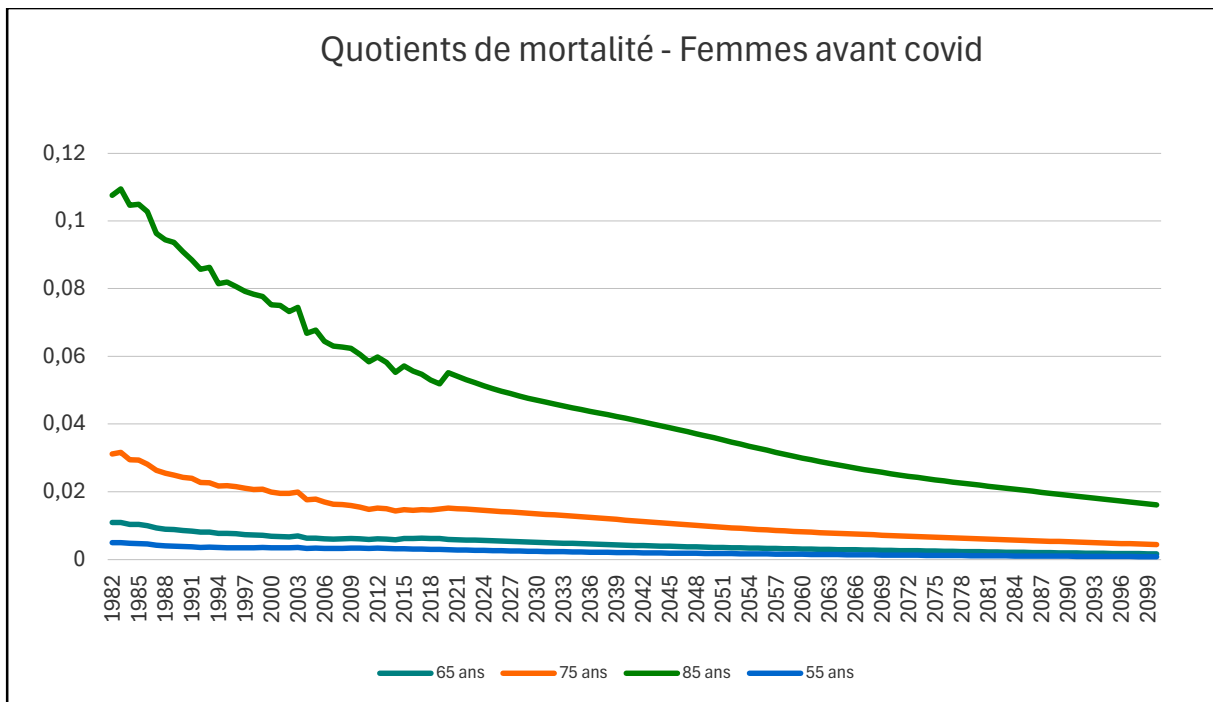


Figure 2-31 : Projection des quotients de mortalité avant covid 19-Femmes.

- Projections « y compris covid » à partir d'un historique de 1982 à 2021

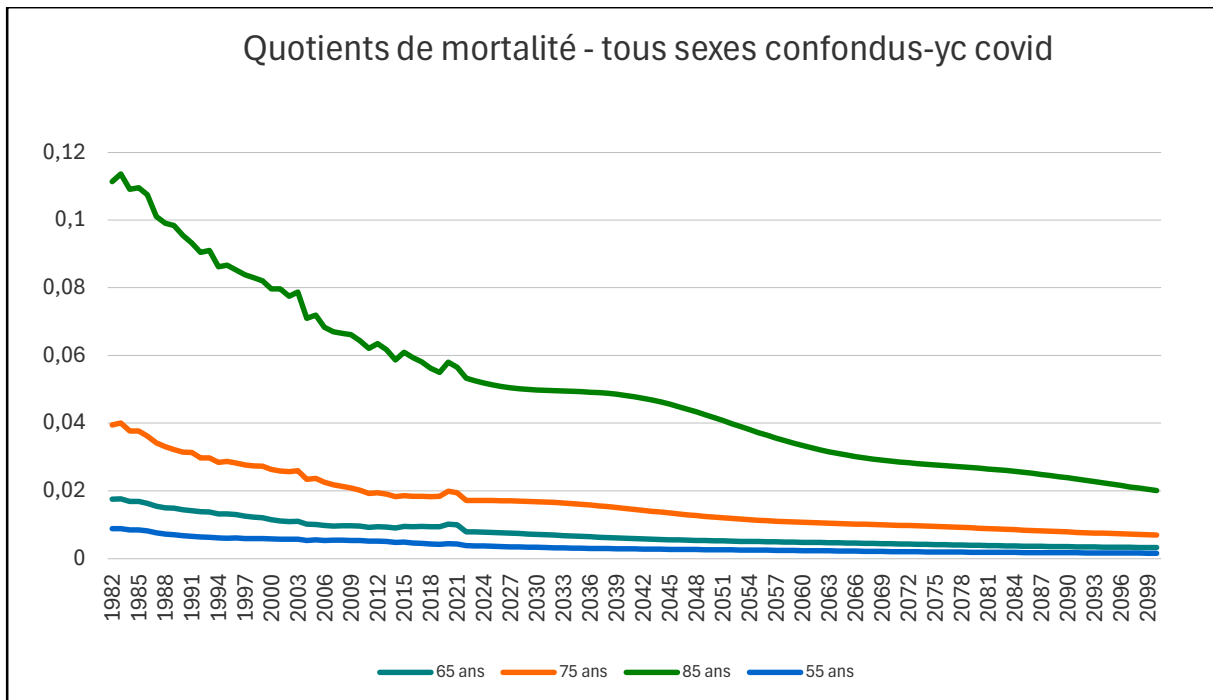


Figure 2-32 : Projection des quotients de mortalité y compris covid 19 –tous sexes confondus.

Après avoir augmenté en 2020 et 2021, les taux de mortalité projetés des personnes très âgées ayant 85 ans notamment du sexe masculin baissent drastiquement en 2023 parce qu'il y a eu une surmortalité de la tranche d'âge 80-84 ans pendant la crise de covid19. Cette tendance est moins prononcée chez les femmes. De manière générale, pour les tranches d'âge plus jeunes ayant 55 ans, 65 ou 75 ans, cette tendance à la baisse est beaucoup plus faible.

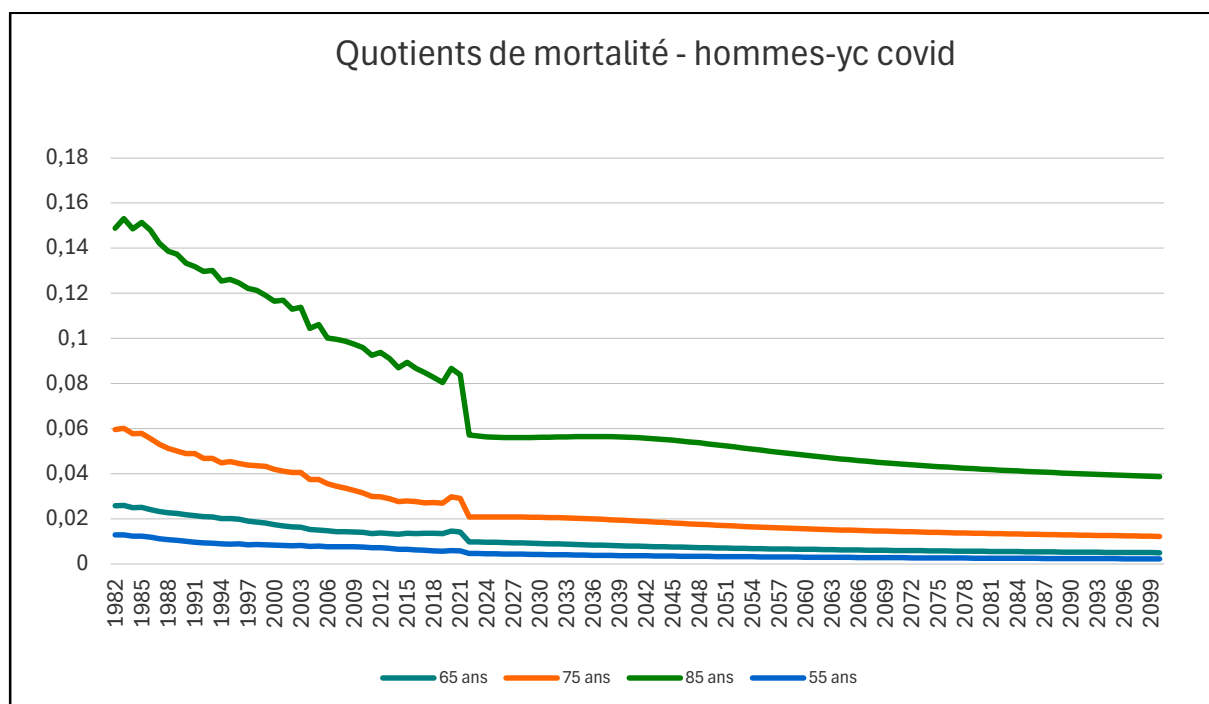


Figure 2-33 : Projection des quotients de mortalité y compris covid 19 – Hommes.

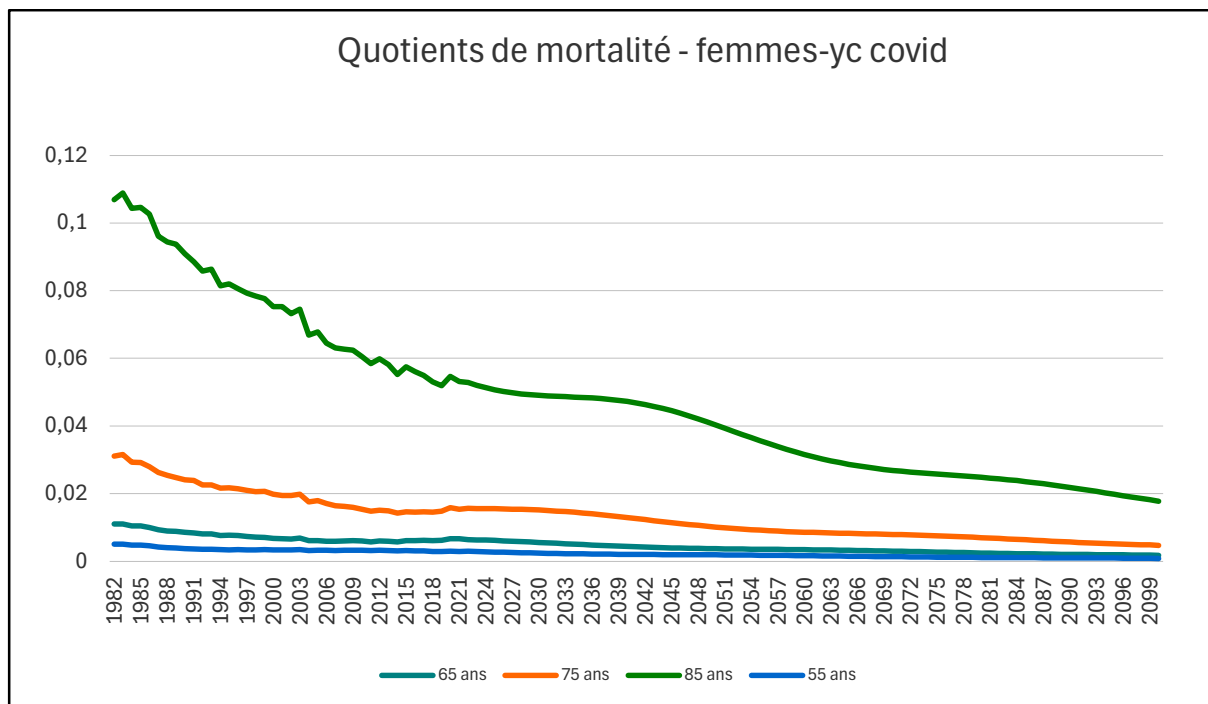


Figure 2-34 : Projection des quotients de mortalité y compris covid 19 – Femmes.

- **Ecart entre les réalisations et les projections en 2020 et 2021**

Les écarts entre les prévisions fondées sur l'historique 1982-2019 et les réalisations en 2020 et 2021 s'expliquent par la crise sanitaire covid19 car ils concernent particulièrement les personnes de la tranche d'âge 70-90 ans.

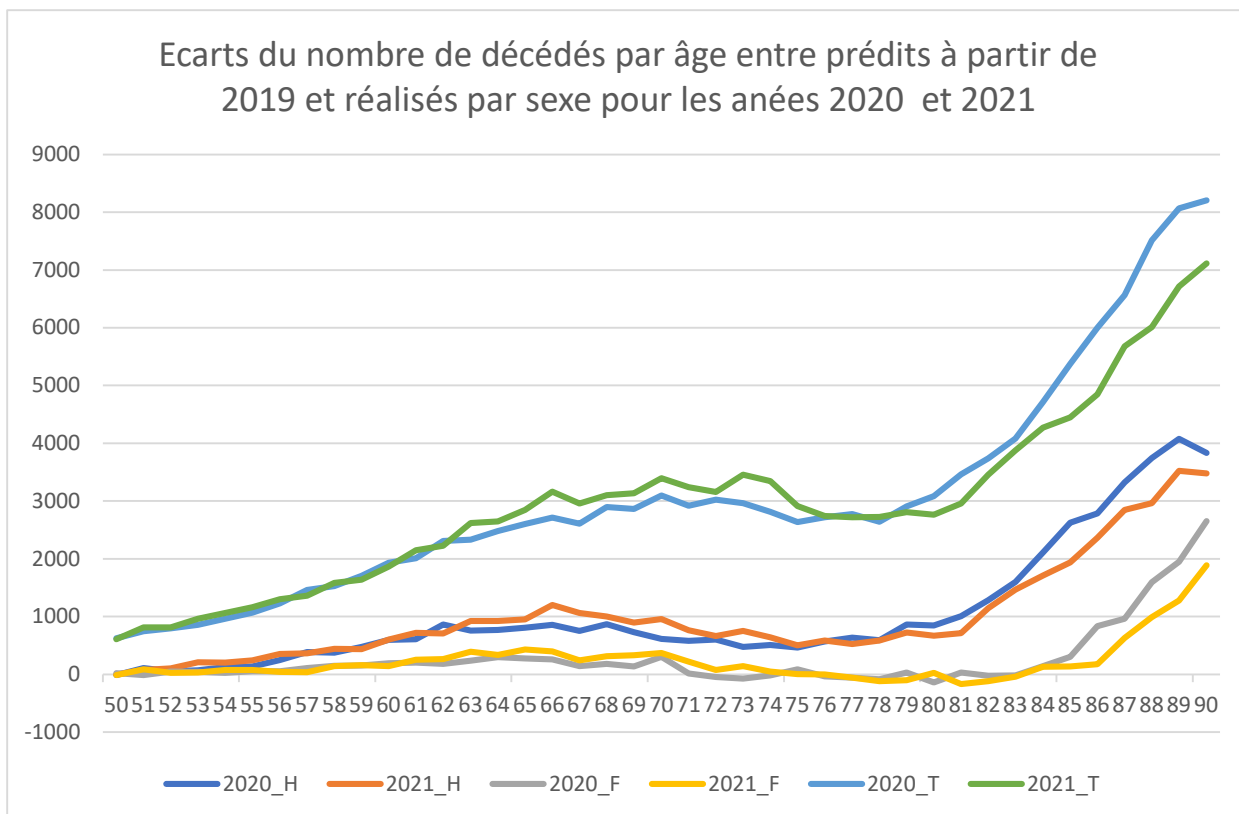


Figure 2-35 : Ecart entre réalisations et prévisions du nombre de décès par âge et selon le sexe des années 2020 et 2021.

Chapitre 3 : Construction d'une table de mortalité d'expérience à partir d'un jeu de données confidentielles d'un organisme de protection sociale

3.1. Présentation et analyse du jeu de données

La base de données est constituée d'un jeu de données d'un organisme de protection sociale. Ces données sont confidentielles. Pour construire la table de mortalité d'expérience, nous disposons dans ce jeu, pour chaque individu, des variables suivantes : date de naissance, sexe, date de début d'observation, statut au début d'observation (cotisant, retraité ou bénéficiaire d'une réversion), date de fin d'observation et motif de sortie (décès, départ de l'organisme de protection sociale au cours de la période d'observation, présent à la date d'extraction) pour les années : 2006, 2007 et 2008. Par conséquent, la période d'observation est de trois ans.

Le volume du portefeuille pris en compte dans l'analyse est conséquent en terme d'effectif pour la période d'étude (1 936 206 ayant entre 50 et 90 ans dont 55,5% des femmes). Donc, on dispose a priori de données suffisantes pour porter un avis sur la mortalité des adhérents à cet organisme de protection sociale.

La pertinence d'une table d'expérience est avant tout conditionnée par la qualité des données utilisées pour sa construction. Il convient, donc, au préalable réaliser une analyse descriptive de la base des données de cet organisme.

Au cours de la période d'observation, 7,3% des personnes sont décédées ; 77,6% sont présents à la date de fin d'observation et 15% ont quitté l'organisme.

Tableau 3-1 : Répartition du motif de sortie des assurés à la date de fin d'observation.

A la date de fin d'observation	Décédé	Présent	Absent (hors décès)
Effectif	177 847	1 880 789	353 735
Proportion	7,3%	77,6%	15,0%

Un peu plus de la moitié des adhérents sont des cotisants, 37% des retraités et 11% titulaires d'une pension de réversion.

Tableau 3-2 : Répartition du statut des assurés à la date de début d'observation.

A la date de début d'observation	Cotisant	Retraité	Réservataire
Effectif	1 261 013	896 528	265 720
Proportion	52%	37%	11%

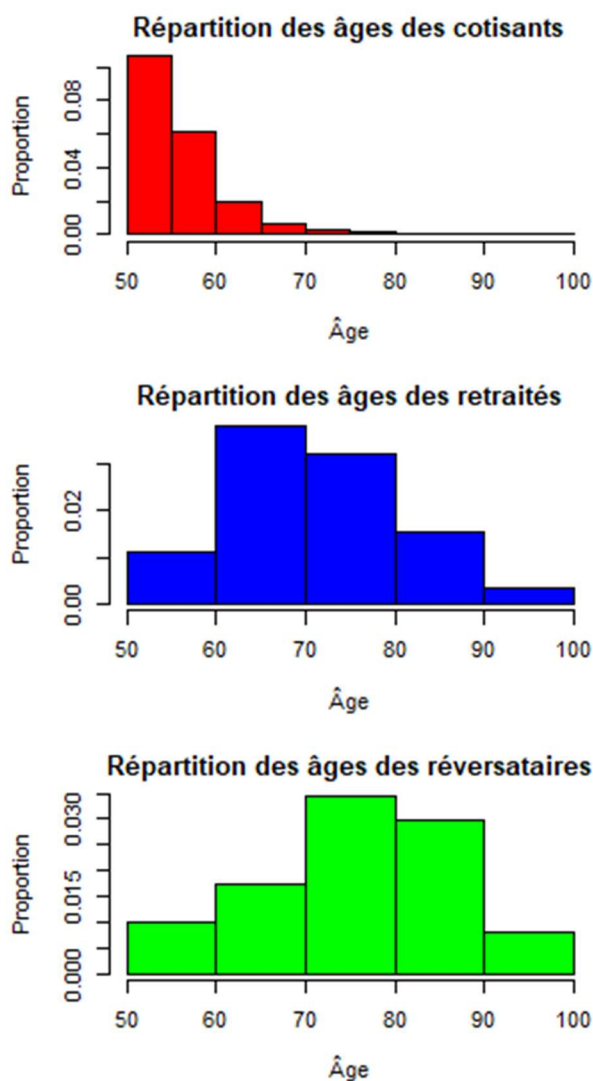


Figure 3-1 : Répartition des adhérents selon leur âge et leur statut.

Les hommes représentent deux tiers des cotisants et plus de la moitié des retraités. En revanche, la quasi-majorité des titulaires des pension de réversion sont des femmes.

Les tableaux ci-dessous présentent les caractéristiques des assurés de notre portefeuille.

Tableau 3-3 : Répartition des assurés par sexe en fonction de leur statut à la date de début d'observation.

Statut	Femme	Homme	Total
Cotisant	37,2 %	62,8 %	100 %
Retraité	44,4 %	55,6 %	100 %
Titulaire de pension de réversion	93,8 %	6,2 %	100 %
Total	51,3 %	48,7 %	100 %

Tableau 3-4 : Répartition des assurés par sexe en fonction de leur motif de sortie à la date de fin d'observation.

Statut	Femme	Homme	Total
Décédé	50,3 %	49,7 %	100 %
Présent	52,0 %	48,0 %	100 %
Absent (hors décès)	42,8 %	57,2 %	100 %
Total	51,3 %	48,7 %	100 %

Tableau 3-5 : Répartition des assurés par motif de sortie en fonction de leur statut à la date de début d'observation.

Statut	Cotisant	Retraité	Titulaire pension de réversion	Total
Décédé	2,4 %	14,2 %	14,4 %	11,4 %
Présent	75,2 %	85,8 %	85,6 %	83,3 %
Absent (hors décès)	22,4 %	0%	0%	5,3%
Total	100 %	100 %	100 %	100 %

3.2. Construction de tables de mortalité d'expérience

3.2.1. Exposition au risque décès : censure et troncature

Un prérequis au calcul des taux bruts est de déterminer, pour une plage d'âges et d'années, les nombres de décès et les expositions. L'exposition représente la population soumise au risque de décès pour chaque année d'observation. Elle est calculée à partir d'informations de chaque individu dans le portefeuille, concernant la date d'entrée dans l'organisme et la date de sortie par rapport à la période d'observation fixée dans notre étude. Or les assurés ne sont pas tous observés depuis la date d'entrée dans l'organisme, ou jusqu'à leur décès, notamment car ils sont encore en portefeuille et ne sont pas décédés, ou qu'ils sont sortis pour une autre cause. D'où l'intérêt de présenter le concept de censures et troncatures, puisque pour toute analyse de données de survie, les données sont généralement incomplètes, la période d'observation étant limitée. Pour autant, même si les données sont incomplètes, il ne faut pas les exclure, étant donné qu'elles contiennent néanmoins de l'information.

Ci-dessous une définition de ces deux concepts :

- Censure à droite : il s'agit du cas où l'assuré n'a pas subi l'évènement (ici le décès) à sa dernière observation. Les données ne sont donc pas complètes. On sait néanmoins qu'elles sont supérieures à une certaine valeur.
- Troncature à gauche : se produit lorsqu'on ne prend en compte que les individus dont les évènements surviennent après un certain point dans le temps, ce qui signifie que les évènements avant ce point ne sont pas observés du tout. Autrement dit, dans l'analyse de la mortalité, on n'observe que les individus qui ont survécu jusqu'à un certain âge (par exemple, 90 ans). Les individus qui sont morts avant cet âge ne sont pas du tout inclus dans l'étude, donc leurs informations ne sont pas disponibles. Cela peut introduire un biais, car les individus ayant subi l'évènement avant ce seuil sont exclus.

Le schéma ci-dessous présente les différentes possibilités d'observation pour un assuré.

