

Mémoire présenté le :

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

Par : Amélie MARTIN

Titre : Construction de tables de mortalité prospectives de référence dans le cadre d'un éventuel ralentissement de l'amélioration de la mortalité

Confidentialité : NON (Durée : 1 an 2 ans)

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

Entreprise :

Nom : GALEA

Signature : //galea

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

B. YON

*Directeur de mémoire en entre-
prise :*

Nom : Irchad MAMODE VALJEE

V. KHONG

Signature :

*Membres présents du jury de
l'ISFA*

Invité :

Nom :

Signature :

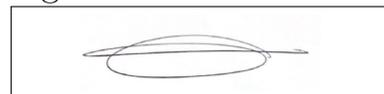
F. PLANCHET

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

Signature du responsable entreprise



Signature du candidat



Résumé

Ce mémoire s'inscrit dans un contexte où les dynamiques de mortalité sont éventuellement affectées par des perturbations récentes. Il propose une démarche prospective visant à construire des tables de mortalité de référence adaptées si les récents changements observés s'avèrent être structurels.

Les taux de mortalité ont été estimés à l'aide des modèles de LEE & CARTER, LOG POISSON et de RENSHAW & HABERMAN. Ces trois modèles ont fourni de bonnes prédictions, avec des estimations très proches des observations. Le modèle de RENSHAW & HABERMAN s'est distingué en capturant les effets de cohorte tout en offrant un bon compromis entre complexité et précision des prédictions.

Pour la projection de la composante temporelle, un modèle classique basé sur une marche aléatoire avec dérive, correspondant à un modèle ARIMA, a été utilisé. Deux scénarios distincts ont été envisagés afin d'évaluer l'impact des perturbations récentes sur les dynamiques de mortalité. Le premier scénario exclut les années marquées par des perturbations, permettant de prolonger les tendances historiques observées avant ces événements. Le second scénario intègre ces années perturbées pour capturer les changements récents dans la mortalité.

Dans une démarche d'innovation, des modèles de projection alternatifs ont été explorés, supposant que, si les changements observés dans les dynamiques de mortalité s'avèrent structurels, les modèles classiques, efficaces dans un contexte de linéarité, pourraient ne plus être adaptés. C'est dans cette optique que des approches plus flexibles, telles que les modèles ETS (Error, Trend, Seasonal) et PROPHET, ont été considérés. Ces modèles permettent de capturer les fluctuations récentes et les tendances complexes des dynamiques de mortalité. Le modèle PROPHET, minimisant l'erreur quadratique moyenne, a ainsi été retenu.

Afin d'illustrer l'utilisation concrète des différentes tables de mortalité construites dans ce mémoire, un exemple a été réalisé sur un portefeuille de retraite supplémentaire. Ce portefeuille, composé de différentes catégories d'assurés et de bénéficiaires, a servi de base pour évaluer l'impact des tables sur les calculs des Provisions Mathématiques Théoriques (PMT). Intégrer des données récentes marquées par des perturbations et/ou utiliser une méthode de projection alternative traduit une diminution de l'amélioration de la mortalité dans le temps. Cette réduction modifie la durée prévisionnelle pendant laquelle les rentes seront versées impactant directement le montant des PMT.

Mots clés : Tables de mortalité - Possible ralentissement de l'amélioration de la mortalité - Construction prospective - Modèle de LEE & CARTER - Modèle de RENSHAW & HABERMAN - Projection de la composante temporelle - Modèle ARIMA - Modèle PROPHET

Abstract

Abstract This thesis is set within a context where mortality dynamics may be affected by recent disruptions. It proposes a forward-looking approach to constructing reference mortality tables adapted to structural changes if the observed recent shifts prove to be enduring.

Mortality rates were estimated using the LEE & CARTER, LOG POISSON, and RENSHAW & HABERMAN models. All three models yielded accurate predictions, with estimates closely matching observations. The RENSHAW & HABERMAN model stood out by capturing cohort effects while offering a good balance between complexity and predictive accuracy.