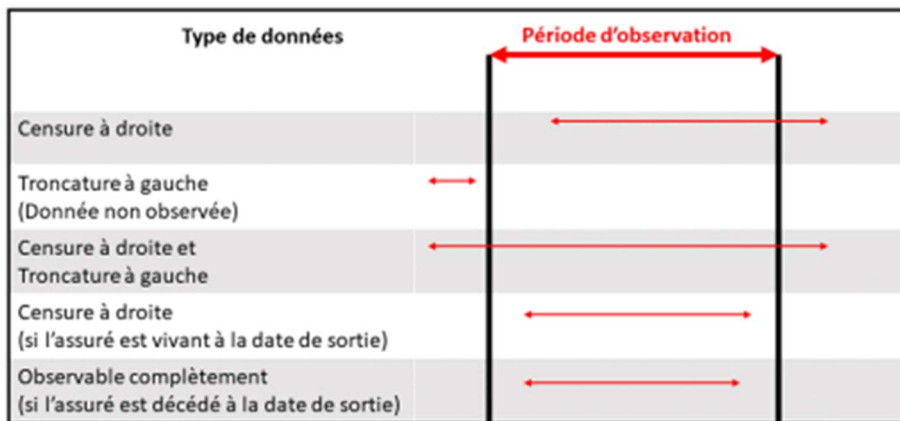


Figure 3-2 : Illustration des censures et troncatures

Ainsi, la détermination pour chaque individu de sa position (observable, tronqué ou censuré) permettra de définir, pour chaque année d'étude et d'âge, le niveau d'exposition et le nombre de décès observés.

Les graphiques suivants nous révèlent que la majorité des expositions et des décès de notre portefeuille ayant des âges compris entre 55 ans et 95 ans.

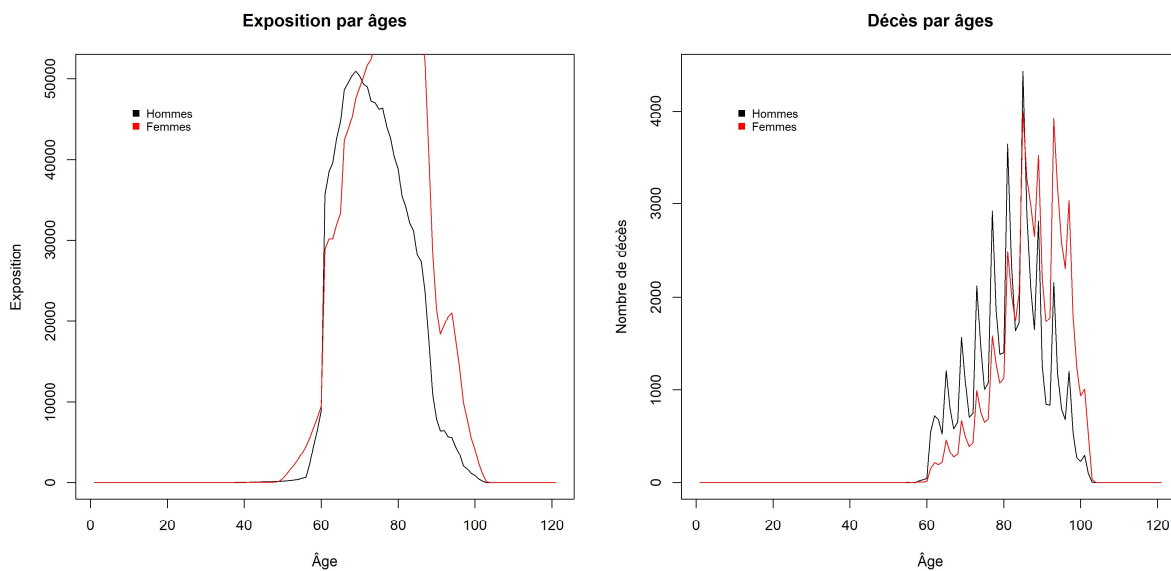


Figure 3-3 : Répartition par âge des expositions et des décès selon le sexe.

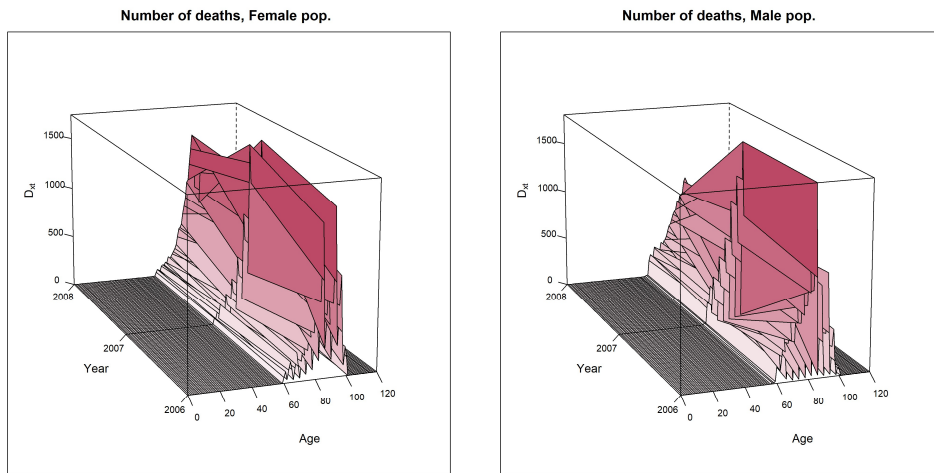


Figure 3-4 : Répartition des décès par âge et année d'observation.

(a) Décès – population féminine

(b) Décès- population masculine

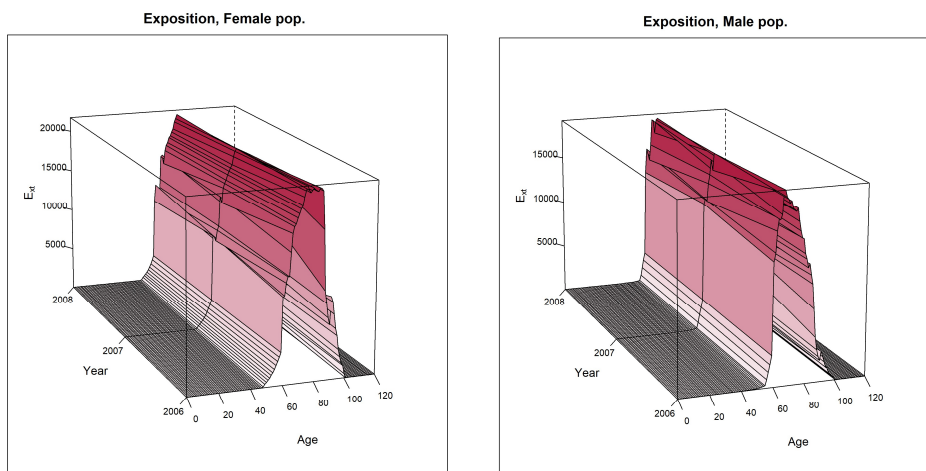


Figure 3-5 : Répartition des expositions par âge et année d'observation.

(c) Exposition – population féminine

(d) Exposition – population masculine

3.2.2. Calcul des taux bruts de mortalité

- **Estimateurs non-paramétriques des taux**

- **Estimateur de Hoem**

La méthode de Hoem se base sur la notion d'exposition au risque. Il s'agit du ratio correspondant au nombre de décès par l'exposition au risque de mortalité pour l'âge correspondant. L'exposition est la quantité en fraction d'années pendant laquelle l'assuré a été observé.

On note $L_{x,t}$ le nombre d'individus d'âge x vivant au 1^{er} janvier de l'année calendaire t . $D_{x,t}$ est le nombre de décès enregistré à l'âge x durant l'année calendaire t d'une exposition au risque $E_{x,t}$. La probabilité qu'un individu d'âge x de décéder durant l'année calendaire t , notée $\hat{q}_{x,t}$, est obtenue à partir de l'estimateur de Hoem :

$$\hat{q}_{x,t} = \frac{D_{x,t}}{E_{x,t}} \quad (3-1)$$

et la probabilité de survie correspondante est :

$$\hat{p}_{x,t} = 1 - \hat{q}_{x,t} \quad (3-2)$$

Cet estimateur a les particularités d'être sans biais et convergent.

➤ Estimateur de Kaplan-Meier

L'estimateur de Kaplan-Meier de la fonction de survie ne fait pas d'hypothèse a priori sur la forme de la loi de mortalité : il s'agit d'un estimateur non paramétrique. Il permet de prendre en compte la censure et la troncature.

Kaplan et Meier proposent une méthode de calcul des lois de survie des individus dans le cas des censures à droite.

Lorsqu'on dispose de plusieurs périodes d'observations, on utilise l'estimateur de Kaplan Meyer.

Supposons qu'on observe N personnes de même âge x pendant k périodes successives (par exemple $k=12$ mois) On suppose que les censures (sorties en vie moins entrées en vie) interviennent à la fin de chaque sous-période et les décès en milieu (autrement dit elles sont uniformément réparties). Si n_{i+1} désigne l'effectif présent à la fin de la sous période i (après départ des censures) :

$$n_{(i+1)} = n_i - d_i - c_{(i+1)} \quad (3-3)$$

Dans ces notations $N = n_0$, d_0 sont les décès de la première sous période et c_1 les premières censures qui interviennent à la fin de la première sous période. L'estimateur de la probabilité de survie à la date t est le produit des probabilités de survie de toutes les sous périodes antérieures. La fonction de survie de Kaplan-Meier s'écrit :

$$\hat{S}_t = \prod_{i=0}^{t-1} \left(1 - \frac{d_i}{n_i}\right) = \prod_{i=0}^{t-1} \frac{(n_i - d_i)}{n_i} \quad (3-4)$$

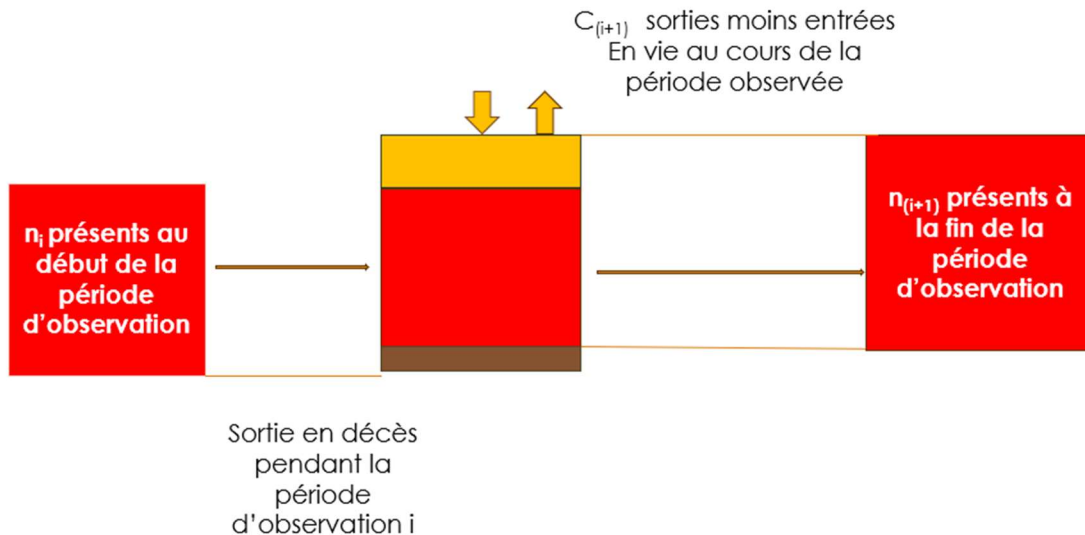


Figure 3-6 : Illustration du nombre des présents et des sorties sur la période d'observation.

Et la variance est donnée par :

$$VAR(S_t) = S_t^2 \times \sum_{i=0}^{t-1} \frac{d_i}{n_i (n_i - d_i)} \quad (3-5)$$

dite formule de Greenwood.

Cette formule traduit tout simplement le fait que pour être survivant à une date t , il faut avoir survécu pendant toutes les périodes $]a_i, a_{i+1}]$ antérieures à cette date t . Cet estimateur de la fonction de survie est également sans biais. De plus sa variance s'écrit :

A partir de l'estimateur de la fonction de survie, on peut aisément déduire les taux bruts de mortalité à travers la relation :

$$q_x = \frac{S_x - S_{x+1}}{S_x} \quad (3-6)$$

Soit :

$$q_x = 1 - \prod_{i=1}^k \frac{n_i - d_i}{n_i} \quad (3-7)$$

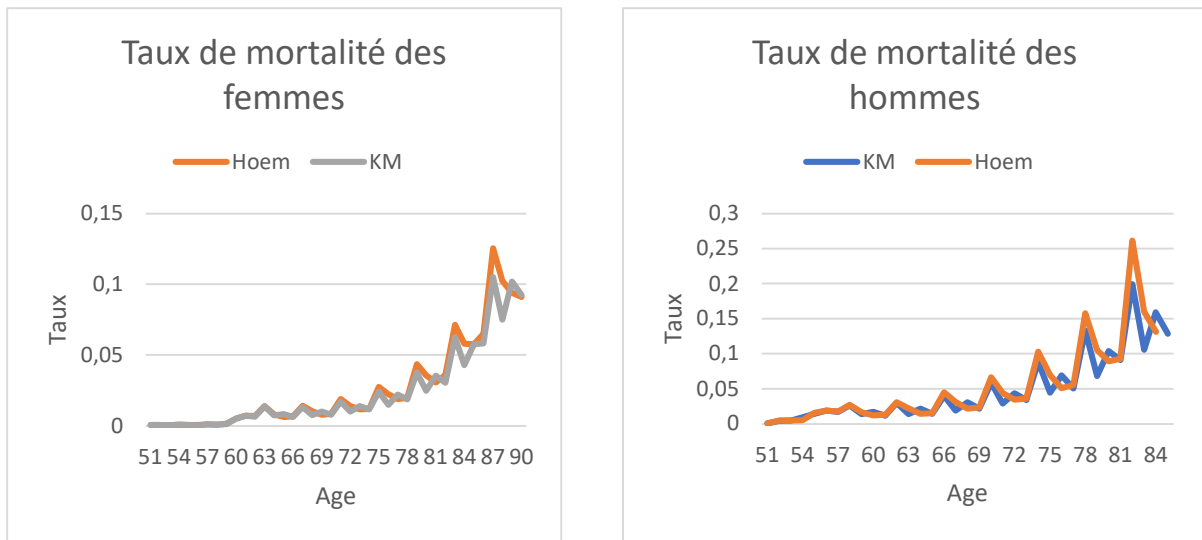


Figure 3-7 : Quotients de mortalité par âge selon les deux méthodes : Koem et Kaplan-Meier.

(a) Population féminine

(b) Population masculine

On constate que les taux bruts de mortalité obtenus par les deux méthodes : Koem et Kaplan-Meier sont proches mais irréguliers. Il est donc nécessaire de régulariser ces valeurs brutes pour obtenir une courbe lisse qui reflète la mortalité observée de notre portefeuille.

Les deux graphiques mettent en évidence des irrégularités en fonction des âges. Aux grands âges, la volatilité des taux de mortalité augmente. La population sous risque diminue et dans le même temps le nombre de décès augmente ce qui provoque des fluctuations d'échantillonnage importantes.

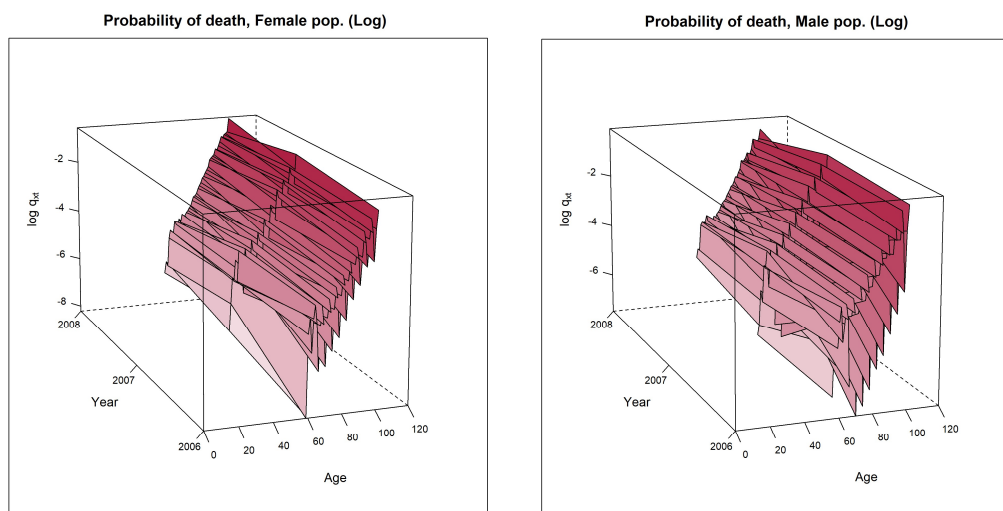


Figure 3-8 : Quotients de mortalité à l'échelle logarithmique par âge et année d'observation.

(a) Population féminine

(b) Population masculine

La volatilité des taux de mortalité au-delà d'un certain âge ne permet pas d'obtenir un ajustement fiable. D'où le recours au critère de Cochran pour fixer la plage d'âges d'ajustement des taux de mortalité.

- **Critère de Cochran**

Pour établir une table de mortalité d'expérience de qualité, il est nécessaire de vérifier que les nombres d'individus exposés et de décès soient suffisamment élevés. Pour ce faire, on utilise le critère de Cochran qui stipule que les conditions suivantes doivent être vérifiées :

condition $N_x \times (1 - \hat{q}_x) \geq 5$ et la condition, $N_x \times \hat{q}_x \geq 5$

où N_x est l'effectif en vie au 1er janvier de l'année d'observation et \hat{q}_x le taux de mortalité empirique à l'âge x obtenue par la méthode de Hoem.

Tableau 3-6 : Critère de Cochran.

Age	2007_H		2007_F		2008_H		2008_F	
	$N_x \times q_x$	$N_x \times (1-q_x)$	$N_x \times q_x$	$N_x \times (1-q_x)$	$N_x \times q_x$	$N_x \times (1-q_x)$	$N_x \times q_x$	$N_x \times (1-q_x)$
50	0	77	0	367	0	61	0	103
55	0	230	0	1 576	0	201	1	1 310
60	226	13 416	61	11 104	173	12 608	43	10 198
65	251	17 970	112	15 693	289	18 461	116	15 897
70	381	17 879	196	18 235	317	18 220	170	17 636
75	31	16 452	133	20 917	531	16 697	315	20 236
80	1 461	12 879	1 049	20 670	699	12 988	631	20 204
85	1 060	9 741	1 343	20 433	887	10 057	1 149	19 558
90	369	2 145	780	6 455	472	2 749	872	7 585

La condition $N_x \times (1 - q_x) \geq 5$ est vérifiée quel que soit l'âge de 50 à 90 ans, pour les hommes et les femmes. Par contre la condition, $N_x \times q_x \geq 5$, n'est vérifiée que pour les âges à partir de 60 ans pour les deux sexes.

Néanmoins, on décide de retenir la tranche d'âge 50-90 ans pour plusieurs raisons. Tout d'abord, cela nous permet d'être en adéquation avec la limite d'âge choisie pour la construction des tables dites « HMD » du chapitre 2. En outre, la liquidation de la rente ne peut être versée qu'au plus tôt à 55 ans. C'est de ce fait l'âge de début d'observation qui a été retenu. L'âge minimal d'ajustement sera donc fixé à 55 ans.

Donc, on réalise l'ajustement sur la plage d'âge 55-90 et pour la période 2007-2008, pour les hommes, femmes et tous sexes confondus. Afin de disposer d'un volume de données suffisant, sans trop restreindre la période d'observation, la tranche d'âge qui sera retenue pour le calibrage de la loi de mortalité est la tranche 50 – 90 ans pour la période 2007-2008, pour les hommes, femmes et tous sexes confondus.

3.3. Synthèse des travaux antérieurs de construction des tables de mortalité dans le cadre des mémoires d'actuariat en France

Dans un contexte post-COVID pour des portefeuilles de rentiers en Angleterre et au Pays de Galles, un mémoire d'actuariat réalisé par Ouhammou A. (2023) porte sur la construction de tables de mortalité

prospectives Best Estimate. L'approche retenue consiste à calibrer des tables de mortalité sur la population générale pour laquelle l'auteur dispose d'un historique important à l'aide d'un modèle stochastique de mortalité. Ensuite, pour tenir compte de la mortalité spécifique des populations rentières, il construit une table de mortalité sur une période de 5 ans à partir de données de 7 portefeuilles avant de procéder à un ajustement à l'aide du modèle de Brass sur les données HMD de la population générale.

Dans le cadre de la nouvelle réglementation Solvabilité 2 et dont le but d'évaluer l'impact des tables de mortalité d'expérience sur l'engagement de l'assureur, Martinez L. (2015) a procédé, à la construction de tables de mortalité d'expérience en ajustant les taux de mortalité bruts par ceux de la table de mortalité de référence « TG05 ». L'étude de la mortalité du portefeuille, composé de deux produits de retraite supplémentaire de la mutuelle « Union Mutualiste Retraite », a révélé une sous-mortalité par rapport aux tables réglementaires TG05.

Aussi, Damasse A. (2015) a procédé à la construction de tables de mortalité d'expérience dans le cadre de son mémoire d'actuariat. Il s'est appuyé sur les données démographiques de l'Human Mortality Database pour construire ses tables de mortalité de référence. L'auteur a testé deux modèles différents : le premier modèle considéré est paramétrique et ajuste les logarithmes des quotients de mortalité à la fonction logistique. Quant au second modèle, il est semi-paramétrique et explore une approche basée sur le modèle SARIMA issu de la théorie des séries temporelles. Ensuite, Damasse A. a examiné l'impact des tables d'expérience précédemment obtenues en examinant leur impact respectif sur les provisions techniques. Cet impact porte sur les rentes en cours de service mais il s'étend également aux rentes en phase d'épargne pour lesquelles une garantie de table de mortalité existe.

Quant à l'étude de Fall N.-Y. (2020), elle a pour objectif la construction des tables de mortalité prospectives pour l'évaluation de la longévité à travers la mise en d'un indicateur de tendance de la longévité utilisé dans le cadre de la certification de la table d'expérience. Le champ de l'étude concerne les données du portefeuille épargne retraite de PREDICA. Les modèles Lee-Carter, LogPoisson, Cairns-Blake-Dowd sont présentés et appliqués à ces données ainsi que des modèles faisant appel aux tables de référence. La méthode de Denuit et Goderniaux est utilisée pour projeter les taux de mortalité aux grands âges. Une projection des taux de mortalité à l'aide d'une modélisation ARIMA (p,d,q) a été réalisée. Aussi deux indicateurs de tendance de la longévité sont présentés. L'un passe par l'évolution du nombre de décès dans le temps pour une exposition fixée, l'autre par un facteur de réduction des taux de mortalité.