For the projection of the temporal component, a classic model based on a random walk with drift, equivalent to an ARIMA model, was employed. Two distinct scenarios were considered to evaluate the impact of recent disruptions on mortality dynamics. The first scenario excluded years marked by disruptions, extending historical trends observed before these events. The second scenario included these disrupted years to capture recent changes in mortality.

In an innovative approach, alternative projection models were explored, based on the assumption that if observed changes in mortality dynamics are structural, traditional models—which are effective in linear contexts might no longer be suitable. In this vein, more flexible approaches, such as ETS (Error, Trend, Seasonal) models and PROPHET, were considered. These models allow for capturing recent fluctuations and the complex trends in mortality dynamics. The PROPHET model, minimizing mean squared error, was ultimately selected.

To illustrate the practical use of the various mortality tables constructed in this thesis, an example was conducted on a supplementary retirement portfolio. This portfolio, consisting of various categories of insured individuals and beneficiaries, served as the basis for assessing the impact of the tables on the calculation of Theoretical Mathematical Reserves (TMR). Incorporating recent data marked by disruptions and/or using an alternative projection method reflects a slowdown in mortality improvement over time. This reduction alters predictions of the duration over which annuities will be paid, directly impacting the amount of TMR.

Keywords : Mortality tables - Possible slowdown in mortality improvement - Prospective construction - LEE & CARTER model - RENSHAW & HABERMAN model - Temporal component projection - ARIMA model - PROPHET model.

Remerciements

J'adresse mes remerciements à toutes les personnes qui ont contribué à la réalisation de ce mémoire.

En premier lieu, je tiens à remercier tout particulièrement Irchad MAMODE VALJEE, sans qui ce mémoire n'aurait pas été possible. Son accompagnement, son implication active, sa disponibilité et ses précieux conseils ont été essentiels tout au long de ce travail.

Je tiens à remercier Norbert GAUTRON, Maud VANNIER-MOREAU et l'ensemble des associés du cabinet, pour m'avoir permis de réaliser ce mémoire au sein de GALEA. Leur confiance et le cadre offert ont été essentiels à l'accomplissement de ce travail. Je remercie également l'ensemble des collaborateurs pour les échanges enrichissants, qui ont contribué à approfondir ce projet.

Je remercie l'ensemble du corps professoral de l'ISFA.

Enfin, je tiens à remercier mes parents, mon conjoint, mon frère ainsi que mes amis pour leur soutien inconditionnel et leurs encouragements tout au long de cette aventure. Leur présence a été une source précieuse de motivation.

Sommaire

Synthèse	7
Synthesis	17
Introduction	27
I Contexte et Exploration des données	28
1 Le contexte historique et juridique	29
1.1 Les risques de mortalité et de longévité	29
1.2 Les tables de mortalité	30
1.2.1 Les premières réflexions sur les tables de mortalité	30
1.2.2 Les types de tables de mortalité	30
1.2.3 Les tables réglementaires ou d'expérience	30
1.2.4 Les autres tables disponibles	33
2 Le contexte mathématique	36
2.1 Les notations	36
2.1.1 La distribution de survie	36
2.1.2 Les notations pour l'analyse de la mortalité	37
2.2 Le diagramme de Lexis	39
2.3 La prise en compte des données manquantes ou incomplètes	40
2.4 La revue de la littérature des modèles de construction	41
3 Les enjeux de l'étude	43
3.1 L'espérance de vie en France	43
3.2 Les facteurs à prendre en compte dans l'analyse des tendances de mortalité dans le temps	46
3.3 La problématique et les perspectives de ce mémoire	48
4 L'étude des données de mortalité utilisées en vue d'une construction prospective	50
4.1 La présentation des bases de données	50
4.1.1 Les données de population française	50
4.1.2 Les données de décès	51
4.2 Le traitement des données	51
4.2.1 La mise en forme des données	51
4.2.2 Le calcul des taux de mortalité bruts	52
4.3 L'analyse des données	53
4.3.1 L'examen des facteurs d'hétérogénéité	53
4.3.2 L'analyse des tendances de mortalité	54
4.3.3 Le périmètre d'étude	58