Le mémoire d'actuariat de Biagini N. (2020) porte sur la modélisation de la mortalité sur un portefeuille de retraite collective. Dans ce cadre, différentes méthodes ont été testées afin d'étudier la mortalité du portefeuille de Retraite Collective de CNP Assurances et réaliser les lois d'expérience. Il s'agit de méthodes de construction endogènes (directement à partir des taux bruts de mortalité observés sur le portefeuille), et exogènes (par positionnement à une table de référence, ici les tables réglementaires des rentes viagères TGH05/TGF05). Néanmoins, les méthodes de construction exogènes ont été privilégiées compte tenu de la volumétrie du portefeuille. Ainsi, trois modèles ont été testés et analysés : une approche par le SMR, une régression logistique et un modèle Log Poisson. L'étude a mis en évidence une surmortalité du portefeuille étudié par rapport aux tables réglementaires.

Chapitre 4 : Construction des tables de mortalité d'expérience prospectives : les modèles de positionnement

Une fois les taux bruts de mortalité déterminés, il existe deux types de méthodes permettant d'obtenir la loi de mortalité d'expérience du portefeuille étudié :

- Les méthodes endogènes, obtenues par un lissage des taux bruts observés. La loi est ainsi obtenue uniquement avec les données intrinsèques au portefeuille.
- Les méthodes exogènes, déterminées par positionnement par rapport à une table de mortalité de référence (Planchet F., Tomas J. (2013a, 2013b, 2013c)).

Ici, le second type de méthodes a été retenue compte tenu de la profondeur du portefeuille étudié qui est constituée de trois années de 2006 à 2008. En effet, la première option nécessite d'avoir un portefeuille relativement conséquent aussi bien dans la profondeur de l'historique, que dans l'exposition par âge et par année. Les tables d'expérience seront construites d'un côté par sexe et d'autre côté tous sexes confondus. Aussi, on prend en compte l'impact de Covid 19 sur ces tables en prenant en compte deux options : l'une avant la covid 19 qui se base sur la table « HMD » construite dans le chapitre 2 de ce mémoire sur la base de l'historique 1982-2019 de la population française ; et l'autre y compris covid en ajoutant les données de deux années 2020 et 2021.

4.1. Les tables de références externes

les deux tables de référence sont d'une part, la table dite de « HMD » que j'ai construite dans le chapitre 2 de ce mémoire et d'autre part la table Insee 2007-2060. Ces deux tables concernent la population française. Néanmoins, la première a été construite en s'appuyant sur l'historique 1982-2019 ou 2021 et projection jusqu'en 2060 par sexe et tous sexes confondus, alors que la seconde est fondée sur les estimations de population et statistiques de l'état civil jusqu'en 2007 et projection de population par sexe sur la période 2007-2060.

- **Table dite « HMD » : avant/y compris covid et par sexe/non générées**

Trois tables construites à partir des données HMD pour les hommes, femmes et tous sexes confondus (voir chapitre 2) selon deux cas de figure :

- Avant covid 19 : fondée sur l'historique 1982-2019,
- Y compris covid qui a pris en compte , l'historique 1982-2021 en intégrant donc les données des 2020 et 2021.

- **Tables Insee prospectives 2007-2060**

Il s'agit du scénario médian des projections démographiques de la population française sur la période 2007 – 2060 fournie par l'INSEE, Blanpain and Chardon (2010). Ces projections sont fondées sur des hypothèses concernant la fécondité, la mortalité et les migrations :

- le taux de fécondité reste à un taux élevé (1.95) jusqu'en 2060.
- La diminution de la mortalité par sexe et âge donné est plus important pour les hommes après 85 ans.
- Une immigration net annuelle constante de 100 000 individus.

4.2. Trois méthodes de positionnement testées

Ces méthodes permettent de lisser les données brutes et de retenir, comme dérive, la dérive de la table de référence.

Dans cette partie, je présente de manière succincte les trois méthodes de positionnement et le processus de validation implémentés dans le package « ELT » du logiciel R.

- **Méthode 1 : méthode SMR (Ratio de Mortalité Standardisé)**

Cette méthode est non paramétrique. Elle ne fait aucune hypothèse sur la loi sous-jacente des décès.

Elle consiste à appliquer un seul facteur de réduction ou majoration aux taux de mortalité de la table de référence q_x^{ref} .

En pratique, ce coefficient est le Ratio de Mortalité Standardisé (SMR) de la population considérée, ici calculé au niveau global en effectuant le rapport entre les décès observés et les décès attendus par la table à abattre. Ensuite, on obtient les probabilités de décès du portefeuille, notées $q_{x,t}$ pour les âges 50-90 et les années 2020 (ou 2022) à 2060 selon les cas (avant covid ou y compris covid) pour les hommes, femmes ou tous sexes confondus par :

$$q_{x,t} = SMR \times q_{x,t}^{ref} \quad (4-1)$$

Avec :

$$SMR = \frac{\sum_{(x,t)} D_{x,t}}{\sum_{(x,t)} E_{x,t} \times q_{x,t}^{ref}} \quad (4-2)$$

Où :

$D_{x,t}$: Nombre de décès observés sur le portefeuille à l'âge x et durant l'année t

$E_{x,t}$: Exposition observée sur le portefeuille à l'âge x , l'année t

$q_{x,t}^{ref}$: Probabilité de décès à l'âge x et l'année t dans la table de référence ; ici, la table « HMD » ou bien celle de Insee prospective 2007-2060.

▪ **Méthode 2 : Variante du modèle relationnel semi-paramétrique de Brass (1971)**

Il s'agit d'une approche à deux paramètres α et β où le paramètre β permet de prendre en compte les évolutions de la mortalité selon l'âge.

L'avantage d'un modèle semi-paramétrique relationnel est d'intégrer l'estimation et l'extrapolation des tendances de mortalité⁵.

Le positionnement est effectué à l'aide de la fonction logistique via une relation linéaire entre le logit des taux ajustés et le logit des taux de référence :

$$\text{logit}(\tilde{q}_{x,t}) = \log\left(\frac{\tilde{q}_{x,t}}{1 - \tilde{q}_{x,t}}\right) \quad (4-3)$$

et le modèle est :

$$\text{logit}(q_{x,t}) = \alpha + \beta \times \text{logit}(q_{x,t}^{ref}) + \varepsilon_{x,t} \quad (4-4)$$

où :

$\varepsilon_{x,t}$ est le terme d'erreur,

α : paramètre qui affecte tous les âges de la même façon ;

x^* , t^* correspondent aux âges et années en communs entre la table des taux bruts de mortalité d'expérience et la table de référence.

L'estimation est effectuée par une minimisation de la distance pondérée par l'exposition entre les probabilités de décès observées ($\hat{q}_{x,t}$) et estimés ($\tilde{q}_{x,t}$) :

$$\sum_{x,t} |E_{x,t} \times (\hat{q}_{x,t} - \tilde{q}_{x,t})| \quad (4-5)$$

Les probabilités de décès ajustées se déduisent sur la plage d'âge 50- 90 ans et sur la période 2020/2022-2060 par :

$$\tilde{q}_{x,t} = \frac{\exp(\hat{\alpha} + \hat{\beta} \times \text{logit}q_{x,t}^{ref})}{1 + \exp(\hat{\alpha} + \hat{\beta} \times \text{logit}q_{x,t}^{ref})} \quad (4-6)$$

Avec $\hat{\alpha}$ et $\hat{\beta}$ les estimateurs des paramètres α et β .

⁵ Planchet, F. and Thérond, P. (2011). Modélisation statistique des phénomènes de durée - Applications actuarielles. Assurance Audit Actuariat. Economica Paris.

▪ **Méthode 3 : un modèle linéaire généralisé de Poisson**

On suppose que les décès suivent une loi de Poisson. Le nombre de décès du portefeuille à l'âge x^* et l'année calendaire t^* est ainsi déterminé par :

$$D_{x^*,t^*} \sim P(E_{x^*,t^*} \times q_{x^*,t^*}) \quad (4-7)$$

Avec :
$$\mu_{x^*,t^*} = \beta_0 + \beta_1 \log q_{x^*,t^*}^{ref} + \beta_2 x^* + \beta_3 t^* + \beta_4 x^* t^* \quad (4-8)$$

et
$$\mu_{x^*,t^*} = \log q_{x^*,t^*} \quad (4-9)$$

est le taux de décès à l'âge x^* et pendant l'année calendaire t^* qui sont respectivement la plage d'âges et la période d'observation commune entre les données observées et la mortalité de référence.

Il est à noter que les interactions entre l'âge et l'année calendaire n'ont significativement de sens que lorsque nous disposons d'un historique commun avec la référence considérée supérieur à 10 ans. Ce n'est pas le cas ici puisque l'étude est réalisée sur la période 2006-2008, donc l'ajustement n'en tiendra pas compte. Autrement dit, les paramètres $\hat{\beta}_3$ et $\hat{\beta}_4$ ne sont pas calculés.

Les taux de mortalité ajustés sont donc obtenus sur la plage d'âges considérée et sur toutes les années de la table de référence par :

$$\tilde{q}_{x,t} = \exp(\hat{\beta}_0 + \hat{\beta}_1 \log q_{x,t}^{ref} + \hat{\beta}_2 x) \quad (4-10)$$

4.3. Application et analyse des résultats des trois méthodes d'ajustement

4.3.1. Estimation du SMR

Tableau 4-1 : SMR selon la table de mortalité de référence retenue.

Méthode 1 : SMR	INSEE		"HMD" avant covid			"HMD" y compris covid		
	H	F	H	F	Total	H	F	Total
SMR	0,9805	1,0163	1,0138	1,0198	1,0231	1,0136	1,0189	1,0233

Nous remarquons que le SMR est inférieur à 1 pour les hommes en optant la table de référence de l'Insee, en conséquence la table de référence dite « Insee » sur-estime le nombre de décès de notre portefeuille pour la population masculine. Alors que la table HMD sous-estime légèrement les décès puisque les SMR sont supérieurs à 1 pour les hommes, les femmes et tous sexes confondus. Comme ce mémoire s'inscrit dans le cadre du risque de longévité, donc un tel modèle nous permettra d'être prudent.

4.3.2. Estimation des paramètres du modèle logistique

Cette méthode a été appliquée à l'aide de l'estimateur de Hoem sur l'ensemble du portefeuille étudié.

Les tableaux suivants présentent les valeurs des paramètres α et β du modèle de positionnement logistique pour les populations masculine, féminine et tous sexes confondus du portefeuille.

Tableau 4-2 : Valeurs des paramètres α et β du modèle de positionnement logistique.

Hommes	Alpha	Bêta
Insee 2007-2060	-0,2927	0,9238
HMD 2019-2060	-0,0152	1,0017
HMD 2021-2060	-0,0226	0,9993

Femmes	Alpha	Bêta
Insee 2007-2060	-0,2998	0,8995
HMD 2019-2060	-0,1131	0,9664
HMD 2021-2060	-0,1234	0,9630

Tous sexes confondus	Alpha	Bêta
HMD 2019-2060	-0,0513	0,9911
HMD 2021-2060	-0,0660	0,9859

Pour les différentes tables de référence et pour les différentes populations à l'exception des hommes avant covid19, le coefficient β (valeur de décalage) est inférieur à 1. Cela signifie une surmortalité des assurés du portefeuille par rapport celle de la table de référence. En effet, plus le coefficient β est élevé, plus le taux de mortalité ainsi transformé est faible via la formule du Logit. Le paramètre α , ici négatif dans tous les cas de figure, affecte tous les âges pour la même valeur.

Les modèles fondés sur les logits des données de mortalité entraînent une marge de sécurité dans le cas du risque de longévité qu'on traite ici dans le cadre de ce mémoire.

4.3.3. Estimation des paramètres du modèle linéaire généralisé de Poisson

Les tableaux suivants présentent les valeurs des paramètres β_0 , β_1 et β_2 du modèle linéaire généralisé de Poisson pour la population masculine et féminine respectivement selon les deux tables de référence « HMD » et « Insee ».

Tableau 4-3 : Paramètres du modèle linéaire généralisé de Poisson pour les assurés.

Hommes	β_0	β_1	β_2
Insee 2007-2060	3.39634	1.24977	-0.03417
HMD 2019-2060	13.1040	2.1969	-0.1215
HMD 2021-2060	12.9736	2.1812	-0.1204

Femmes	β_0	β_1	β_2
Insee 2007-2060	-2.50568	0.74507	0.02098
HMD 2019-2060	0.80437	1.01819	-0.00888
HMD 2021-2060	0.93064	1.02540	-0.01015

Tous sexes confondus	β_0	β_1	β_2
HMD 2019-2060	6.15892	1.48816	-0.05762
HMD 2021-2060	5.92128	1.46724	-0.05547

Ces résultats seront analysés dans la section suivante.

4.4. Validation des tables et choix du modèle

4.4.1. Validation visuelle des ajustements de la mortalité passée au niveau local

▪ Comparaison des ajustements pour l'année 2008

Les graphiques suivants permettent de mettre en exergue la forte cohérence des taux de mortalité calculés de notre portefeuille avec ceux issus des tables INSEE Prospectives 2007-2060 prises en compte par le Conseil d'Orientation des Retraites (COR) pour les projections à moyen et long terme des régimes de retraite en France.

Ces graphiques présentent les probabilités de décès selon les trois méthodes, pour la plage d'âges 50-90 sur l'année 2008, pour la population masculine et féminine. Ils nous donnent une première indication de la qualité de l'ajustement au niveau local.

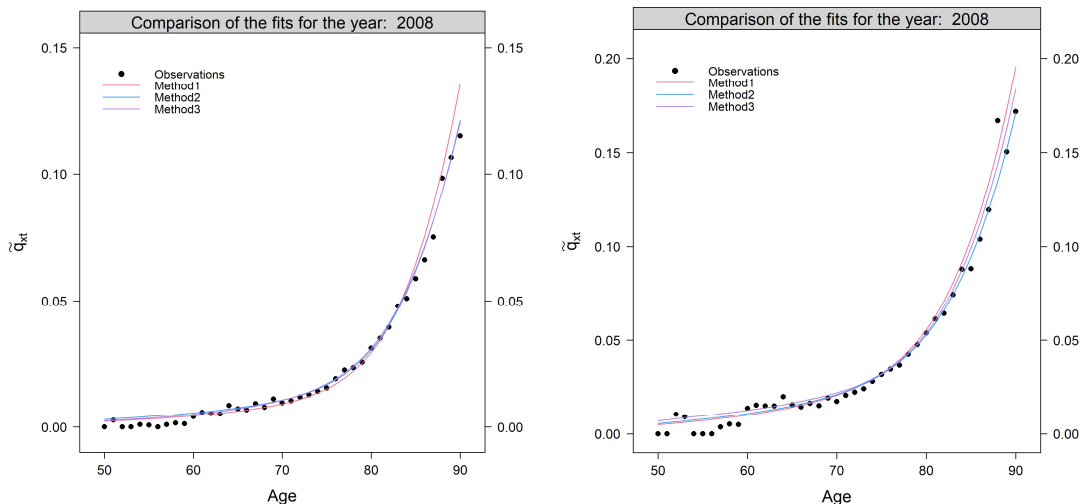


Figure 4-1 : Comparaison des quotients de mortalité par âge pour l'année 2008 observés et ajustés selon les trois méthodes

(a) Insee – population féminine

(b) Insee - population masculine

Il n'est à priori pas aberrant de construire des tables de mortalité prospectives propres aux adhérents de notre portefeuille en ajustant les tables de mortalité de référence : INSEE prospectives vues en 2007 (tables instantanées) ou bien les tables dites « HMD » en fonction de tables instantanées construites pour le portefeuille à notre disposition sur la période 2006 à 2008.

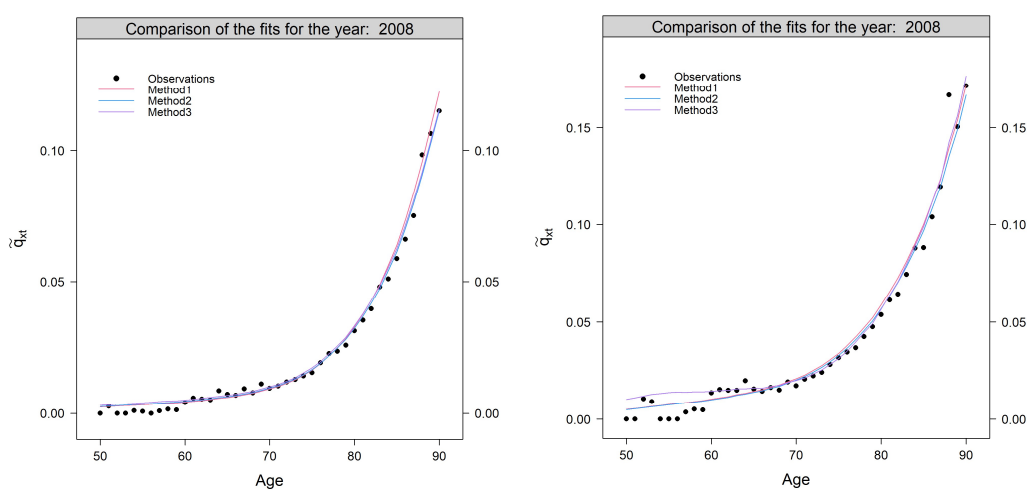


Figure 4-2 : Comparaison des quotients de mortalité par âge pour l'année 2008 observés et ajustés selon les trois méthodes

(c) HMD - population féminine

(d) HMD - population masculine

4.4.2. Résultats des tests de proximité

- Test SMR

Comme évoqué précédemment, si le SMR > 1 les décès ajustés sont sous-estimés et réciproquement si SMR < 1. Nous pouvons aussi appliquer un test afin de déterminer si le SMR est significativement différent de 1.

La statistique du test s'écrit de la façon suivante :

$$\text{Si SMR} > 1, N^{SMR} = 3 \times D^{\frac{1}{2}} \times \left(1 - (9D)^{-1} - (D/E)^{\frac{1}{3}} \right) \quad (4-11)$$

$$\text{Si SMR} < 1, N^{SMR} = 3 \times D^{\frac{1}{2}} \times \left((D^*/E)^{\frac{1}{3}} + (9D^*)^{-1} - 1 \right) \quad (4-12)$$

Avec :

$$D = \sum_{(x,t)} D_{(x,t)} \quad (4-13)$$

$$D^* = \sum_{(x,t)} D_{(x,t)} + 1$$

$$E = \sum_{(x,t)} -E_{(x,t)} \tilde{q}_x(t)$$

Sous l'hypothèse nulle H0, le SMR n'est pas significativement différent de 1. La statistique N^{SMR} suit une loi Normale centrée réduite :

$$N^{SMR} \sim N(0,1) \quad (4-14)$$

Ainsi, l'hypothèse nulle H0 sera rejetée si :

$$N^{SMR} > N_{1-\alpha}(0,1) \quad (4-15)$$

Où $N_{(1-\alpha)}(0,1)$ est le $(1-\alpha)$ quantile de la distribution Normale centrée réduite. La p-valeur est alors donnée par :

$$p - \text{valeur} = 1 - F_{N(0,1)}(N^{SMR}) \quad (4-16)$$

Nous chercherons à avoir la p-valeur la plus proche de 1.

▪ Test de Wilcoxon

Le test de Wilcoxon est une adaptation du test des signes qui consiste à prendre en compte la valeur de l'écart entre les taux de mortalité bruts et lissés pour chaque âge. Ainsi, si on note $w +$ la somme des écarts positifs et $w -$ la somme des écarts négatifs, le test consiste à tester l'hypothèse nulle suivante : $w + = w -$.

Tableau 4-4 : Résultats des tests de SMR et Wilcoxon.

			Test de Wilcoxon			Test de SMR		
			W	Hyp. Retenue	P-value	SMR	Hyp. Retenue	P-value
Hommes	Insee	Méthode-SMR	2 149	H1	0,0388	0,9805	H0	0,4993
		Méthode-BRASS	1 858	H0	0,4708	1,0491	H1	0
		Méthode-GLM-Poisson	2 255	H1	0,0106	1	H0	0,4985
	MD avant covid	Méthode-SMR	2 188	H1	0,0247	1,0138	H0	0,4993
		Méthode-BRASS	1 879	H0	0,4132	1,0332	H1	0
		Méthode-GLM-Poisson	2 346	H1	0,0029	1	H0	0,4985
	HMD yc covid	Méthode-SMR	2 199	H1	0,0216	1,0136	H0	0,4993
		Méthode-BRASS	1 893	H0	0,3772	1,0333	H1	0
		Méthode-GLM-Poisson	2 349	H1	0,0028	1	H0	0,4985
Femmes	Insee	Méthode-SMR	1 941	H0	0,2692	1,0163	H0	0,4993
		Méthode-BRASS	2 235	H1	0,0137	1,0037	H0	0,258
		Méthode-GLM-Poisson	2 216	H1	0,0175	1	H0	0,4985
	MD avant covid	Méthode-SMR	2 057	H0	0,1008	1,0198	H0	0,4993
		Méthode-BRASS	1 875	H0	0,4238	1,0276	H1	0
		Méthode-GLM-Poisson	2 167	H1	0,0316	1	H0	0,4985
	HMD yc covid	Méthode-SMR	2 054	H0	0,1037	1,0189	H0	0,4993
		Méthode-BRASS	1 885	H0	0,3976	1,0263	H1	0
		Méthode-GLM-Poisson	2 163	H1	0,0331	1	H0	0,4993
Tous sexes confondus	MD avant covid	Méthode-SMR	2 204	H1	0,0203	1,0231	H0	0,4995
		Méthode-BRASS	1 804	H0	0,6373	1,0466	H1	0
		Méthode-GLM-Poisson	2 293	H0	0,0063	1	H0	0,499
	HMD yc covid	Méthode-SMR	2 212	H1	0,0184	1,0233	H0	0,4995
		Méthode-BRASS	1 815	H0	0,6014	1,0457	H1	0
		Méthode-GLM-Poisson	2 308	H1	0,0051	1	H0	0,499

Le premier test effectué est le test du SMR qui valide l'ajustement global des données dans le cas de deux méthodes : SMR et GLM-Poisson et uniquement la méthode BRASS dans le cas de la table Insee pour les femmes. Dans tous les cas : pour les hommes, femmes, tous sexes confondus, avant et y compris covid 19, les p-value sont très proches.

Le test de Wilcoxon, qui permet d'affirmer que les probabilités de décès observés et ajustés sont équivalentes, valide la méthode de BRASS.

▪ Valeurs d'intérêt

Afin de déterminer le modèle final, nous allons définir certaines métriques de choix du modèle. Le modèle choisi sera celui qui détient la meilleure valeur de critères définis dans la majorité.

➤ La statistique de R^2

D'une valeur comprise entre 0 et 1, le coefficient de détermination mesure l'adéquation entre le modèle et les données observées. Le R^2 se définit comme la part de variance expliquée par rapport à la variance totale.

$$R^2 = 1 - \left(\frac{\sum_{(x,t)} \left(\frac{D_{x,t}}{E_{x,t}} - q_x(t) \right)^2}{\sum_{(x,t)} \left(\frac{D_{x,t}}{E_{x,t}} - \sum_{(x,t)} \frac{D_{x,t}}{E_{x,t} N} \right)^2} \right) \quad (4-17)$$

où N est le nombre d'observations.

- La statistique de χ^2

Le test d'adéquation du χ^2 va nous permettre de voir qu'il n'y a pas eu de déformation non aléatoire lors du lissage des taux bruts de mortalité.