II	Construction de tables de mortalité prospectives	61
1	L'estimation des taux	62
1.1	Le modèle de LEE & CARTER	62
1.2	Le modèle LOG POISSON	70
1.3	Le modèle de RENSHAW & HABERMAN	74
1.4	La comparaison des paramètres	78
2	La validation des résultats	81
2.1	L'étude des taux de mortalité observés et prédits	81
2.2	Les intervalles de confiance du nombre de décès	81
2.3	L'étude des résidus	84
2.4	L'étude du <i>Standardised Mortality Ratio</i>	87
2.5	L'étude du χ^2	89
2.6	L'étude de l'erreur moyenne des prédictions	91
2.7	L'étude du coefficient de détermination	93
2.8	L'étude du Critère d'Information Bayésien	94
3	La projection des taux de mortalité	96
3.1	Les processus $ARIMA(p, d, q)$	96
3.1.1	Définition	96
3.1.2	La méthodologie BOX & JENKINS	98
3.2	Les taux de mortalité projetés	107
4	La fermeture de la table	110
4.1	La méthode de COALE & KISKER	110
4.2	La méthode de DENUIT & GODERNIAUX	112
4.3	La comparaison des méthodes de fermeture	113
III	Modélisation alternative de la composante temporelle d'un modèle prospec-	116
tif		
1	Les modèles de projection alternatifs	117
1.1	Le modèle ETS	117
1.1.1	La présentation du modèle	117
1.1.2	L'application aux données	121
1.2	Le modèle PROPHET de Facebook	123
1.2.1	La présentation du modèle	123
1.2.2	L'application aux données	128
1.3	La comparaison des modèles alternatifs : ETS et PROPHET	131
1.4	La comparaison du modèle classique et du modèle alternatif retenu	133
2	Les nouvelles tables de mortalité	135
2.1	La projection des taux de mortalité	135
2.2	La fermeture de la table de mortalité	137

IV Impact quantitatif sur le provisionnement d'un portefeuille de retraite supplémentaire	140
1 Contextualisation	141
1.1 La retraite supplémentaire	141
1.2 Les régimes en points	142
1.3 Le régime étudié	142
1.3.1 Les assurés actifs ou radiés	142
1.3.2 Les bénéficiaires de droits directs	143
1.3.3 Les bénéficiaires de droits indirects	144
2 Le calcul des provisions	145
2.1 Les Provisions Mathématiques Théoriques (PMT)	145
2.2 Les hypothèses de calcul	145
2.3 Les méthodologies de calculs	146
2.3.1 Les assurés actifs ou radiés	146
2.3.2 Les bénéficiaires de droits directs	149
2.3.3 Les bénéficiaires de droits indirects	149
2.4 Les montants de PMT	150
Conclusion	153
Références	154
Annexes	157

Synthèse

Construction de tables de mortalité prospectives de référence dans le cadre d'un éventuel ralentissement de l'amélioration de la mortalité

Amélie MARTIN

Mots clés : Tables de mortalité - Possible ralentissement de l'amélioration de la mortalité - Construction prospective - Modèle de LEE & CARTER - Modèle de RENSHAW & HABERMAN - Projection de la composante temporelle - Modèle ARIMA - Modèle PROPHET

Les perspectives de ce mémoire

Basées sur des données historiques, les tables de mortalité capturent les dynamiques de mortalité au sein d'une population, en particulier dans un cadre prospectif où l'extrapolation des taux de mortalité permet d'analyser les tendances futures. En reflétant l'impact des conditions de vie, des soins de santé, des politiques publiques et d'autres facteurs socio-économiques sur la longévité, l'espérance de vie constitue un indicateur essentiel pour évaluer la santé globale et le bien-être d'une population. Celle-ci s'est considérablement améliorée au fil des siècles. Cependant, ces dernières années, un ralentissement de cette progression a été observé, soulevant des questions quant aux causes profondes de ce phénomène : est-il principalement imputable à des facteurs conjoncturels ou bien reflète-t-il un changement structurel plus profond ? Si ce ralentissement résulte effectivement de causes structurelles, il est essentiel d'en comprendre les implications à long terme pour la mortalité des populations.

Cette étude se propose d'explorer différents scénarios en matière de mortalité et ainsi construire des tables de référence permettant de mieux anticiper les défis futurs.

Plusieurs axes de recherche sont proposés :

- **La construction de nouvelles tables de mortalité intégrant des données récentes :** L'utilisation de données relatives aux dernières années permet de tenir compte du changement dans les dynamiques de mortalité.
- **La modélisation alternative de la composante temporelle pour la projection des taux de mortalité :** L'utilisation de modèles alternatifs dans la projection de la composante temporelle, efficaces dans un contexte de non-linéarité, permet de tenir compte des tendances historiques tout en intégrant les dernières années.

En plus d'analyser l'impact sur les taux de mortalité, ce mémoire présente un exemple d'impact de l'utilisation de ces nouvelles tables de référence sur le provisionnement d'un régime de retraite supplémentaire.

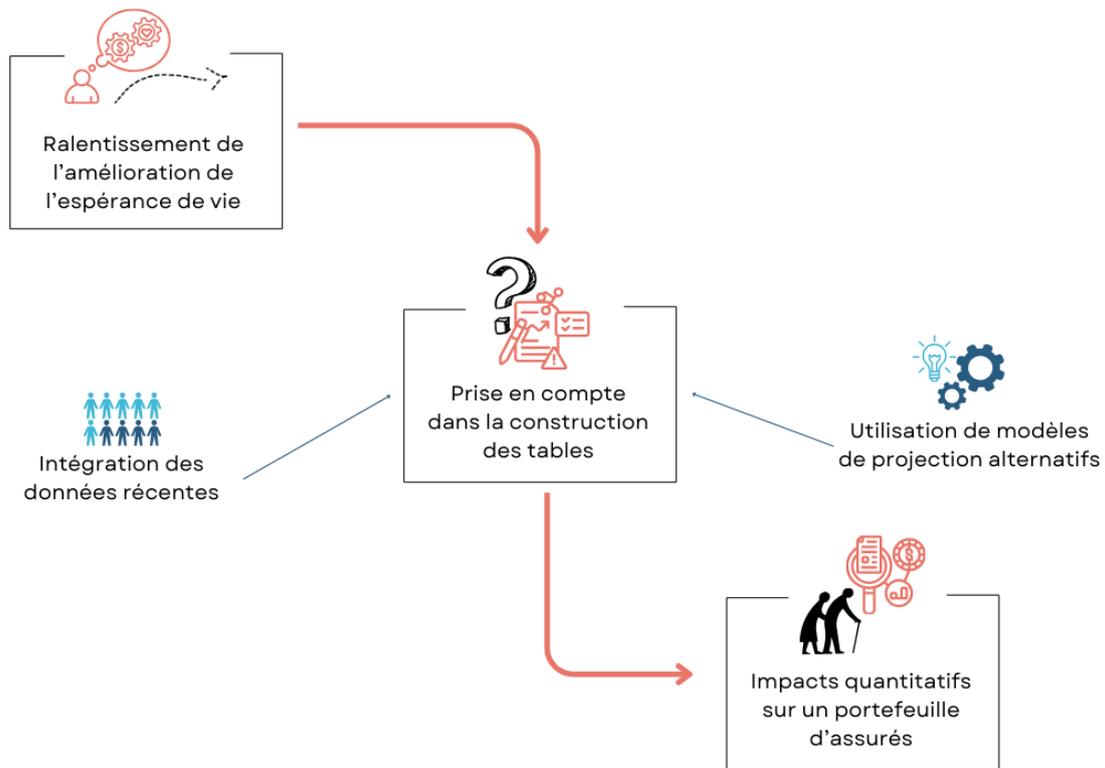


FIGURE 1 – Les perspectives du mémoire

Les données de l'étude

L'étude s'appuie sur les données de population et de décès fournies par l'Institut National de la Statistique et des Études Économiques (INSEE). À partir de ces informations, les taux de mortalité bruts par sexe ont été calculés.