Cet indicateur permet de mesurer la qualité de l'ajustement du modèle. La valeur du χ^2 est :

$$\chi^2 = \sum_{(x,t)} \frac{(D_{x,t} - E_{x,t} q_x(t))^2}{E_{x,t} q_x(t)} \quad (4-18)$$

Cet indicateur doit suivre une loi du χ^2 à 41-(nombre de paramètres estimés) degrés de liberté puisque nous avons 90-50 + 1 = 41 âges pour lesquels on lisse les taux bruts.

On privilégiera le modèle ayant la valeur du χ^2 la plus faible.

- La statistique de MAPE

Il s'agit d'une mesure de l'exactitude de l'ajustement par rapport aux observations.

Cet indicateur correspond à la moyenne des écarts en valeur absolue par rapport aux valeurs observées.

$$MAPE = \frac{\sum_{(x,t)} \left| \frac{\frac{D_{x,t}}{E_{x,t}} - q_x(t)}{\frac{D_{x,t}}{E_{x,t}}} \right|}{\sum_{(x,t)} \frac{D_{x,t}}{E_{x,t}}} \times 100 \quad (4-19)$$

C'est donc un pourcentage et par conséquent un indicateur pratique de comparaison. Néanmoins, en présence d'observations nulles il y aura une division par zéro et ces observations doivent être retirées.

- Application numérique : valeurs d'intérêt

Tableau 4-5 : Valeurs d'intérêt selon les tables de référence retenues et les trois méthodes d'ajustement.

			SMR	Khi-2	R ²	MAPE en %
Hommes	Insee	Méthode-SMR	0,9805	8126,06	0,7492	148,26
		Méthode-BRASS	1.0491	8359,62	0,7463	146,36
		Méthode-GLM-Poisson	1	7878,11	0,7505	157,58
	HMD avant covid	Méthode-SMR	1,0138	7 995,450	0,7480	151,14
		Méthode-BRASS	1,0332	8285,35	0,7448	145,88
		Méthode-GLM-Poisson	1	7687,61	0,7445	164,79
	HMD yc covid	Méthode-SMR	1,0136	7 997,030	0,7482	151,08
		Méthode-BRASS	1,0333	8282,07	0,7448	145,97
		Méthode-GLM-Poisson	1	7684,88	0,7439	165,28
Femmes	Insee	Méthode-SMR	1,0163	3039,77	0,9093	58,61
		Méthode-BRASS	1.0037	2691,47	0,9163	74,18
		Méthode-GLM-Poisson	1	2687,66	0,9168	63,58
	HMD avant covid	Méthode-SMR	1,0198	2 923,060	0,9155	57,78
		Méthode-BRASS	1,0276	2922,33	0,9115	60,38
		Méthode-GLM-Poisson	1	2786,53	0,9123	68,94
	HMD yc covid	Méthode-SMR	1,0189	2 945,450	0,9155	57,52
		Méthode-BRASS	1,0263	2925,59	0,9115	60,6
		Méthode-GLM-Poisson	1	2795,5	0,9122	69,42
Tous sexes confondus	HMD avant covid	Méthode-SMR	1,0231	10 319,020	0,8455	95,04
		Méthode-BRASS	1.0466	10847,52	0,8419	91,5
		Méthode-GLM-Poisson	1	9831	0,8408	132,8
	HMD yc covid	Méthode-SMR	1,0233	10 310,020	0,8458	94,94
		Méthode-BRASS	1,0457	10784,46	0,842	92,31
		Méthode-GLM-Poisson	1	9828,93	0,8414	131,54

Nous obtenons des résultats globalement satisfaisants. En effet, un MAPE faible correspond à un faible écart entre les valeurs observées et celles ajustées notamment ici pour la population féminine quel que soit la méthode choisie. De plus, la valeur du $R^2 > 0,75$ permet de conclure à un bon modèle au vu de ce critère notamment dans le cas des femmes où elle est supérieure à 0,91.

La valeur du χ^2 indique que les ajustements sont proches par sexe entre les trois méthodes quel que soit la table de mortalité retenue et donnent des résultats quasi équivalents. Le χ^2 calculé sur l'ajustement par les différentes méthodes pour la population féminine sont les plus faibles. Cela révèle que les valeurs estimées se rapprochent le plus des données observées.

4.4.3. Résultats des tests de régularité et d'ajustement

- **Tests des runs**

Le test des runs est un test non paramétrique. Il détermine si les éléments d'une séquence sont mutuellement indépendants. Un run se définit comme le segment d'une séquence, construit avec des éléments avoisinants qui sont égaux.

On teste ainsi les hypothèses suivantes :

H_0 : le nombre de runs d'une séquence de n éléments est une variable aléatoire dont la distribution conditionnelle sachant le nombre n_+ de signes positifs et n_- de signes négatifs est approximativement normale avec comme paramètres :

$$\mu = \frac{2n_+n_-}{n_+ + n_-} + 1 \quad (4-20)$$

et

$$\sigma^2 = \frac{2n_+n_-(2n_+n_- - (n_+ + n_-))}{(n_+ + n_-)^2(n_+ + n_- - 1)} \quad (4-21)$$

La statistique du test des runs ϑ^{RUN} s'écrit :

$$\vartheta^{RUN} = \frac{\text{Nombre de runs} - \mu}{\sigma^2} \quad (4-22)$$

L'hypothèse H_0 est rejetée si :

$$|\vartheta^{RUN}| > N_{1-\frac{\alpha}{2}}(0,1) \quad (4-23)$$

Avec : $N_{1-\frac{\alpha}{2}}(0,1)$ est le $(1-\alpha/2)$ est le quantile de la distribution normale centrée réduite.

Le tableau ci-dessous nous révèle que le test des runs est validé uniquement par les deux méthodes : BRASS et GLM-Poisson pour la population féminine quel que soit la table de mortalité retenue. Alors que ce test n'est pas validé pour la population masculine et tous sexes confondus par les trois méthodes et quel que soit la table de mortalité.

Tableau 4-6 : Résultats des tests des runs selon la méthode d'ajustement et la table de mortalité retenue.

Test des runs			Nb. Tuns	signes(-)	signes(+)	Hyp. Retenue	P-value
Hommes	Insee	Méthode-SMR	27	51	31	H1	0,003
		Méthode-BRASS	32	47	35	H1	0,0382
		Méthode-GLM-Poisson	27	53	29	H1	0,0052
	MD avant covid	Méthode-SMR	26	53	29	H1	0,0024
		Méthode-BRASS	28	44	38	H1	0,0021
		Méthode-GLM-Poisson	29	55	27	H1	0,033
	HMD yc covid	Méthode-SMR	25	54	28	H1	0,0014
		Méthode-BRASS	30	45	37	H1	0,0092
		Méthode-GLM-Poisson	29	55	27	H1	0,033
Femmes	Insee	Méthode-SMR	27	41	41	H1	9e-04
		Méthode-BRASS	33	53	29	H0	0.1817
		Méthode-GLM-Poisson	37	52	30	H0	0.6233
	MD avant covid	Méthode-SMR	27	45	37	H1	0,001
		Méthode-BRASS	33	47	35	H0	0,065
		Méthode-GLM-Poisson	33	53	29	H0	0,1817
	HMD yc covid	Méthode-SMR	29	45	37	H1	0,0047
		Méthode-BRASS	33	47	35	H0	0,065
		Méthode-GLM-Poisson	31	54	28	H0	0,0888
Tous sexes confondus	MD avant covid	Méthode-SMR	23	51	31	H1	1e-04
		Méthode-BRASS	24	45	37	H1	1e-04
		Méthode-GLM-Poisson	29	54	28	H1	0,028
	HMD yc covid	Méthode-SMR	25	52	30	H1	8e-04
		Méthode-BRASS	24	45	37	H1	1e-04
		Méthode-GLM-Poisson	29	54	28	H1	0,028

▪ **Tests des signes**

Il s'agit d'un test non-paramétrique qui examine la fréquence des changements de signes de la différence entre les probabilités de décès observées et ajustées. Sous l'hypothèse nulle H_0 , la médiane entre les signes positifs et négatifs de cette différence est nulle.

Soit le nombre n_+ de signes positifs et n_- de négatifs, avec $n = n_+ + n_-$, la statistique du test des signes \mathcal{E}^{SIG} s'écrit :

$$\mathcal{E}^{SIG} = \frac{|n_+ - n_-| - 1}{\sqrt{n}} \quad (4-24)$$

Sous l'hypothèse nulle H_0 , la statistique du test des signes suit une loi normale centrée réduite :

$$\mathcal{E}^{SIG} \sim \mathcal{N}(0,1) \quad (4-25)$$

L'hypothèse H_0 est rejeté si : $|\mathcal{E}^{SIG}| > \mathcal{N}_{1-\frac{\alpha}{2}}(0,1)$

Avec $\mathcal{N}_{1-\frac{\alpha}{2}}(0,1)$ est le quantile de la distribution normale centrée réduite.

Le test des signes consiste à compter le nombre de signes positifs et négatifs de la différence entre les taux de mortalité bruts et lissés. Sous l'hypothèse nulle, les valeurs des taux de mortalité bruts sont supérieures aux valeurs des taux de mortalité lissés dans 50% des cas.

Tableau 4-7 : Résultats des tests de signes selon la table de mortalité de référence et la méthode d'ajustement.

Test des signes			signes(+)	signes(-)	Hyp. Retenue	P-value
Hommes	Insee	Méthode-SMR	31	51	H1	0,0359
		Méthode-BRASS	35	47	H0	0,2245
		Méthode-GLM-Poisson	29	53	H1	0,0111
	MD avant covid	Méthode-SMR	29	53	H1	0,0111
		Méthode-BRASS	38	44	H0	0,5808
		Méthode-GLM-Poisson	27	55	H1	0,0029
	HMD yc covid	Méthode-SMR	28	54	H1	0,0058
		Méthode-BRASS	37	45	H0	0,4395
		Méthode-GLM-Poisson	27	55	H1	0,0029
Femmes	Insee	Méthode-SMR	41	41	H0	0,9121
		Méthode-BRASS	29	53	H1	0,0111
		Méthode-GLM-Poisson	30	52	H1	0,0204
	MD avant covid	Méthode-SMR	37	45	H0	0,4395
		Méthode-BRASS	35	47	H0	0,2245
		Méthode-GLM-Poisson	29	53	H1	0,0111
	HMD yc covid	Méthode-SMR	37	45	H0	0,4395
		Méthode-BRASS	35	47	H0	0,2245
		Méthode-GLM-Poisson	2	54	H1	0,0058
Tous sexes confondus	MD avant covid	Méthode-SMR	31	51	H1	0,0359
		Méthode-BRASS	37	45	H0	0,4395
		Méthode-GLM-Poisson	28	54	H1	0,0058
	HMD yc covid	Méthode-SMR	30	52	H1	0,0204
		Méthode-BRASS	37	45	H0	0,4395
		Méthode-GLM-Poisson	28	54	H1	0,0058

Quel que soit le sexe de la population, les résultats du test des signes montrent que le lissage des taux de mortalité, dans le cadre de la méthode BRASS, est correctement effectué et n'engendre pas de sur-lissage ou sous-lissage global par rapport à la population observée. Alors que les méthodes du SMR et GLM Poisson sont rejetées.

4.4.4.Choix du modèle fondé sur les résultats des tests des différentes méthodes

Pour pouvoir trancher sur le choix du modèle d'ajustement pertinent, il convient de synthétiser les différents résultats des tests dans un tableau en affectant des poids à chaque méthode selon ses résultats aux différents tests. A titre d'illustration, on pourrait affecter trois points à la méthode fournissant le meilleur résultat pour un test donné, deux points à la suivante, et un point à la méthode donnant le moins bon résultat.

Tableau 4-8 : Affectation d'un poids à chaque méthode selon ses résultats aux différents tests.

Type de test			Proximité				Régularité et ajustement		Poids total associé	
			SMR	Wilcoxon	Khi-2	R ²	MAPE	Runs		Signes
Hommes	Insee	Méthode-SMR	3	2	2	2	2	1	2	14
		Méthode-BRASS	1	3	1	1	3	3	3	15
		Méthode-GLM-Poisson	2	1	3	3	1	2	1	13
	HMD avant covid	Méthode-SMR	3	2	2	3	2	1	2	15
		Méthode-BRASS	1	3	1	1	3	3	3	15
		Méthode-GLM-Poisson	2	1	3	2	1	2	1	12
	HMD yc covid	Méthode-SMR	3	2	2	3	2	1	2	15
		Méthode-BRASS	1	3	1	2	3	2	3	15
		Méthode-GLM-Poisson	2	1	3	1	1	3	1	12
Femmes	Insee	Méthode-SMR	3	3	1	1	3	1	3	15
		Méthode-BRASS	1	1	2	2	1	2	1	10
		Méthode-GLM-Poisson	2	2	3	3	2	3	2	17
	HMD avant covid	Méthode-SMR	3	2	1	3	3	1	3	16
		Méthode-BRASS	1	3	2	1	2	2	2	13
		Méthode-GLM-Poisson	2	1	3	2	1	3	1	13
	HMD yc covid	Méthode-SMR	2	2	1	3	3	1	3	15
		Méthode-BRASS	1	3	2	1	2	2	2	13
		Méthode-GLM-Poisson	3	1	3	2	1	3	1	14
Tous sexes confondus	HMD avant covid	Méthode-SMR	3	2	2	3	2	2	2	16
		Méthode-BRASS	1	3	1	2	1	1	3	12
		Méthode-GLM-Poisson	2	1	3	1	3	3	1	14
	HMD yc covid	Méthode-SMR	3	2	2	3	2	2	2	16
		Méthode-BRASS	1	3	1	2	3	1	3	14
		Méthode-GLM-Poisson	2	1	3	1	1	3	1	12

On constate que la méthode « SMR » a été privilégié dans le cadre de la table de référence « HMD » dans tous les cas de figure : avant et y compris covid et pour toutes populations : hommes, femmes et tous sexes confondus.

Alors que dans le cadre d'une table de référence « Insee prospectives 2007 », d'une part, c'est le modèle logistique qui a été sélectionné pour la population masculine ; d'autre part, c'est le modèle GLM poisson qui a été sélectionné pour la population féminine.

Donc, j'opte pour la méthode « SMR » pour l'ajustement des taux de mortalité pour notre portefeuille.

4.4.5. Fermeture des tables : Méthode de DENUIT & GODERNIAUX

La méthode de Denuit & Goderniaux (2005) est conçu pour ajuster et prolonger les taux de mortalité dans les âges élevés jusqu'à un âge ultime où la probabilité de survie devient nulle. Cette méthode repose sur la régression quadratique suivante :

$$\log \hat{q}_{x,t} = a_t + b_t x + c_t x^2 + \varepsilon_{x,t} \quad (4-26)$$

Où :

$$\varepsilon_{x,t} \sim N(0, \sigma^2) \quad (4-27)$$

Deux contraintes sont suggérées par les auteurs de la méthode :

$q_{130} = 1$; où 130 est un âge limite choisi de fin de vie humaine.

et $q'_{130} = 0$

Où q'_{130} désigne la dérivée première de q_x au point $x=130$. Il s'agit d'une contrainte d'inflexion.

Ces deux contraintes permettent de garantir la fermeture de la table à 130 ans en imposant une allure concave des taux de mortalité aux âges élevés et l'existence d'une tangente horizontale au point $x=130$ ans. Cette contrainte a pour conséquence d'éviter une décroissance éventuelle des quotients de mortalité q_x aux âges élevés.

Ces deux contraintes entraînent l'égalité suivante :

$$a_t + b_t x + c_t x^2 = c(130 - x^2) \quad (4-28)$$

Le calibrage de la loi selon les trois méthodes choisies a été réalisé sur la tranche d'âge 50-90 ans et l'année 2008. Par conséquent, le prolongement des tables est testé à partir de 91 ans. Les graphiques ci-dessous présentent les fins de tables, en comparant les logarithmes des taux de décès réels, de la table « HMD » et des ceux projetés de la table Insee pour les deux populations masculine et féminine, et des différentes méthodes retenues avec une fermeture de table par l'approche de DENUIT & GODERNIAUX. Seule les résultats de la 1^{ère} méthode « SMR » est présentée ici.

Les graphiques suivants nous montrent que la fermeture des tables respectent correctement la tendance des taux de mortalité observés et la poursuite de la tendance initiale reste cohérente que soit la méthode d'ajustement initiale que ce soient pour les hommes ou pour les femmes.

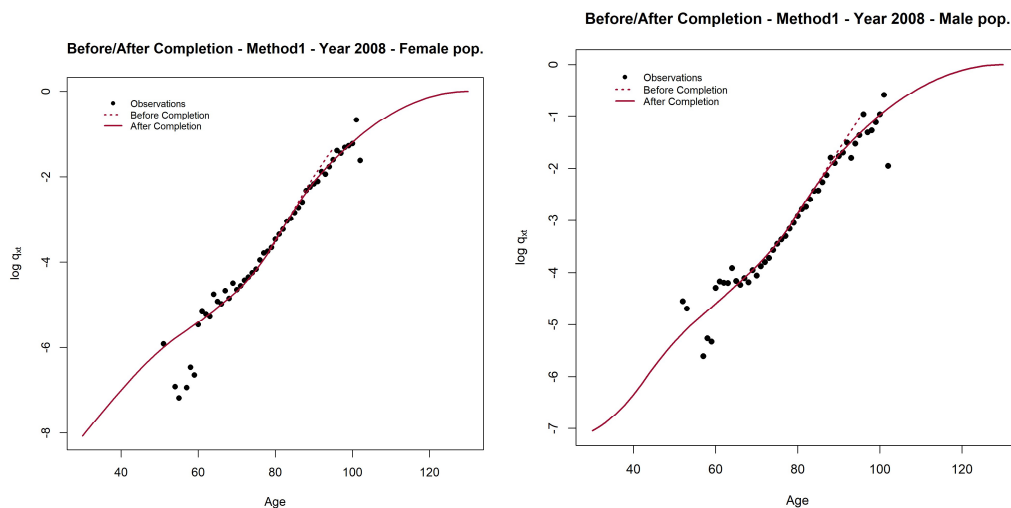


Figure 4-3 : Taux de mortalité (échelle logarithmique) avant et après complétude de table de mortalité.

(a) Insee-population féminine

(b) Insee-population masculine

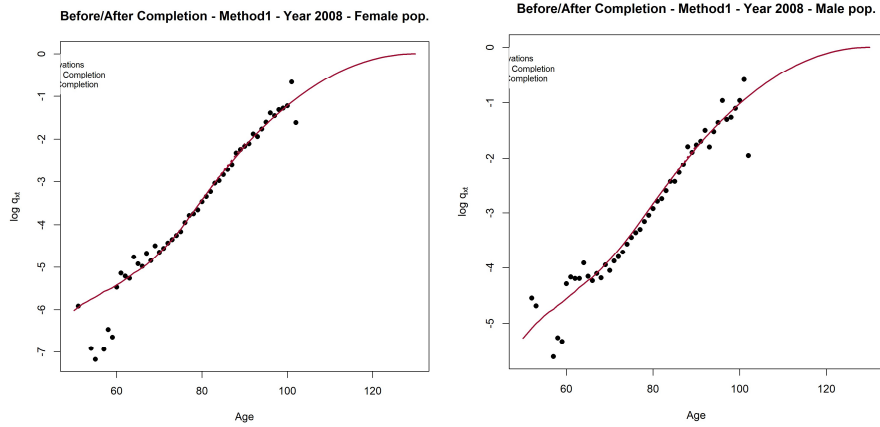


Figure 4-4 : Taux de mortalité (échelle logarithmique) avant et après complétude de table de mortalité.

(c) « HMD » -population féminine

(d) « HMD » -population masculine

4.4.6. Cohérence des tendances d'évolution de la mortalité

- **Analyse de l'espérance de vie résiduelle**

L'espérance de vie résiduelle sur ω années d'un individu âgé x à la date t s'écrit :

$$\omega e_x^{\uparrow}(t) = \sum_{\delta=1}^{\omega} \prod_{j=0}^{\delta-1} (1 - q_{x+j}(t)),$$

(4-29)

Avec ω le nombre d'années sur lesquelles sont calculées l'espérance de vie résiduelle, pour un individu d'âge x , à la date t .

Les taux de survie sont présentés chaque année et pour chaque âge de la manière suivante:

Tableau 4-9 : Présentation des taux de survie par âge et année.

Année \ Âge	2007	2008	„„	2060
50	$1-q_{50}(2007)$	$1-q_{50}(2008)$	„„	$1-q_{50}(2060)$
51	$1-q_{51}(2007)$	$1-q_{51}(2008)$	„„	$1-q_{51}(2060)$
„„	„„	„„	„„	„„
90	$1-q_{90}(2007)$	$1-q_{90}(2008)$	„„	$1-q_{90}(2060)$

Cette méthode de comparaison consiste à multiplier et à additionner les probabilités de survie ($1-q_x(t)$) colonne par colonne, dans le sens de la flèche (schéma ci-dessous).

Ainsi, chaque année t , le résultat obtenu est l'espérance de vie périodique résiduelle à l'année t ($\omega e_x^1(t)$).

L'âge x dans $\omega e_x^1(t)$

correspond à l'âge à partir duquel l'espérance de vie est calculée, ω est l'écart entre x et le dernier âge de la table.

Le dernier âge de la table est fixé à 90 ans dans notre application pour permettre une comparaison avec les résultats sur les taux bruts.

Tableau 4-10 : Illustration du sens de calcul des quotients de survie par âge et année .

Année \ Âge	2007	2008	„	2060
50	↓	↓	↓	↓
51	↓	↓	↓	↓
„	↓	↓	↓	↓
90	↓	↓	↓	↓

Ci-dessous un graphique représentant l'espérance de vie périodique résiduelle pour différentes années avec comme table de référence Insee pour les hommes :

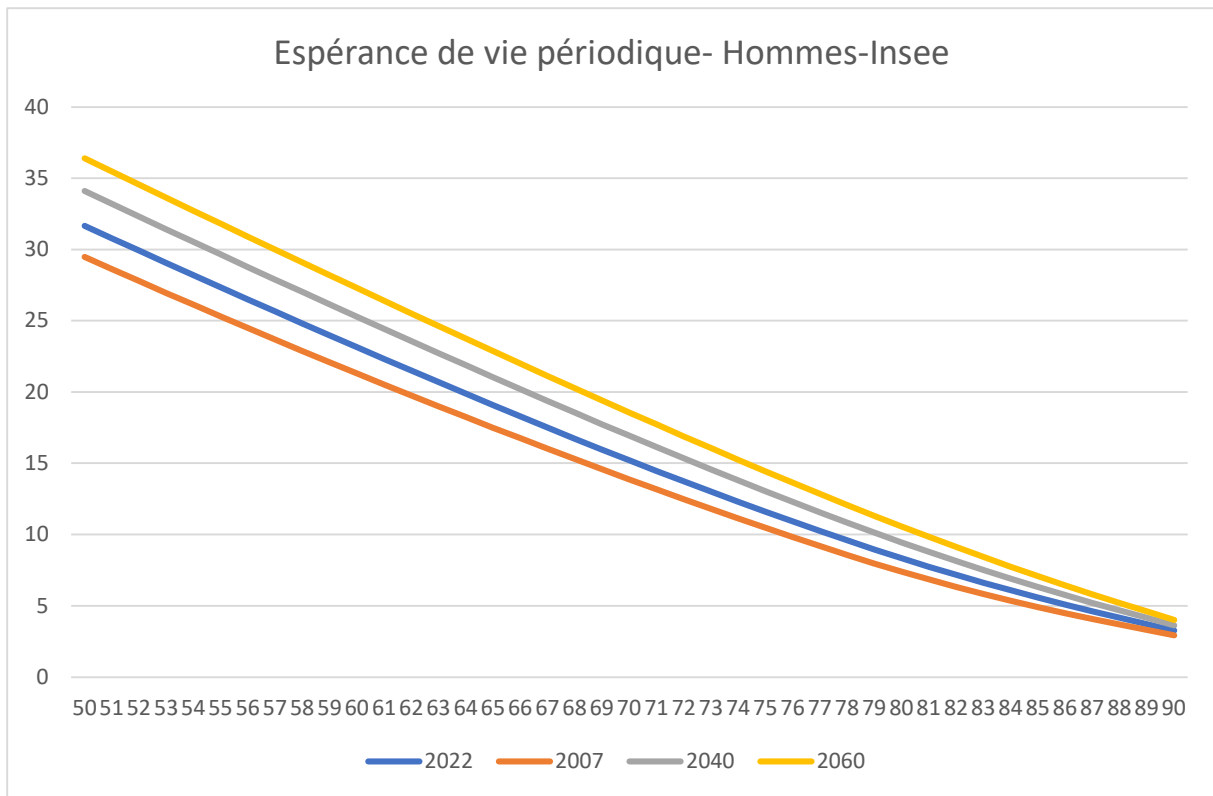


Figure 4-5 : Espérance de vie par âge selon les années pour la population masculine.

▪ **Vision prospective : Cohérence de l'évolution de la mortalité, hors périodes d'observation**

Il s'agit d'une part de comparer les tendances d'évolution des espérances de vie résiduelles totales pour les âges 60, 70, 80 et 90 pour la population masculine, féminine et tous sexes confondus avant et y compris covid respectivement.

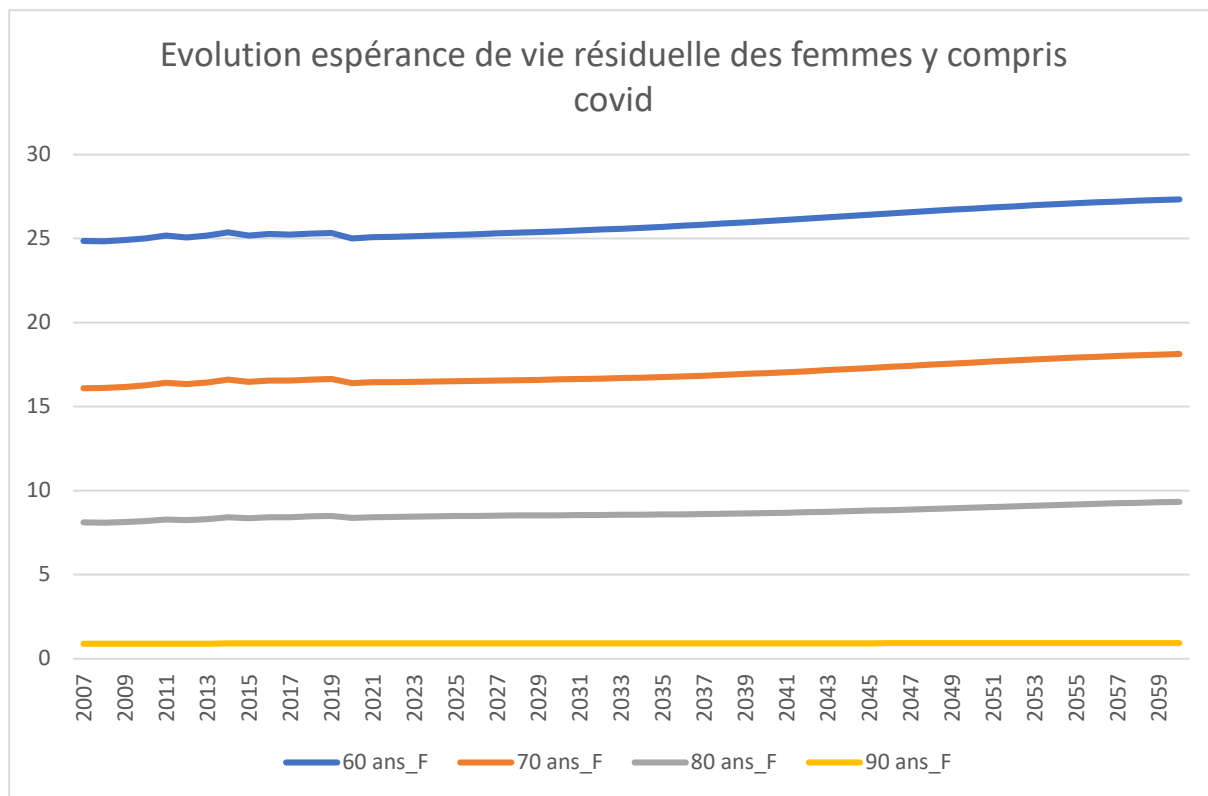


Figure 4-6 : Evolution de l'espérance de vie résiduelle pour certains âges de la population féminine.

On note un rapprochement de l'espérance de vie résiduelle à 60 ans entre les hommes et les femmes dans la table dite « HMD » avant covid19.

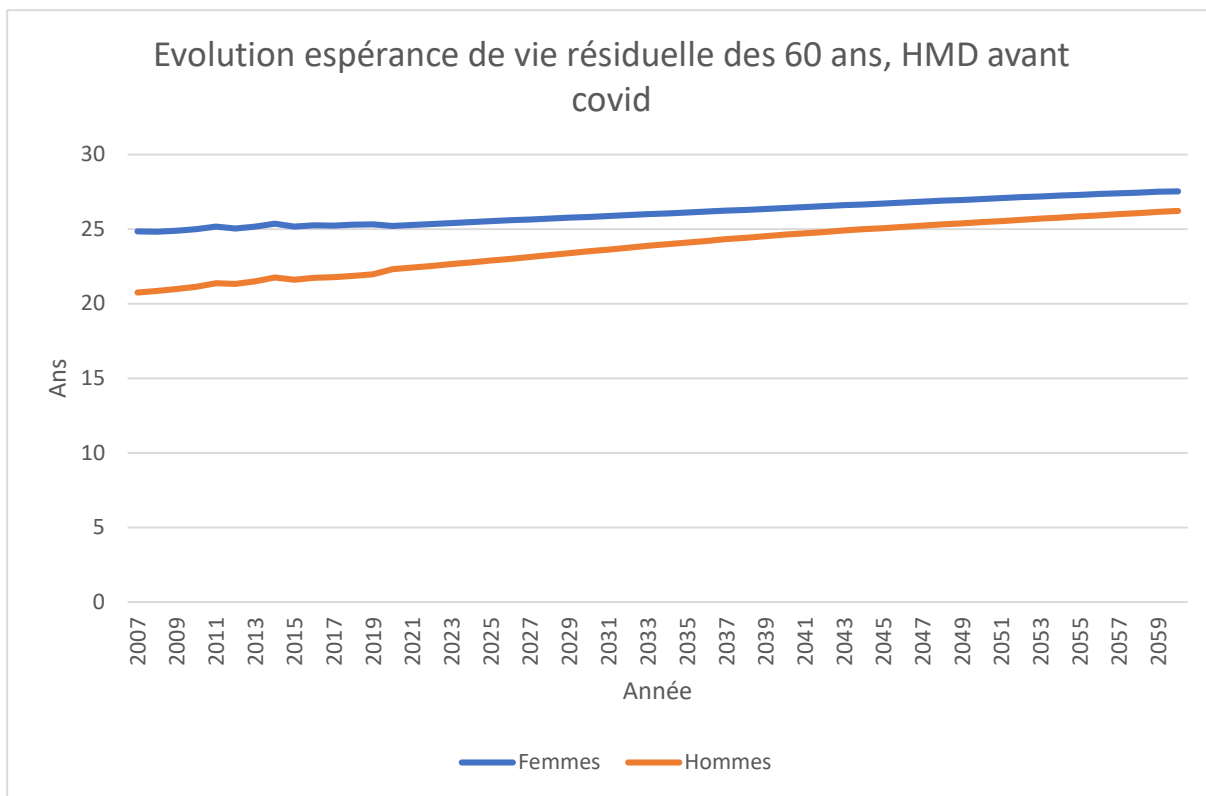


Figure 4-7 : Evolution de l'espérance de vie résiduelle des personnes âgées de 60 ans par sexe.

- Présentation graphique des tables prospectives finales

On présentera graphiquement les tables prospectives finales pour les différentes populations selon les différents cas de figure : « HMD » avant covid/ y compris covid ou Insee.

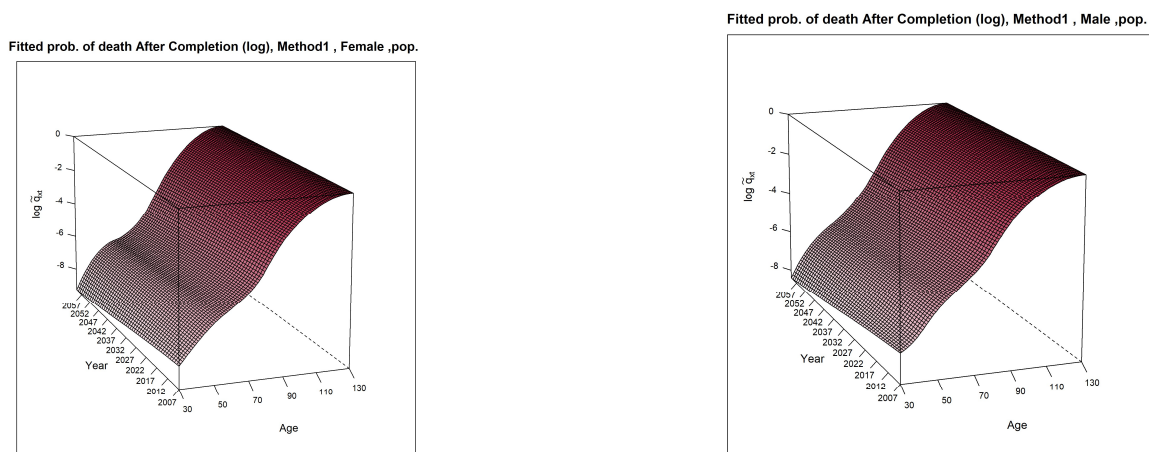
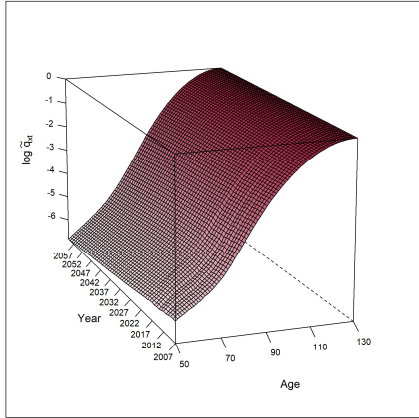


Figure 4-8 : Tables de mortalité prospectives.

(a) Insee-table finale femmes

(b) Insee -table finale hommes

Fitted prob. of death After Completion (log), Method1 , Female ,pop.



Fitted prob. of death After Completion (log), Method1 , Female ,pop.

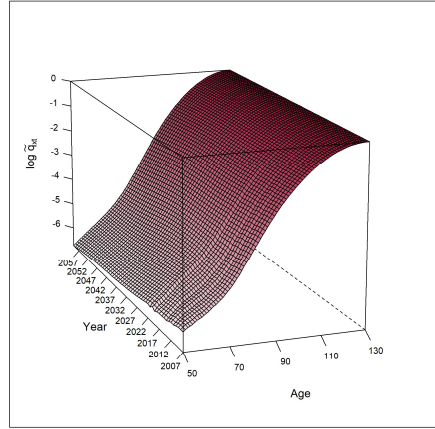
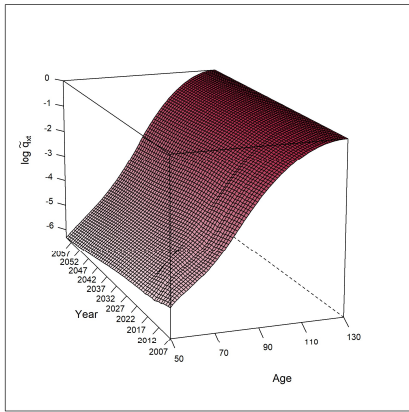


Figure 4-9 : Tables de mortalité prospectives.

(c) « HMD » - table finale femmes avant covid

(d) « HMD » - table finale femmes yc covid

Fitted prob. of death After Completion (log), Method1 , Male ,pop.



Fitted prob. of death After Completion (log), Method1 , Male ,pop.

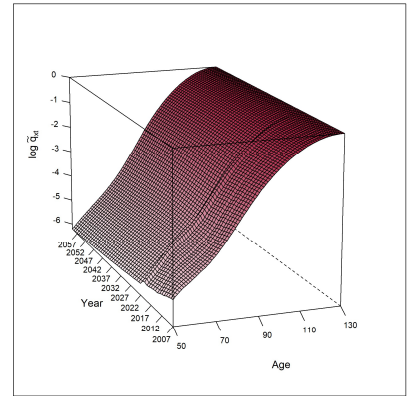
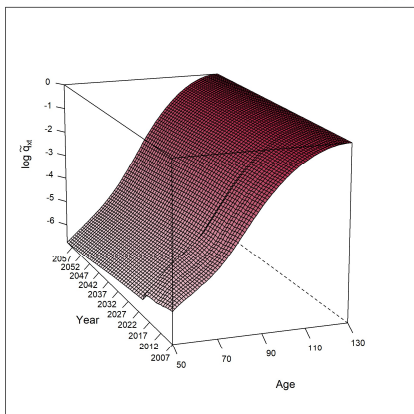


Figure 4-10 : Tables de mortalité prospectives.

(e) « HMD » - table finale hommes avant covid

(f) « HMD » - table finale hommes yc covid

Fitted prob. of death After Completion (log), Method1 , Total ,pop.



Fitted prob. of death After Completion (log), Method1 , Total ,pop.

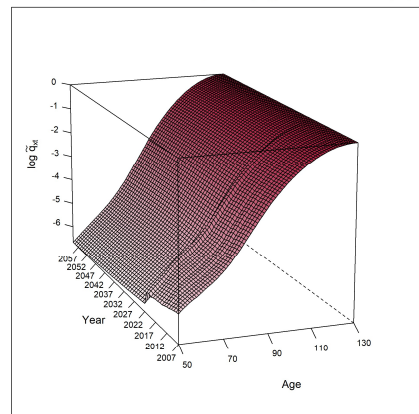


Figure 4-11 : Tables de mortalité prospectives.

(g) « HMD »-tous sexes confondus avant covid

(h) « HMD »-tous sexes confondus yc covid

4.4.7. Conclusion sur le choix du modèle

Les différents éléments de validation présentés ci-dessus avaient pour but de comparer les estimations faites par les trois modèles pour les hommes, femmes, tous sexes confondus selon trois tables de référence : Insee prospectives 2007-2060 et tables « HMD » construites dans le chapitre 2 selon deux périodes d'historique : l'une de 1982-2019 (dite avant covid) et l'autre sur la période 1982-2021 (dite y compris covid) .

La sélection d'un modèle doit obéir à plusieurs critères, notamment :

- La qualité de l'ajustement : les tests réalisés ont montré que les trois modèles apportaient un ajustement correct des données brutes
- La tendance d'évolution issue du portefeuille

Le modèle retenu pour les femmes et pour les hommes est le suivant : il s'agit d'une application d'un coefficient de passage dit « SMR » aux taux de mortalité issus de la table de référence.

Tableau 4-11 : Modèle retenu pour les projections de long terme des hommes, femmes et tous sexes confondus.

Méthode 1 : SMR	INSEE		"HMD" avant covid			"HMD" y compris covid		
	H	F	H	F	Total	H	F	Total
SMR	0,9805	1,0163	1,0138	1,0198	1,0231	1,0136	1,0189	1,0233

Les résultats obtenus confirment la grande proximité de la mortalité des assurés de notre portefeuille de celle de la population générale en France. Ce qui est cohérent avec le fait qu'il s'agit d'une population d'un organisme de protection sociale en France.

Chapitre 5 : Impact des tables de mortalité d'expérience non genrées sur la tarification et le provisionnement des contrats d'épargne retraite d'une compagnie d'assurance vie

Pour établir les tarifs des contrats de rente viagère et comptabiliser les provisions correspondant à ces contrats, les assureurs peuvent, soit utiliser des tables de mortalité réglementaires, soit utiliser des tables d'expérience certifiées, conformément aux dispositions de l'article A.132-18 du code des assurances. Cette approche permet de garantir une tarification et un provisionnement prudents, tenant notamment compte de l'augmentation continue de l'espérance de vie.

Si les primes sont trop basses, la compagnie peut se retrouver rapidement décapitalisée. Si elles sont trop élevées, l'assureur ne sera pas compétitif. D'où l'importance pour les assureurs de déterminer correctement le prix de leurs produits.

Comme évoqué en introduction de ce mémoire, à partir du 24 octobre 2024, et du 1er janvier 2025 pour les contrats renouvelés annuellement par tacite reconduction, les rentes liquidées dans le cadre de contrats d'assurance collectifs portant sur des engagements de retraite supplémentaire devront être calculées sur la base d'une table de mortalité identique pour les hommes et les femmes.

Notre objectif ici est de mesurer d'une part, l'impact d'intégration par les compagnies d'assurance de l'expérience de leurs portefeuilles dans la détermination de leurs tarifs et l'estimation de leurs provisions mathématiques, et d'autre part de comparer les résultats obtenus à l'aide de ce type de table avec ceux obtenus en utilisant les tables réglementaires.

5.1. Impact sur la tarification

De manière synthétique, la tarification se base sur trois éléments à savoir le taux d'intérêt, les frais de gestion et la table de mortalité. Cette dernière devrait permettre à l'entreprise d'assurance d'effectuer une tarification prenant en considération le risque viager. Le principe de tarification en assurance vie se présente par l'égalisation, à l'instant de la souscription t_0 , entre les valeurs actuelles probables des engagements pris par l'assureur et ceux pris par l'assuré. Nous allons comparer la prime unique d'un contrat retraite.

5.1.1 Exemple d'un contrat retraite

Considérons une rente annuelle viagère à terme échu différée, avec un premier versement à l'âge de 65 ans, de 100 €, financée par une prime unique.

Pour calculer le tarif d'une rente viagère annuelle à terme échu de 100 € versée à partir de 65 ans pour un assuré d'âge x au moment de la souscription en 2021, en tenant compte d'une table de mortalité prospective, il est nécessaire de déterminer la valeur actuelle des paiements futurs de la rente en tenant compte de la probabilité de survie de l'assuré, selon la table de mortalité retenue et d'un taux d'intérêt technique.

- Hypothèses retenues ici dans le cadre de ce chiffrage
 - Montant de la rente : 100 € par an.
 - Âge de souscription : x (ici, x prend trois valeurs : 50 ans, 55 , ou 60 ans (en 2024)).
 - Âge de début de la rente : 65 ans (rente différée de 15 ans (10 ou 5 ans selon l'âge de l'assuré au moment de la souscription en 2024)).
 - Table de mortalité prospective : on testera plusieurs tables : tables réglementaires TGH05/TGF05 et les tables d'expérience prospectives construites dans le cadre de ce mémoire.
 - Taux d'intérêt : Taux d'actualisation pour calculer la valeur actuelle des paiements futurs, ici à titre d'illustration, je prends $i=3\%$, compatible avec les taux de l'année 2021.
 - Âge ultime : 130 ans (ou l'âge maximal dans la table de mortalité).
 - Probabilités de survie : Calculées à partir de la table de mortalité à partir de 50 ans, car nous devons d'abord tenir compte de la probabilité de survie de l'assuré jusqu'à l'âge de 65 ans, avant de commencer les paiements.

- Étapes du calcul

- Probabilité de survie de x à 65 ans : Nous devons d'abord calculer la probabilité que l'assuré survive entre l'âge x à la souscription et le début des paiements de rente à 65 ans. Ces probabilités sont extraites de la table de mortalité prospective prise en compte. Si l'assuré décède avant 65 ans, il n'y aura pas de rente à verser.
- Valeur actuelle des paiements de rente à partir de 65 ans : ensuite, nous calculons la valeur actuelle de la rente viagère annuelle versée à partir de 65 ans.
- Actualisation à l'âge de 50 ans : Une fois la valeur actuelle calculée à 65 ans, il faut actualiser cette valeur pour l'amener à l'âge de 50 ans (c'est-à-dire tenir compte de l'attente de 15 ans (65 ans-x) avant le début des paiements). Cela se fait avec la formule suivante :

- ❖ Le prix d'une rente viagère immédiate d'arrérage annuel r qui est aussi la prime unique pure est donné par :

$$\pi = r \times a_x = r \times \frac{N_{x+1}}{D_x} \quad (5-1)$$

Où :

- π : prime pure unique. Ce prix est aussi appelé capital constitutif de la rente.

- x représente l'âge actuel de l'assuré
 - r est l'arrérage,
 - N_{x+1} : Somme des valeurs actuelles des paiements futurs à terme échu à partir de l'âge x .
 - D_x : Valeur actuelle d'une personne âgée de x ans.
 - a_x représente la valeur actuelle des paiements futurs à terme échu pour une personne âgée de x ans.
- Dans le cas d'une rente viagère différée (versement de $r=100\text{€}$ par an) à partir de l'âge de 65 ans, la formule générale pour la valeur actualisée des paiements à partir de 65 ans est :

$$\pi = 100 \times \frac{1}{(1+i)^{65-x}} \times {}_{(65-x)}P_x \sum_{k=1}^{\omega} \frac{kP_{65}}{(1+i)^k} \quad (5-2)$$

Le tableaux ci-dessous illustre la valeur de la prime pure unique en fonction de l'âge de l'assuré à la souscription et la table de mortalité, à taux technique identique. Nous ne tenons pas compte ici des frais de gestion dans notre chiffrage.

Tableau 5-1 : Prime unique d'une rente viagère de 100 € selon la table de mortalité retenue et l'âge de l'assuré à la souscription

Date de souscription en 2024		Age de l'assuré à la souscription		
		50 ans	55 ans	60 ans
Hommes	Insee	817	968	1 157
	HMD avant covid	827	979	1 169
	HMD yc covid	878	1 036	1 234
Femmes	Insee	1 006	1 179	1 386
	HMD avant covid	973	1 140	1 343
	HMD yc covid	958	1 124	1 325
Tous sexes confondus	HMD avant covid	957	1 123	1 324
	HMD yc covid	996	1 166	1 373
Tables réglementaires	TGH05	972	1 135	1 333
	TGF05	1 067	1 247	1 462

Plus l'âge de souscription est jeune, plus la prime à payer est moindre pour un montant de rente identique. Aussi, la prime des hommes est moins élevée que celle des femmes car ils vivent moins longtemps.

De manière générale, à âge de souscription donné, la prime est plus élevée avec la table réglementaire TGF05, qui correspond à la table préconisée par le Code des Assurances pour ce genre de contrat. Cela s'explique par l'espérance de vie plus élevée des assurés des tables réglementaires car elles sont plus prudentes et ne reflètent pas forcément la mortalité réelle des assurés du portefeuille de l'assureur.

La même explication à retenir dans la comparaison de la prime des hommes, femmes et tous sexes confondus selon les tables de mortalité « HMD » avant ou y compris covid 19.

Comme le montre le graphique ci-dessous, les espérances de vie des assurés ayant 50 ans sur les 40 prochaines années sont plus élevées dans les tables réglementaires TGH05 et TGF05 sur la période 2007 et 2021 par rapport aux autres tables. Cela s'explique par la prudence des tables réglementaires par rapport aux tables d'expérience prospectives qui reflètent la mortalité des assurés du portefeuille de l'assureur.

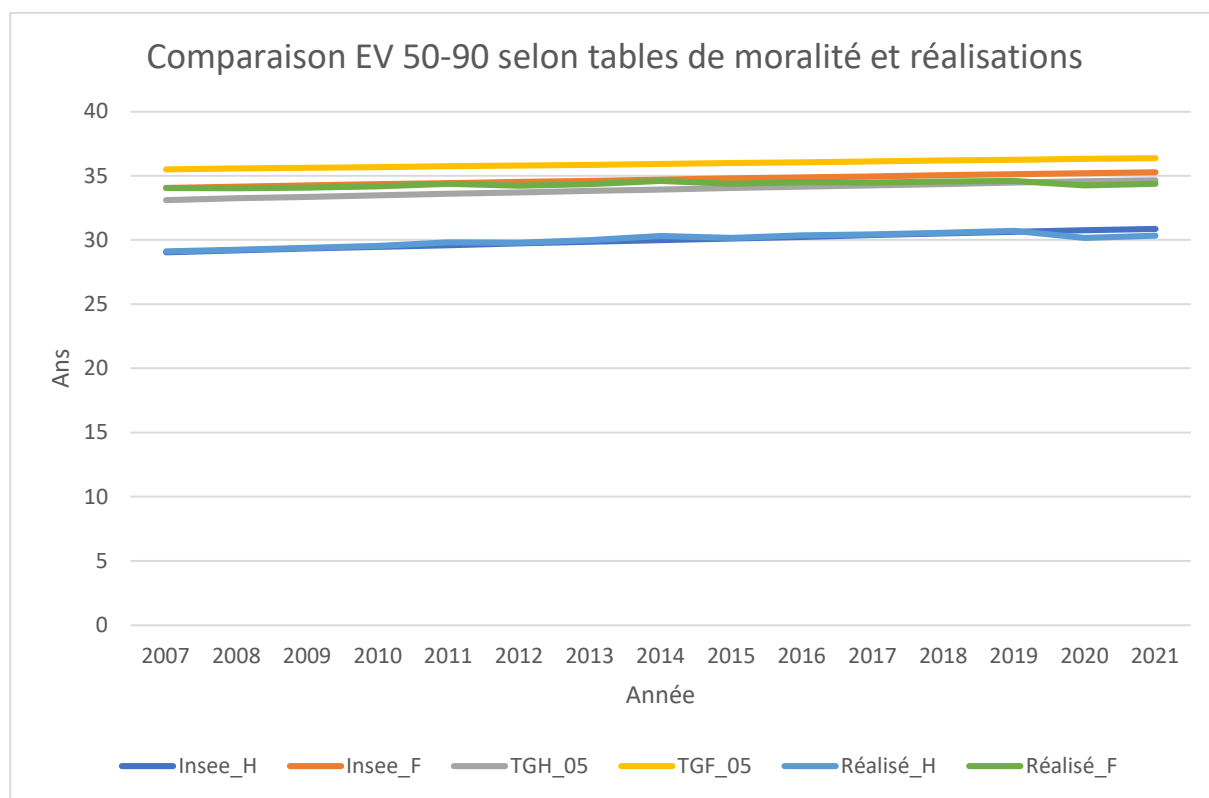


Figure 5-1 : Espérance de vie des 50-90 ans selon la table de mortalité.

Ainsi, si l'assureur utilise des bases techniques qui ne reflètent pas la sinistralité réelle, il se verra très probablement confronté à des pertes techniques et sera, à fortiori, dans l'obligation d'utiliser ses fonds propres pour pouvoir respecter les termes du contrat, le capital constitutif initial devenant insuffisant à financer les arrérages réels à servir.

5.1.2. Comparaison des capitaux constitutifs de rente selon la table de mortalité retenue

Le calcul du capital constitutif d'une rente viagère est essentiel dans la gestion des contrats d'épargne retraite. Il est utilisé pour déterminer le montant qu'un assureur doit avoir en réserve (capital constitutif) afin de pouvoir verser une rente viagère, qui commence par exemple à 65 ans, et ce, jusqu'à la fin de la vie du bénéficiaire, en fonction de tables de mortalité. Ce calcul est lié à la prime unique à la souscription d'un contrat d'épargne retraite.

En effet, la prime unique à la souscription du contrat correspondra au capital constitutif nécessaire pour assurer le paiement de la rente viagère future. Par exemple, si nous choisissons de percevoir une

rente viagère de 1 euro par an à partir de 65 ans, la prime unique devra donc être égale au capital constitutif correspondant à ce montant.

- Calcul du capital constitutif d'une rente viagère

Le capital constitutif d'une rente viagère correspond au montant que l'assureur doit mettre de côté pour assurer le paiement de la rente tout au long de la vie du rentier. Ce montant dépend de plusieurs facteurs :

- L'âge du rentier (65 ans dans ce cas),
- Le montant de la rente annuelle (ici 1 euro),
- La table de mortalité utilisée pour évaluer les probabilités de décès à chaque âge,
- Le taux technique utilisé par l'assureur pour capitaliser le capital constitutif.

Le capital constitutif C , dans le cas d'un arrérage $r= 1\text{€}$, peut être exprimé comme suit :

$$C = a_x = \sum_{k=1}^{\omega} kP_x \times \left(\frac{1}{1+i}\right)^k \quad (5-3)$$

où ${}_kP_x$ est la probabilité pour cet individu d'âge x d'être encore en vie à l'âge $x+k$. On suppose i est le taux technique des rentes, c'est-à-dire le taux d'escompte.

Le capital constitutif d'une rente viagère à terme échu immédiate correspond à la valeur actuelle probable des flux de rentes de 1€ versé tant que l'assuré d'âge x est vivant.

où ${}_kP_x$ désigne la probabilité pour un individu d'âge x d'être en vie dans k années. On suppose que le taux d'actualisation est constant et égal à 3%.

Les tables de mortalité permettent de calculer les probabilités de survie à chaque âge à partir de 65 ans. Elles sont cruciales car elles influencent la durée de paiement prévue de la rente. Plus l'espérance de vie est élevée, plus le capital constitutif doit être important pour garantir le versement de la rente à vie.

Dans cet exemple, on suppose qu'on versera 1€ annuellement à un retraité à partir de l'âge de 65 ans (on considère cet âge comme celui d'entrée à la retraite). Ce versement se poursuit tant que la personne est en vie. L'objectif est de comparer le coût de ce versement avec les différentes tables de mortalité d'expérience prospective : TGH05/ TGF05 ; « Insee_H » et « Insee-F » pour les hommes et les femmes et les tables dites « HMD » pour les hommes, femmes et tous sexes confondus avant et y compris covid19.

Les capitaux constitutifs moyens à constituer, pour le versement d'un 1 € par pour un retraité à partir de 65 ans en 2025 figurent dans le tableau suivant :

Tableau 5-2 : Comparaison des capitaux constitutifs obtenus les différentes tables de mortalité

Table de mortalité		Coût constitutif d'une rente d'un 1€	Ecart du coût par rapport aux tables réglementaires
Hommes	Insee	13,99	-11,35%
	HMD avant covid	14,22	-9,84%
	HMD yc covid	14,89	-5,59%
Femmes	Insee	16,42	-4,77%
	HMD avant covid	15,97	-7,40%
	HMD yc covid	15,78	-8,49%
Tous sexes confondus	HMD avant covid	15,79	-8,40%
	HMD yc covid	16,29	-5,54%
Tables réglementaires	TGH05	15,78	
	TGF05	17,24	

L'écart du coût, pour tous sexes confondus, et pour les femmes est calculé par rapport à la table TGF05. Pour les hommes, par rapport à la table TGH05.

Ce tableau confirme le fait qu'en moyenne, les tables d'expérience prospective construites dans le cadre de ce mémoire coûtent moins cher en terme de capital constitutif pour verser une pension d'un euro à partir de 65 ans à un assuré par rapport aux tables réglementaires actuelles dans tous les cas de figure.

5.2. Impact sur les provisionnements

Les provisions mathématiques constituent les plus importantes provisions techniques des organismes d'assurance-vie. Leur mode de calcul est encadré par les prescriptions comptables de l'Autorité des Normes Comptables (ANC).

Dans le chapitre précédent, nous avons construit différentes tables de mortalité d'expérience prospective. Le but ici est d'étudier l'impact en termes de provisionnement des contrats d'épargne dans le cas d'un changement de table de mortalité.

5.2.1. Contrat d'épargne et provisionnement

L'Article R343-3 du Code des Assurances stipule que les provisions techniques correspondant aux opérations d'assurance sur la vie, d'assurance nuptialité-natalité et de capitalisation, sont les suivantes :

« 1° Provision mathématique : différence entre les valeurs actuelles des engagements respectivement pris par l'assureur et par les assurés. Pour des contrats faisant intervenir une table de survie ou de mortalité, les montants des provisions mathématiques doivent inclure une estimation des frais futurs de gestion qui seront supportés par l'assureur pendant la période de couverture au-delà de la durée de paiement des primes ou de la date du prélèvement du capital constitutif ; l'estimation de ces frais est égale au montant des chargements de gestion prévus dans les conditions tarifaires de la prime ou du capital constitutif et destinés à couvrir les frais de gestion ;

2° Provision pour participation aux bénéfices : montant des participations aux bénéfices attribuées aux bénéficiaires de contrats lorsque ces bénéfices ne sont pas payables immédiatement après la liquidation de l'exercice qui les a produits ;

3° Réserve de capitalisation : réserve destinée à parer à la dépréciation des valeurs comprises dans l'actif de l'entreprise et à la diminution de leur revenu ;

4° Provision de gestion : provision destinée à couvrir les charges de gestion future des contrats non couvertes par ailleurs ;

5° Provision pour aléas financiers : provision destinée à compenser la baisse de rendement de l'actif ;

6° Provision pour risque d'exigibilité : provision destinée à faire face aux engagements dans le cas de moins-value de l'ensemble des actifs mentionnés à l'article [R. 343-10](#). La provision à constituer est évaluée dans les conditions définies à l'article [R. 343-5](#) ;

7° Provision pour frais d'acquisition reportés : provision destinée à couvrir les charges résultant du report des frais d'acquisition constaté ;

8° Provision pour égalisation : provision destinée à faire face aux fluctuations de sinistralité afférentes aux opérations d'assurance de groupe contre le risque décès ;

9° Provision de diversification : pour les engagements relevant de l'article [L. 134-1](#), provision destinée à absorber les fluctuations des actifs affectés à ces engagements et sur laquelle les souscripteurs ou adhérents détiennent des droits individualisés sous forme de parts.

10° Provision collective de diversification différée : pour les engagements relevant de l'article [L. 134-1](#), provision destinée au lissage de la valeur de rachat des contrats.

11° Provision pour garantie à terme : pour les engagements relevant du 2° de l'article [L. 134-1](#), provision destinée à faire face à une insuffisance d'actifs au regard des garanties à échéance contractées. »

Un engagement ne peut être provisionné qu'au titre d'une seule des catégories mentionnées au présent article.

5.2.2. La provision mathématique diffère donc selon la phase du contrat de rente viagère considéré

L'alinéa 2 de l'article R331-3 du Code des assurances dispose que la provision mathématique est « la différence entre les valeurs actuelles des engagements respectivement pris par l'assureur et par l'assuré ».

L'engagement de l'assuré consiste à payer les primes, uniques ou périodiques (y compris les chargements d'acquisition et de gestion compris dans les primes futures. En revanche, l'engagement

de l'assureur inclut le paiement du capital promis ou le service de la rente acquise à la date de survenance de l'aléa viager (la survie ou le décès) prévu au contrat.

La provision mathématique est ainsi constituée pour que l'assureur puisse honorer son engagement à tout moment (en cas de rachat du contrat par exemple).

On distingue deux cas de figure : la phase de constitution de l'épargne et celle de la restitution des rentes. La première concerne les actifs et la seconde les retraités. Pour ces derniers, l'engagement de l'assureur est alors donné par la valeur actuelle probable des rentes restantes au moment de l'évaluation. Le coût pour l'assureur provient de l'écart entre la valeur réelle des rentes et leur valeur estimée dans la phase de constitution de l'épargne.

$$E_t = D_t \times P_t \times A_t \quad (5-4)$$

E_t est l'engagement évalué à la date t au bénéfice de l'actif,

D_t est la valeur estimée des droits à la date d'exigibilité,

P_t est la probabilité de survie de l'assuré ; les droits ne seront servis que si l'assuré est vivant,

A_t est l'actualisation des flux. Il est à noter que l'actualisation des flux sur la période de constitution du capital se fait au taux d'actualisation (droits escomptés pour avoir leur valeur actuelle), alors que l'actualisation des flux sur la période de service des rentes se fait au taux technique des rentes.

Dans le cas d'un contrat de rentes viagères, la provision mathématique diffère donc selon la phase du contrat considéré :

- Phase de constitution de l'épargne :

Durant cette phase de constitution du capital (CC), l'assureur s'engage uniquement à conserver l'ensemble des primes versées par l'assuré en vue de la conversion (on suppose qu'il n'existe pas de revalorisation des provisions). De son côté, aucun engagement n'incombe à l'assuré.

Dès lors, la provision mathématique $PM(t)$ à un instant t de la phase de constitution du de l'épargne est exactement égale à l'ensemble du capital épargné par l'assuré :

$$PM(t) = CC(t) \quad (5-5)$$

- Phase de restitution de l'épargne :

Si l'assuré a survécu jusqu'à la date de sortie en rentes, la conversion du capital épargné en rentes survient et amorce la phase de restitution. Dès lors, la situation du contrat est figée : l'assuré perçoit de l'assureur une rente fixe, calculé au moment de la conversion, jusqu'à ce qu'il décède. Aucun apport de capital émanant de l'assuré ne peut survenir.

Donc, la provision mathématique $PM(t)$ à un instant t de la phase de rentes est égale à la valeur actuelle probable de l'assureur :

$$PM(t) = VA \text{ Assureur}(t) \quad (5-6)$$

Le niveau de ces rentes est déterminé par le capital constitué pendant la phase d'épargne mais surtout par la table de mortalité en vigueur au moment de la conversion de ce capital en rentes. Cependant, un changement de table de mortalité peut survenir entre la souscription du contrat et la conversion du capital. Dans ce cas, à capital épargné et taux technique équivalents, le montant des rentes prévu à la souscription devrait différer du montant des rentes perçues. Pour se prémunir de ce risque de changement de table de mortalité et donc d'une baisse du montant des arrérages, il est possible pour l'assuré de souscrire initialement une garantie de table de mortalité. Cette option lui permettra de bénéficier de l'application par l'assureur de la table de mortalité en vigueur au moment de la souscription lors de la conversion du capital épargné en rentes.

Mais, cette garantie fait notamment apparaître pour l'assureur une provision supplémentaire que l'on appelle "provision pour écart capital constitutif - provision mathématique" (provision pour écart CC-PM).

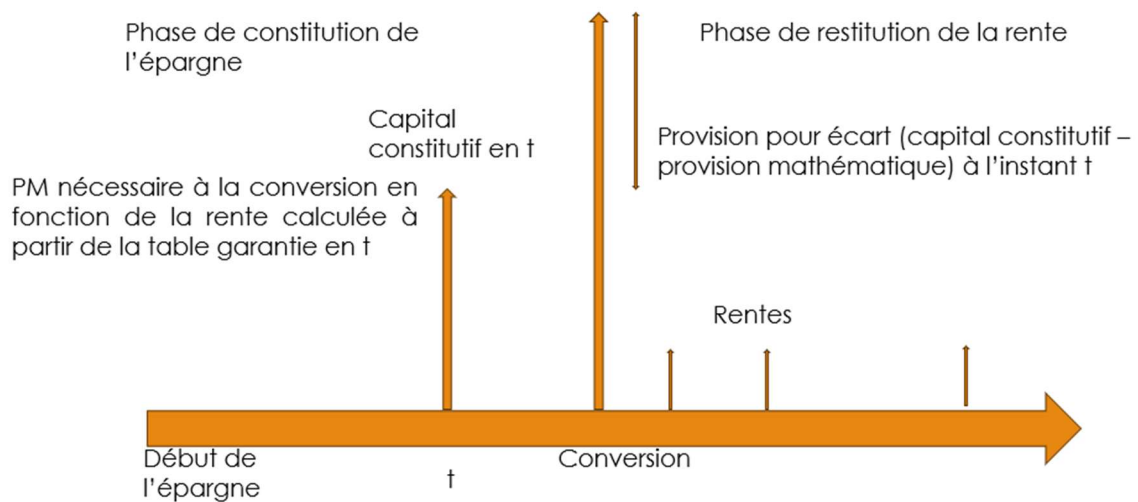


Figure 5-2 : Schéma de déroulement de vie d'un contrat de rente viagère.

5.2.3. Illustration de l'impact d'un changement de table de mortalité sur le montant de provisionnement à travers des cas types

Le principe d'évaluation des engagements consiste en l'estimation à la date t des flux des rentes.

Dans cette section, nous allons étudier l'impact dans le calcul des provisions dites mathématiques, c'est-à-dire les dettes qui estiment au bilan de l'assureur ce que ce dernier devra payer en moyenne aux assurés, sous certaines hypothèses, en général celles du tarif. Il est à noter qu'il n'y a provision que s'il existe un engagement.

On se propose un cas d'étude de trois femmes et trois hommes ayant respectivement 50, 55 et 60 ans en 2025. Au moment de la souscription de leurs contrats d'épargne au cours de la décennie 2010, les tables de mortalité TGH05 et TGF05 étaient en vigueur. Les caractéristiques techniques du contrat sont les suivantes :

- taux technique garanti : $i = 3\%$;
- frais d'arrérage : $f = 0\%$;
- sortie en rentes viagères annuelles différées ;
- âge de sortie prévu : 65 ans ;
- table de mortalité au moment de la souscription : table TGF05 pour les femmes et TGH05 pour les hommes ;
- souscription d'une garantie de table de mortalité.

Supposons qu'en $t = 01/01/2024$, le capital épargné pour chaque assuré parmi nos cas types, noté K , s'élève à 10 000 € par assuré et qu'aucun versement complémentaire ne viendra gonfler ce capital avant l'échéance prévue.

Le contrat étant en phase d'épargne, il y a deux provisions à calculer : la provision mathématique et la provision pour écart CC-PM (capital constitutif – provisions mathématiques).

- Le calcul de la provision mathématique correspond exactement à ce qui a été épargné à l'instant où la provision est calculée donc $PM(31/12/2024) = K = 6 \cdot 10000 = 60\,000$ €.
- Calcul de la provision pour écart : CC-PM

Pour obtenir le montant de cette provision, il faut d'abord déterminer l'arrérage issu du capital constitué et de la table de mortalité garantie pour chaque assuré, puis déterminer la provision mathématique nécessaire pour assurer le même montant de rentes en fonction de la table en vigueur et enfin mesurer l'écart : (capital constitutif - provision mathématique).

Le montant de l'arrérage r provenant des caractéristiques du contrat de l'assuré se calcule au moyen de l'égalité : $r = K/a_x$

Le montant de la rente dépend du facteur de rente viagère a_x , qui représente la valeur actuelle des paiements futurs à terme échu pour une personne âgée de x ans.

$$a_x = \frac{N_{x+1}}{D_x} \quad (5-7)$$

Où :

- N_{x+1} : Somme des valeurs actuelles des paiements futurs à terme échu à partir de l'âge x .
- D_x : Valeur actuelle d'une personne âgée de x ans.
- x représente l'âge actuel de l'assuré (50, 55, ou 60 ans dans ce cas).

- **Calcul d'une provision pour écart CC-PM**

- **Calcul de la provision mathématique**

La provision mathématique correspond exactement à ce qui a été épargné à l'instant où la provision est calculée donc : $PMc(t) = K = 10\ 000 * 6 = 60\ 000\ €$

- **Calcul de la provision pour écart CC-PM**

Pour obtenir le montant de cette provision, il faut d'abord déterminer l'arrérage issu du capital constitué et de la table de mortalité garantie, puis déterminer la provision mathématique nécessaire pour assurer le même montant de rentes en fonction de la table en vigueur et enfin mesurer l'écart entre le capital constitutif et la provision mathématique.

Le montant de l'arrérage r provenant des caractéristiques du contrat de l'assuré se calcule au moyen de l'égalité :

$$r = K/a_x \quad (5-8)$$

Tableau 5-3 : Calcul des a_x en fonction de l'âge selon les tables réglementaires actuelles

Age à la souscription en 2020		ax		
		50	55	60
Tables réglementaires	TGH05	9,77	10,71	13,38
	TGF05	10,71	11,82	14,67

Tableau 5-4 : Calcul du montant de l'arrérage r correspondant à un capital constitutif de 10000 € par assuré

Age à la souscription en 2020		Montant de l'arrérage r		
		55	55	60
Tables réglementaires	TGH05	1 023	934	747
	TGF05	933	846	681

L'application de tables de mortalité genrées (TGH05 pour les hommes et TGF05 pour les femmes) permet d'adapter le calcul de la rente viagère selon le sexe de l'assurée. Les femmes ayant en moyenne une espérance de vie supérieure à celle des hommes, l'assureur doit donc verser la rente plus longtemps à une femme qu'à un homme. Ceci implique que les rentes viagères versées aux femmes sont mécaniquement inférieures à celles versées aux hommes comme le montre le tableau ci-dessus.

Concrètement : avec une table de mortalité « homme », un homme de 55 ans qui partirait à la retraite en 2025 avec un capital constitutif de 10 000 euros toucherait une pension de 1023 euros par an. Avec

la table de mortalité « femmes », une femme de même âge et avec un capital constitutif de même niveau touchera une pension de 933 euros par mois, soit un écart de 9%.

Pour déterminer la provision mathématique $PMr(t,x)$ qui permettrait d'assurer ce niveau de rentes au moment de la conversion en utilisant la table de mortalité retenue selon le cas : hommes, femmes, tous sexes confondus avant et y compris covid 19, on procède comme suit :

$PMr(t,x)$ est donné par la relation suivante :

$$PMr(t,x) = r_x \times a_x^{tabmort} \quad (5-9)$$

Les tableaux suivants permettent de refléter la différence entre les provisions calculées sur la base de la réglementation actuelle (table TGH05 pour les hommes et TGF05 pour les femmes) et les différentes tables prospectives construites dans le cadre de ce mémoire. Aussi, une hypothèse de taux technique égal à 3% a été appliqué.

Tableau 5-5 : Ecart de provisionnement par rapport aux tables réglementaires, en %.

Age à la souscription en 2020		50 ans			55 ans			60 ans		
		Montant des provisions	ax	Ecart du montant des provisions par rapport aux tables réglementaires	Montant des provisions	ax	Ecart du montant des provisions par rapport aux tables réglementaires	Montant des provisions	ax	Ecart du montant des provisions par rapport aux tables réglementaires
Hommes	Insee	8 360	8,17	-16%	8 422	9,02	-16%	8 643	11,57	-14%
	HMD avant covid	8 527	8,33	-15%	8 586	9,19	-14%	8 791	11,77	-12%
	HMD yc covid	8 995	8,79	-10%	9 064	9,70	-9%	9 229	12,35	-8%
	TGF05	10 964	10,71	10%	11 038	11,82	10%	10 963	14,67	10%
Femmes	Insee	9 426	10,10	-6%	9 424	11,14	-6%	9 478	13,91	-5%
	HMD avant covid	9 113	9,76	-9%	9 095	10,75	-9%	9 179	13,47	-8%
	HMD yc covid	8 970	9,61	-10%	8 949	10,58	-11%	9 049	13,28	-10%
Tous sexes confondus	HMD avant covid	8 969	9,61	-10%	8 955	10,58	-10%	9 057	13,29	-9%
	HMD yc covid	9 305	9,97	-7%	9 292	10,98	-7%	9 362	13,74	-6%
Tables réglementaires	TGH05	10 000			10 000			10 000		
	TGF05	10 000			10 000			10 000		

Avec l'interdiction d'utiliser les tables de mortalité différenciées par sexe et dans l'attente d'éventuelles tables de mortalité réglementaires, l'une des options consiste à appliquer la table de mortalité des femmes (dite « TGF05 ») pour tous par les assureurs. Or, l'utilisation de cette table pour les hommes implique une augmentation de provisionnement de 10% pour l'assureur quel que soit l'âge à la souscription (dans notre étude de cas ici : 50, 55 ou 0 ans).

Comme expliqué précédemment, on dispose ici de trois ensembles de tables d'expérience (les tables "Insee" et "HMD" avant et y compris covid 9). Les tables dites « HMD » ne diffèrent que par l'historique de données choisi. Il est logique que tout changement de tables impacte le provisionnement : des contrats en phase d'épargne pour lesquels une garantie de table existe ; de l'ensemble des contrats en phase de restitution. L'utilisation d'une nouvelle table d'expérience va engendrer pour l'assureur un montant de provisions propre. Il y a un écart de 2 à 5% entre les provisionnements basés sur les tables « HMD » avant et y compris covid 19 quel que soit le sexe de l'assuré et l'âge de l'assuré à la souscription. Nous remarquerons toutefois qu'en moyenne, ces tables d'expérience dites « HMD »

conduisent à provisionnement inférieur à celui avec les tables réglementaires en vigueur actuellement TGH05/TGF05.

En effet, nous constatons que l'utilisation des tables de mortalité d'expérience prospectives conduirait à un besoin de provisionnement moindre allant de 6% à 16% selon la table de mortalité retenue par rapport aux tables réglementaires.

Donc, les assureurs ont tout intérêt de construire une table d'expérience prospective qui reflète la mortalité de leur portefeuille que d'appliquer la table TGF05 dans le cadre de la suppression des tables de mortalité générées des contrats d'épargne retraite collectifs.

5.3. Conséquences financières pour l'assureur

5.3.1. Risque d'insuffisance des provisions

Si l'assureur a un portefeuille très genré (fortement masculin ou féminin) et qu'il passe à une table unisexe, il existe un risque d'insuffisance des provisions :

- Dans un portefeuille principalement masculin, l'assureur devra provisionner davantage avec la nouvelle table unisexe, ce qui peut entraîner un besoin accru en capital.
- Dans un portefeuille principalement féminin, les provisions seront réduites, mais cela pourrait entraîner un risque de sous-provisionnement si les femmes vivent plus longtemps que prévu par la nouvelle table.

5.3.2. Réévaluation des engagements à long terme

L'assureur doit réévaluer ses engagements à long terme avec la nouvelle table de mortalité unisexe. Cela peut entraîner :

- Des ajustements des réserves techniques (provisions).
- Un réajustement du mix de portefeuille pour éviter les déséquilibres entre les sexes.
- Un renforcement des tests de stress pour s'assurer que l'assureur peut faire face à l'incertitude liée à l'utilisation d'une table unisexe.

Cela entraînera des ajustements administratifs et techniques.

Il est à noter que toute variation de provision doit être constituée immédiatement après le changement de table en vertu du principe selon lequel l'assureur peut répondre à tout moment à ses engagements. Toutefois, à l'échelle d'un portefeuille, cette variation représente en pratique un montant conséquent. Le législateur (article A331-1-2 du Code des assurances) autorise donc la répartition de ce provisionnement supplémentaire sur quinze années.

5.4. Synthèse des impacts de suppression d'une table de mortalité genrée

La suppression d'une table de mortalité genrée dans le cadre des contrats épargne retraite collectifs, au profit d'une table de mortalité unisexe (non différenciée par sexe), a des impacts importants sur les provisions mathématiques et la gestion des risques pour l'assureur. Voici une analyse des principaux effets de ce changement.

Historiquement, les tables de mortalité utilisées pour les contrats de retraite et d'assurance vie étaient souvent genrées, c'est-à-dire différenciées selon le sexe, car les femmes ont généralement une espérance de vie plus longue, ce qui entraîne des paiements de rentes plus longs. Alors que les hommes ont une espérance de vie plus courte, ce qui signifie que les paiements de rentes sont effectués sur une durée plus courte.

5.4.1. Effet sur les provisions mathématiques

▪ **Impact pour les femmes**

- Dans un système genré, les femmes, avec une espérance de vie plus longue, ont des provisions mathématiques plus élevées. Cela est dû au fait qu'elles perçoivent la rente sur une période plus longue.
- Avec la suppression des tables genrées, la table unisexe réduit l'espérance de vie moyenne utilisée pour les femmes. Cela entraîne une réduction des provisions mathématiques pour les femmes, car les engagements futurs (rentes viagères) sont désormais calculés sur une période plus courte, en moyenne.

▪ **Impact pour les hommes**

- Dans un système genré, les provisions mathématiques pour les hommes sont généralement plus faibles, car leur espérance de vie est plus courte.
- Avec une table unisexe, l'espérance de vie utilisée pour les hommes augmente, car elle reflète maintenant la durée de vie moyenne combinée des deux sexes. Cela conduit à une augmentation des provisions mathématiques pour les hommes, car les rentes devront être versées sur une période plus longue que celle prévue par les anciennes tables spécifiques aux hommes.

▪ **Effet global sur les provisions**

L'impact global dépend de la structure démographique du portefeuille d'assurés :

- Si le portefeuille comporte plus de femmes (par exemple, dans des secteurs où la proportion de femmes est élevée), la suppression des tables genrées pourrait entraîner une réduction des

provisions globales, car les engagements futurs diminuent pour les femmes (bien qu'ils augmentent pour les hommes).

- Si le portefeuille contient plus d'hommes, les provisions globales pourraient augmenter, car les engagements futurs seront plus élevés pour les hommes.
- Dans un portefeuille équilibré entre hommes et femmes, l'impact global pourrait être neutre en termes de provisions, car les augmentations pour les hommes compenseraient les réductions pour les femmes.

5.4.2. Effet sur la tarification : incertitude accrue

Avec une table de mortalité unisexe, l'assureur doit gérer un risque plus élevé d'incertitude, car les différences de mortalité entre les hommes et les femmes ne sont plus prises en compte directement :

- Si un portefeuille contient principalement des hommes ou des femmes, l'assureur risque de sous-provisionner ou de sur-provisionner en fonction de la structure démographique réelle.
- Le mélange des deux sexes pourrait entraîner des distorsions dans la tarification des contrats futurs si la composition du portefeuille n'est pas équilibrée entre hommes et femmes.
- Dans le cas de nouveaux contrats, les assureurs devront revoir leur tarification pour tenir compte du fait que les différences de mortalité entre hommes et femmes ne sont plus explicitement utilisées. Cela peut entraîner des ajustements des primes pour les nouveaux souscripteurs.
- Les anciens contrats pourraient également être affectés si l'assureur décide de revoir les provisions mathématiques pour tenir compte du passage à une table unisexe.

Il convient de noter que la suppression de tables de mortalité genrées en assurance pour les contrats épargne retraite collectifs se traduirait par une mutualisation de la situation inégale entre ces les deux sous-populations masculine et féminine en termes d'espérance de vie. La tarification fondée sur l'espérance de vie des femmes statistiquement plus élevée, se traduisait par un coût d'acquisition des rentes plus onéreux que pour les hommes. Il n'en demeure pas moins que les tables de mortalité réglementaires TGH05 et TGF05 sont prudentes. En outre, le Code des Assurances impose que les excédents techniques qui en découlent soient reversés aux assurés, dans des conditions précisées par arrêté (article A.132-11 du code des assurances), de sorte que les assurés ne soient pas pénalisés par des tarifs excessivement conservateurs. Dans le cas de l'utilisation de la table TGF05, cela conduirait à ce que les sociétés d'assurance dégagent un bénéfice technique. Les assureurs ont néanmoins l'obligation de restituer dans les 8 ans une part importante de ce bénéfice technique aux assurés, dans le cadre d'une répartition que l'assureur peut orienter en fonction de sa politique commerciale, par exemple sous la forme d'une revalorisation des rentes.

Conclusion

Loi n°2023-973 du 23 octobre 2023 relative à l'industrie verte, publiée au Journal Officiel le 24 octobre 2023 a acté la suppression des tables de mortalité générées aux contrats d'épargne retraite collectifs.

▪ Démarche poursuivie

Dans le cadre de ce travail, des tables de mortalité prospectives de la population française ayant entre 50 et 90 ans ont été construites. L'établissement de ces tables a été fondé sur les données de la base HMD. Pour tenir compte des effets de la Covid 19 à court et long terme, deux types de projections réalisés qui ne diffèrent que par l'étendu de l'historique de données : l'un « dit avant covid » basé sur l'historique 1982-2019 et l'autre dit « y compris covid » basé sur l'historique 1982-2021.

Nous avons choisi les données de mortalité pour la population française publiées par la HMD pour des raisons de cohérence et de disponibilité d'un historique conséquent. Pour des projections stochastiques de la mortalité, nous avons limité notre analyse pour les hommes, femmes et tous sexes confondus à une tranche d'âge de base de 50 à 90 ans et inclus l'expérience de 1982 à 2019 ou 2021 selon les cas de figure : avant ou y compris Covid19.

L'objectif de notre modélisation est de prédire les taux de mortalité futurs. Pour ce faire, plusieurs modèles ont été testés : Lee-Carter, RH, APC, BCD, M5 et M7. Les résultats de ces modèles n'étaient pas concluants. C'est pour cette raison qu'on a opté pour une modélisation de la mortalité selon une approche bayésienne. Dans ce cadre, deux méthodes, le Stacking et pseudo-BMA, basées sur la « leave-future-out validation » ont été testées et comparées au modèle bayésien standard de moyenne (BMA).

Ensuite, une table de mortalité d'expérience a été construite en se basant sur des données confidentielles d'un organisme de protection sociale. En outre, trois méthodes (SMR, BRASS, GLM-Poisson), déterminées par positionnement par rapport à une table de mortalité de référence ont été testées. Finalement, la méthode « SMR » a été retenue pour réaliser la table de mortalité prospective d'expérience pour les femmes, les hommes et tous sexes confondus avant et y compris covid 19. Il s'agit d'une application d'un coefficient de passage dit « SMR » aux taux de mortalité issus de la table de référence.

Certes, des travaux antérieurs ont élaboré des tables de mortalité d'expérience prospective en France ; mais ce mémoire a mis en œuvre une méthode récente fondée sur l'approche bayésienne et une méthode ensembliste qui permettrait de fournir des résultats plus pertinents et moins biaisés qu'une approche fréquentiste basée sur un modèle unique.

▪ Impact de la suppression des tables de mortalité générées des contrats épargne retraite collectifs

L'impact d'un changement de table de mortalité sur les provisions mathématiques d'un contrat de retraite en assurance vie est significatif, car les provisions mathématiques sont directement liées aux hypothèses actuarielles utilisées pour évaluer les engagements futurs de l'assureur. La table de mortalité reflète les probabilités de survie et de décès, qui influencent la durée des paiements de

rentes viagères, l'estimation des flux futurs, et donc le montant à provisionner pour garantir ces engagements.

La suppression des tables de mortalité genrées et le passage à une table unisexe a des effets notables sur les provisions mathématiques dans les contrats de retraite en assurance vie. En général :

- Les provisions pour les femmes diminuent, car la nouvelle table unisexe prévoit une durée de vie plus courte pour elles.
- Les provisions pour les hommes augmentent, car la nouvelle table prévoit une durée de vie plus longue pour eux.
- L'impact global dépendra de la composition du portefeuille d'assurés, mais aussi des hypothèses de gestion des risques, des taux d'actualisation, et des ajustements nécessaires pour les primes des nouveaux souscripteurs.

Cela nécessite une gestion prudente et une révision des modèles de provisionnement, tout en tenant compte de l'incertitude accrue et des exigences réglementaires.

Nous remarquons, en comparant les coûts du contrat retraite calculés à partir de la table d'expérience et ceux calculés selon la table Insee 2007-2060 ou bien TGH05/TGF05, des divergences importantes. En effet, les tarifs calculés à partir des tables réglementaires sont plus chers que ceux calculés selon la table d'expérience du portefeuille.

L'application de tables de mortalité genrées (TGH05 pour les hommes et TGF05 pour les femmes) permet d'adapter le calcul de la rente viagère selon le sexe de l'assurée. Les femmes ayant en moyenne une espérance de vie supérieure à celle des hommes, l'assureur doit donc verser la rente plus longtemps à une femme qu'à un homme. Ceci implique que les rentes viagères versées aux femmes sont mécaniquement inférieures à celles versées aux hommes.

Avec l'interdiction d'utiliser les tables de mortalité différenciées par sexe et dans l'attente d'éventuelles tables de mortalité réglementaires, ce sera très probablement la table de mortalité des femmes (dite « TGF05 ») qui sera appliquée pour tous par les assureurs. Néanmoins, cela générerait une augmentation de provisionnement de 10% au titre des engagements liés à la population masculine.

▪ Répercussions de la Covid 19

Au vu des statistiques portant sur la mortalité récente en France et publiées par l'Insee, il paraît peu probable que la Covid19 n'ait pas d'incidence sur la mortalité. Il devrait y avoir au minimum un effet à court terme. En effet, il s'avère que la Covid19 n'a pas affecté tous les sous-groupes de population de la même manière. Les nouvelles données mettent en évidence des différences basées sur l'âge et le sexe. Ces disparités peuvent avoir des répercussions à court et à long terme sur un contrat d'épargne retraite collectif.

Aussi, ce mémoire a mis en évidence l'impact à long terme de la Covid 19 sur l'évolution de l'espérance de vie en France ; où on a réalisé des projections avant covid partant d'un historique de la période 1982-2019 et des projections en intégrant deux années supplémentaires touchées par la crise de covid 19, soient les données de deux années 2020 et 2021.

Il est important de tenir compte dans nos projections de la covid19 parce que selon Swiss Re, « La Covid-19 pourrait entraîner la plus longue période de surmortalité en temps de paix ». Cela est

confirmée par les statistiques de l'Insee qui met en évidence, trois ans après le pic de la pandémie de la Covid-19, une surmortalité en France toutes causes confondues élevée par rapport aux niveaux d'avant la pandémie.

▪ **Limites du travail réalisé**

L'ancienneté de la base de données du portefeuille limité à un historique de trois ans de 2006 à 2008. Aussi, cette base n'intègre pas des données concernant l'épargne des assurés. D'où le travail réalisé sur des cas types au lieu d'être réalisé sur les données épargne réelles d'un portefeuille.

▪ **Perspectives**

- Etendre ce travail réalisé dans le cadre de ce mémoire à des données récentes d'un régime de retraite supplémentaire incluant des données réelles sur l'épargne retraite collective obligatoire.
- Développer une méthodologie pour intégrer l'expérience historique, y compris les pandémies dans les modèles de projection de la mortalité, afin de tenir compte de la possibilité de chocs futurs en matière de mortalité. Cela passerait par l'adaptation de la technique de validation au cas où l'évolution de la mortalité présenterait un changement de régime. En fait, comme discuté avec l'impact de type Covid 19, l'approche de modèle ensembliste devrait attribuer plus de poids aux modèles qui sont non seulement efficaces pour représenter le passé mais aussi pour prévoir l'avenir. Il nécessiterait dès lors une introduction de certaines techniques potentielles de changement de régime dans les modèles considérés afin de résoudre un tel problème.
- Même si les données sont encore récentes et qu'il est prématuré de tirer des conclusions définitives, il est important de garder à l'esprit qu'une grande incertitude dans la prévision de la mortalité future suite aux répercussions de la covid 19 subsiste. D'où la nécessité d'une mise à jour régulière des modèles de projection de mortalité des compagnies d'assurance.

Bibliographie

I. OUVRAGES

1. Delwarde A., Denuit M. (2006), Construction de tables de mortalité périodiques et prospectives. *Economica*.
2. Gérard F., De Courson C. (2024), Rapport d'information sur la fiscalité de l'épargne retraite par capitalisation- Assemblée Nationale.
3. Planchet F.,Thérond, P. (2011), Modélisation statistique des phénomènes de durée - Applications actuarielles. Assurance Audit Actuariat. *Economica* Paris.

II. ARTICLES

4. Barigou K., Goffard P.-O., Loisel S., Salhi Y. (2022). Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting*, In press, <10.1016/j.ijforecast.2022.01.011>. (hal-03175212v3).
5. Blanpain N., Chardon O. (2010), « Projections de population à l'horizon 2060 Un tiers de la population âgé de plus de 60 ans », *Insee Première*, n° 1320.
6. Brass W. (1971). On the scale of mortality. In W. Brass (ed.), editor, *Biological aspects of demography*. London : Taylor & Francis.
7. Cairns, A. J. G., D. Blake and K. Dowd (2006a) : Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin*. 36(1): 79-120.
8. Cairns, A. J. G., D. Blake and K. Dowd (2006b) : A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *The Journal of Risk and Insurance*. 73(4): 687-718.
9. Cairns A.J.C.(2008) : Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal* 2-3, p79-113.
10. Lee, R. D. , Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.
11. Lee, R. D. (2000) : The Lee-Carter Method of Forecasting Mortality, With Various Extensions and Applications. *North American Actuarial Journal*, 4(11): 80-93.
12. Lee, R. Miller. T. (2001) : Evaluating the Performance of Lee-Carter Mortality Forecasts. *Demography*. 38(4): 537-549.
13. Papon S. (2024), « Bilan Démographique 2023 », *Insee Première*, n° 1978.
14. Planchet F., Lelieur Vincent (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*, Vol. 7, n°14, pp. 118-146.
15. Planchet F., 2007, Tables de mortalité d'expérience pour des portefeuilles de rentiers (Tables TGH 05 et TGF 05), Notice de présentation, Institut des actuaires.
16. Planchet F., Tomas J. (2013a). Construction et validation des références de mortalité de place. Note de travail II1291-11 v1.4, Institut des Actuaires.
17. Planchet F., Tomas J. (2013b). Critères de validation : aspects méthodologiques. Note de travail II1291-14 v1.2, Institut des Actuaires.
18. Planchet F. , Tomas J. (2013c). Méthodes de positionnement : aspects méthodologiques. Note de travail II1291-12v1.5, Institut des Actuaires.Planchet F., Tomas J., et Youssef W., *ELT* (2016), *Experience Life Tables*, R package version 1.6.
19. Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570.

20. A.M. Villegas, P. Millossovich, V.K. Kaishev. (2018) StMoMo : An R Package for Stochastic Mortality Modeling. Journal of Statistical Software, Volume 84, Issue 3.

III. COURS

21. Demesttre O. (2022), Actuariat-Vie, CNAM.
22. Faure D. (2023) : Réglementation des entreprises et droit du contrat d'assurance, CNAM.
23. Planchet F. (2010) : Modèles de durée. IFSA.

IV. MEMOIRES

24. Biagini N.(2020), « Modélisation de la mortalité sur un portefeuille de retraite collective », Mémoire d'actuariat, Institut du Risque Management.
25. Damasse A. (2015), « Mortalité prospective et construction de tables de mortalité d'expérience », Mémoire d'actuariat, ISFA.
26. Fall N.- Y. (2020), « Construction de tables de mortalité prospectives et mise en place d'un indicateur de tendance de la longévité », Mémoire d'actuariat, ENSAE.
27. Martinez L. (2015), « Construction de tables de mortalité d'expérience et analyse du risque de longévité : application à un régime de retraite supplémentaire », Mémoire d'actuariat, Centre d'études actuarielles.
28. Ouhammou A. (2023), « Construction d'une table de mortalité prospective de référence dans un contexte post-COVID pour des portefeuilles de rentiers en Angleterre et au Pays de Galles », Mémoire d'actuariat, EURIA.

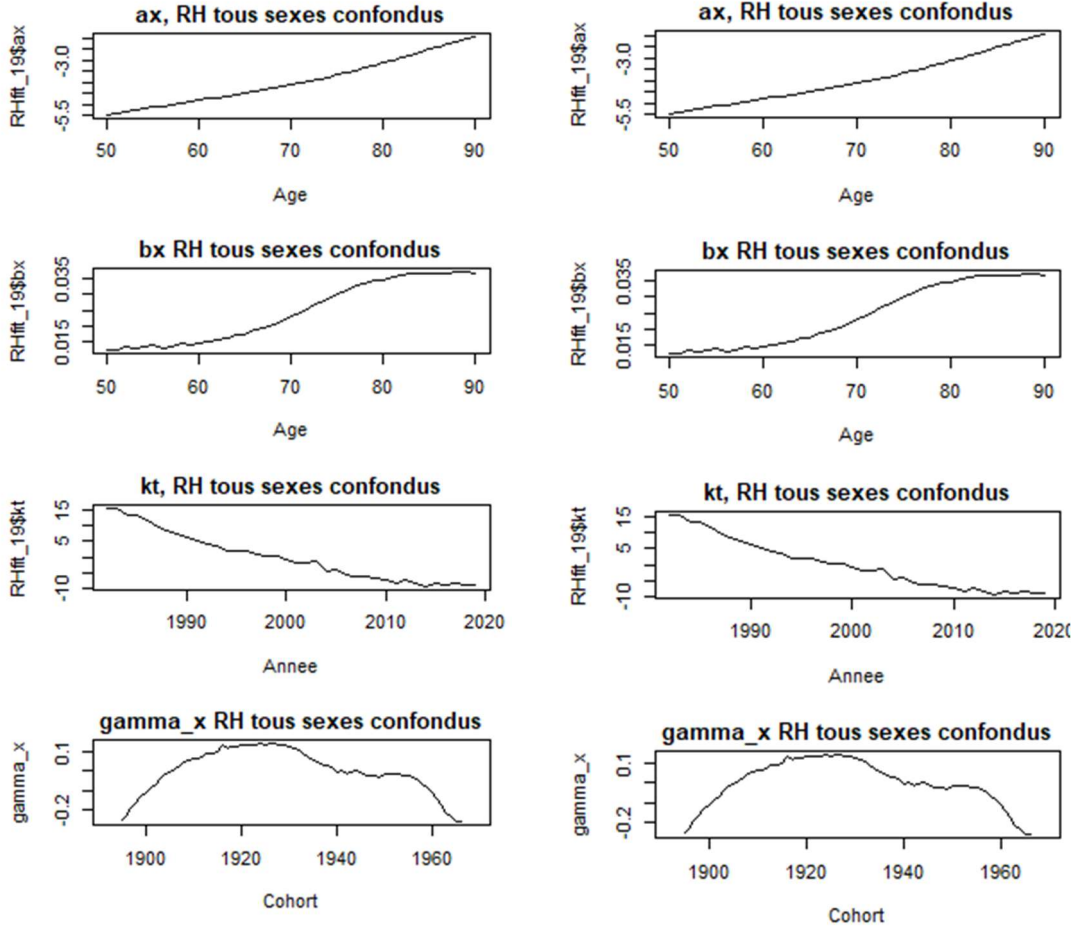
V. SITE INTERNET

29. www.ined.fr
30. www.insee.fr
31. www.mortality.org
32. www.ressources-actuariales.net/gtmortalite
33. <https://actudactuaires.typepad.com/>
34. [Code des assurances - Légifrance](#)

Annexes

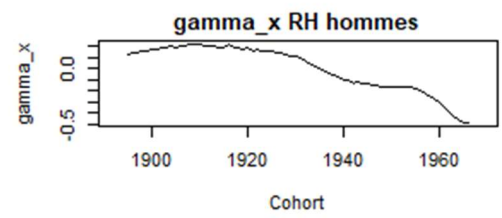
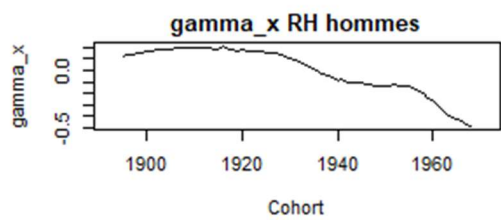
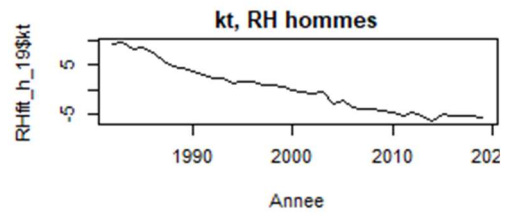
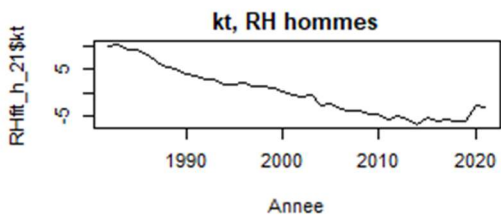
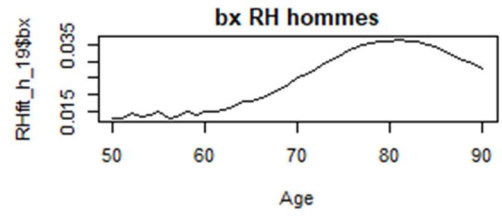
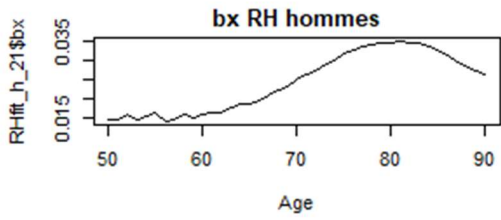
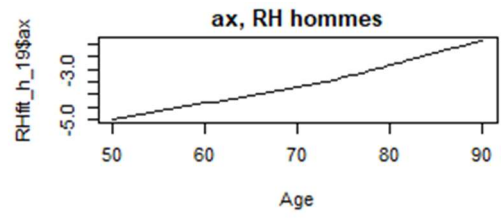
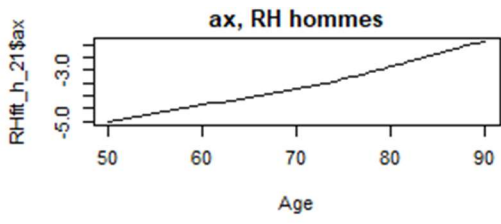
Annexe A

A-1/ Graphiques des paramètres du modèle RH pour les hommes, femmes, tous sexes confondus, avant et y compris Covid pour les modèles RH



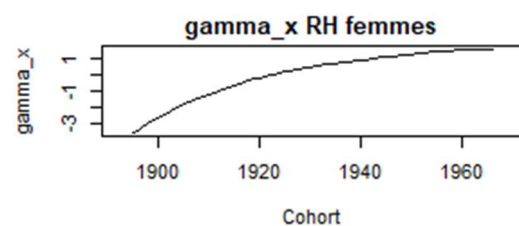
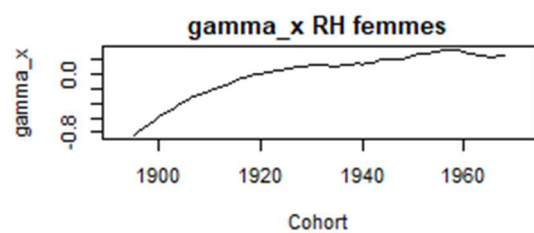
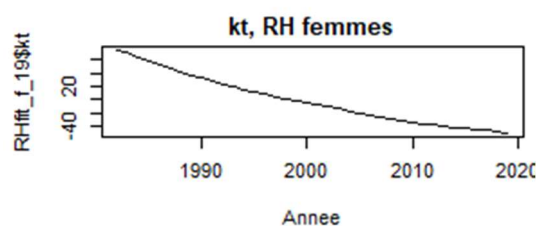
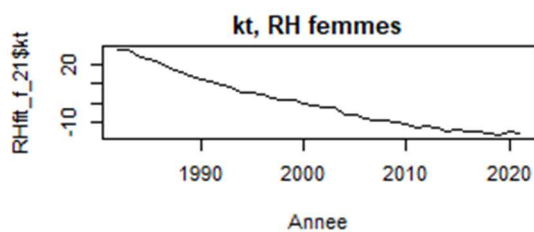
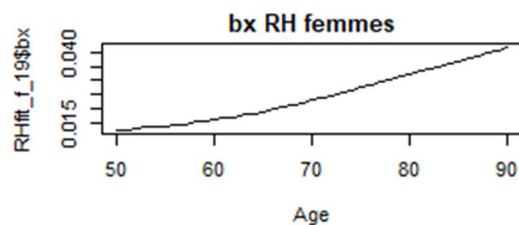
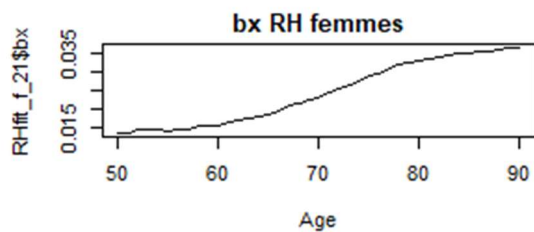
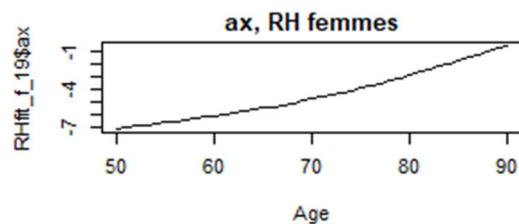
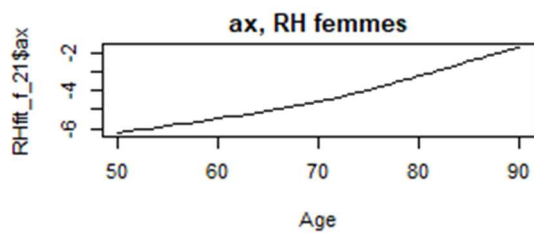
(a) y compris Covid -tous sexes confondus

(b) avant Covid- tous sexes confondus



(c) Y compris Covid -hommes

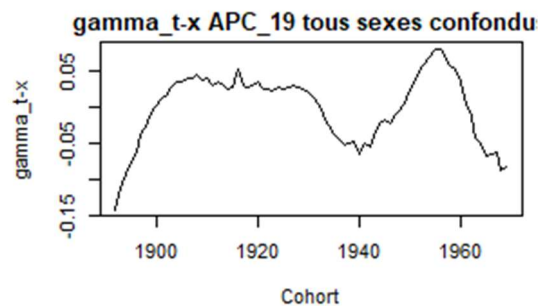
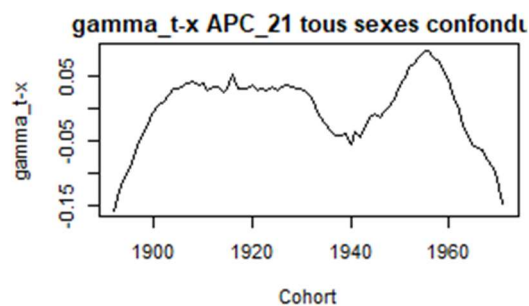
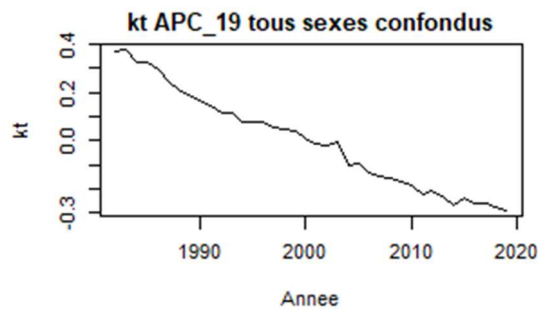
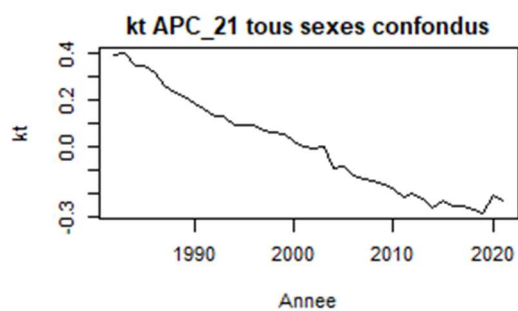
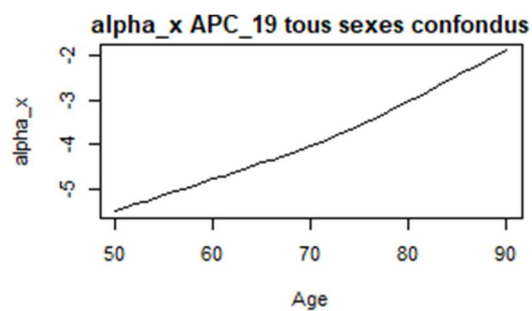
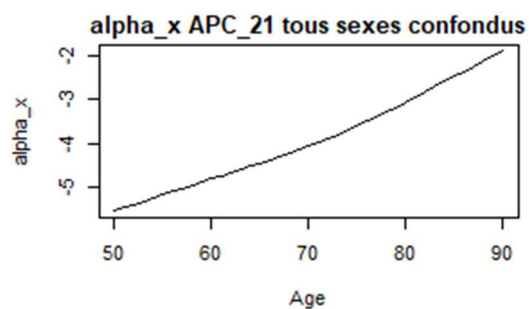
(d) avant Covid-hommes



(e) Y compris Covid -femmes

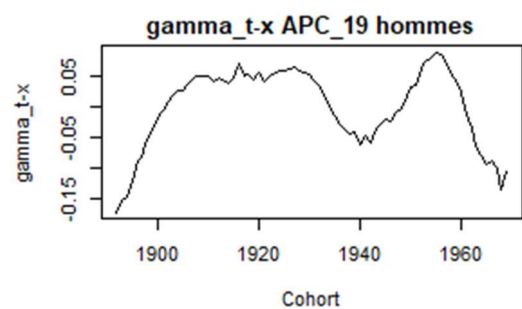
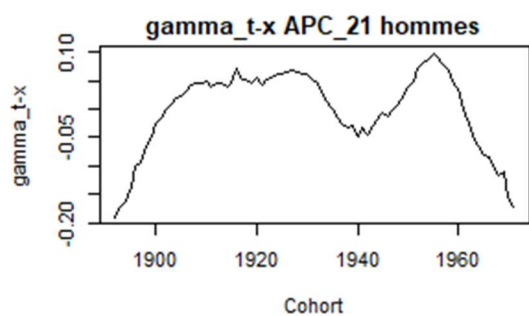
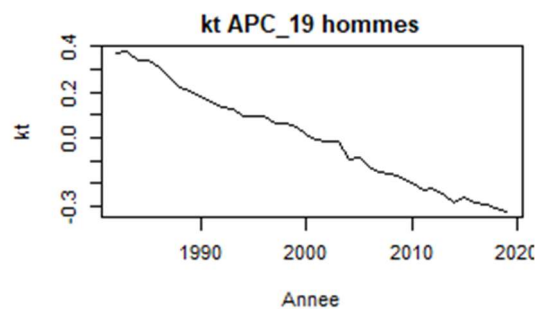
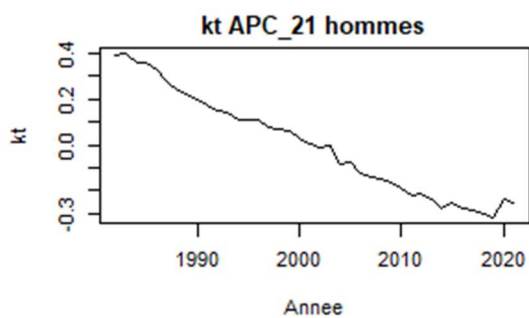
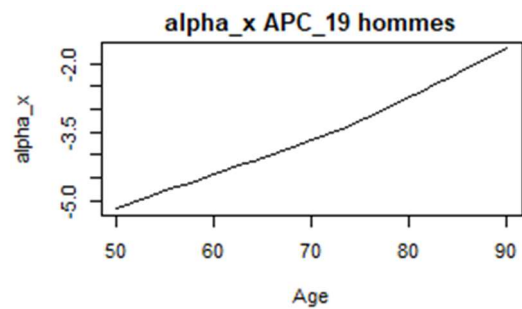
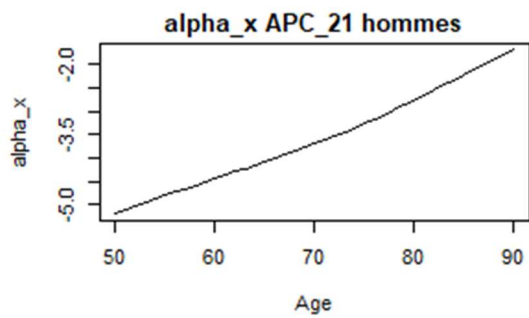
(f) Avant Covid -femmes

A-2/ Graphiques des paramètres du modèle RH pour les hommes, femmes, tous sexes confondus, avant et y compris Covid pour les modèles APC



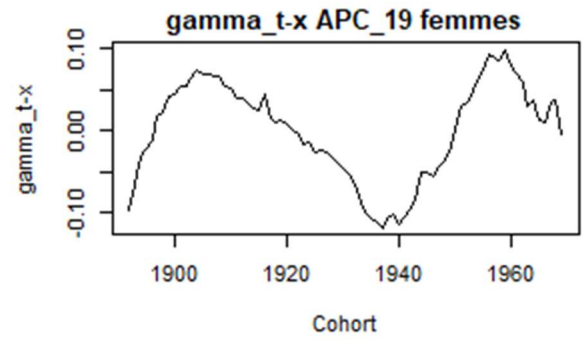
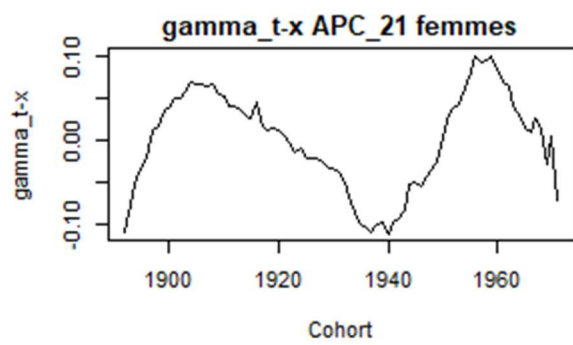
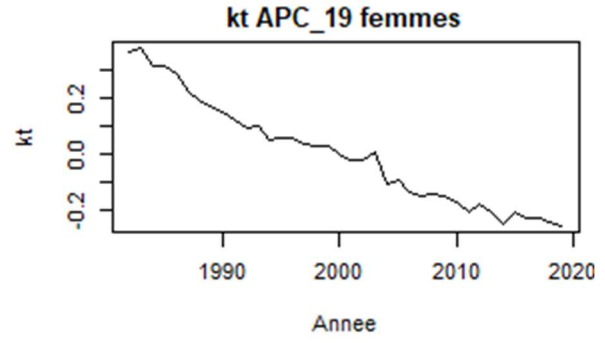
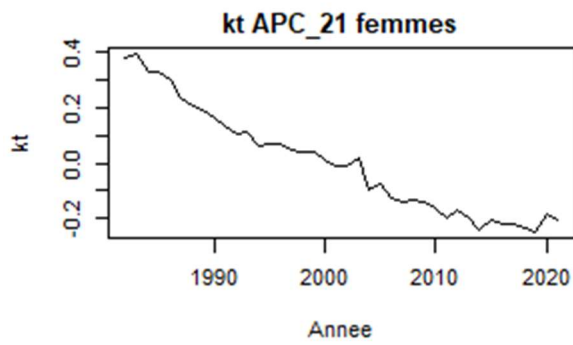
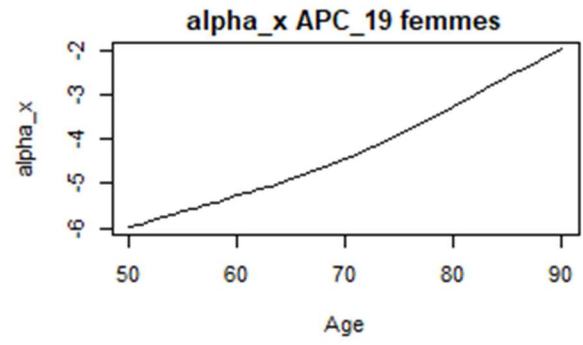
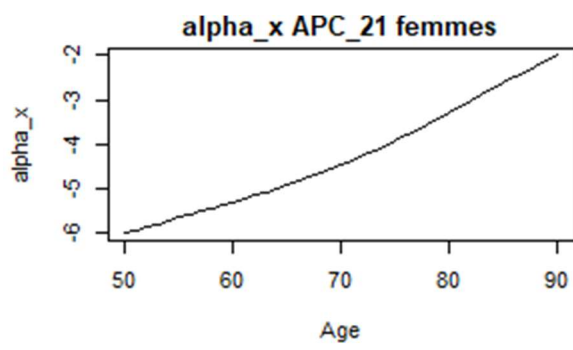
(a) y compris Covid -tous sexes confondus

(b) avant Covid- tous sexes confondus



(c) Y compris Covid -hommes

(d) avant Covid-hommes

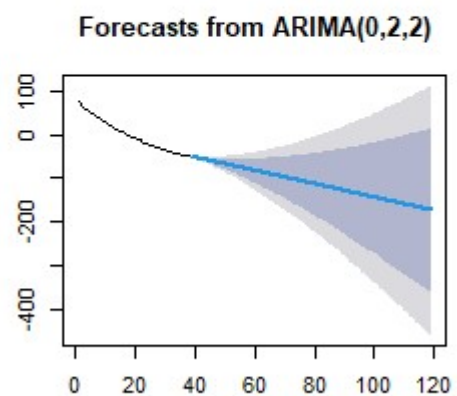
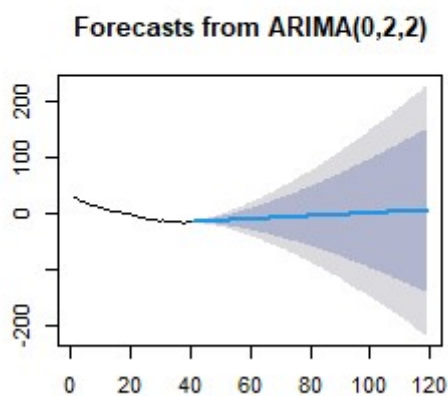
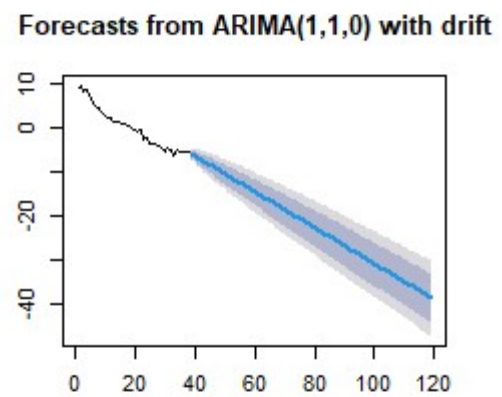
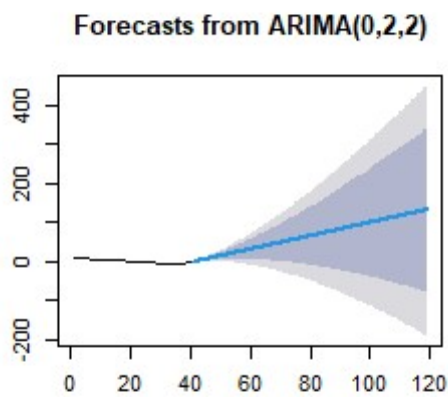
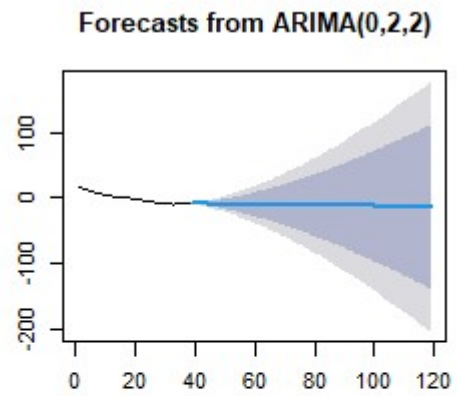
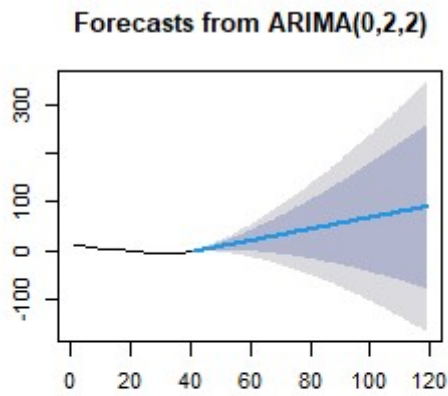


(e) Y compris Covid -femmes

(f) Avant Covid -femmes

Annexe B

Evolution de la composante temporelle des modèles RH avant covid et y compris covid, pour tous sexes confondus, hommes et femmes



RH – yc covid

RH – avant covid