L'analyse de ces taux de mortalité bruts révèle une tendance générale à la baisse sur le long terme. Cependant, une perturbation récente est observée, marquée par une hausse des taux de mortalité depuis 2019. Ce constat corrobore les observations précédentes et souligne l'importance de l'actualisation des données pour mieux appréhender les dynamiques actuelles. L'étude de d'autres approches de projections pourrait également permettre une meilleure prise en compte de ces changements s'ils s'avèrent être structurels.

Les taux de mortalité estimés

Les taux de mortalité ont été estimés selon trois méthodes : le modèle de LEE & CARTER et deux de ses extensions, le modèle LOG POISSON, qui modélise le nombre de décès par une loi de Poisson, et le modèle de RENSHAW & HABERMAN, intégrant un terme de cohorte. Ces modèles sont composés de plusieurs composantes, dont une composante temporelle qui représente l'évolution de la mortalité en fonction du temps.

En appliquant ces trois modèles, les taux de mortalité estimés sont obtenus et par conséquent les nombres de décès prédits. Ceux-ci sont comparés au nombre de décès observés à l'aide de plusieurs critères, dont le taux de ratio de mortalité standardisé (SMR). Celui-ci est proche de 1 pour les trois modèles signifiant que les nombres de décès estimés sont proches du nombre de décès observés.

Le modèle de RENSHAW & HABERMAN se distingue ensuite des autres modèles en présentant des résidus de Pearson aléatoires, indiquant un ajustement plus précis.

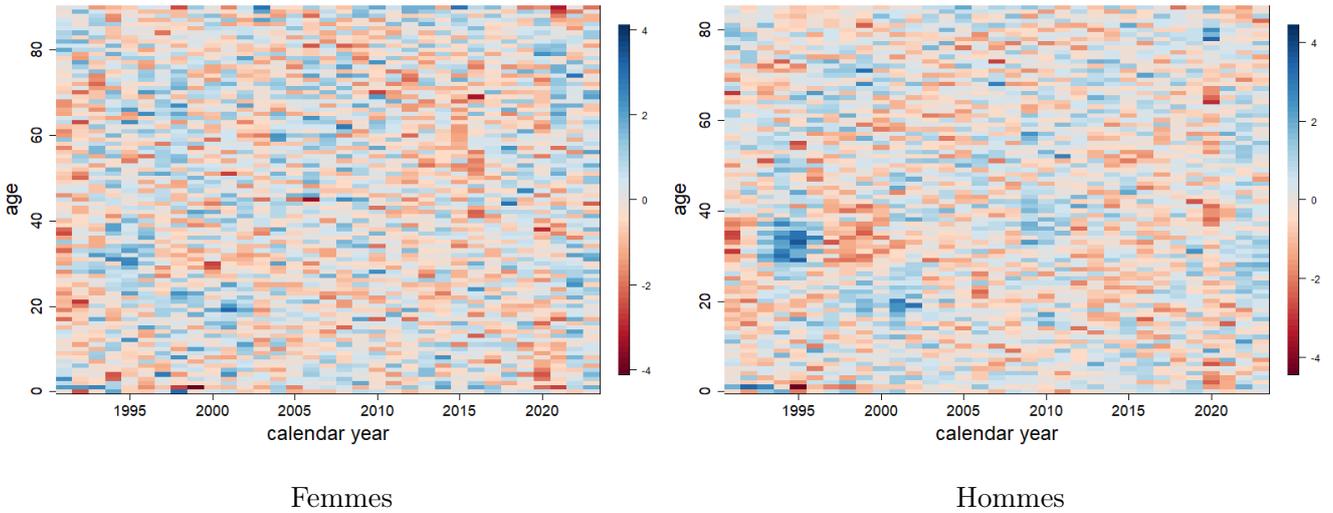


FIGURE 2 – Heatmap des résidus de Pearson du modèle RENSHAW & HABERMAN

De plus, l'analyse du critère d'information bayésien (BIC) met en évidence un meilleur compromis entre complexité et précision des prédictions avec ce modèle, ce qui conduit à le privilégier pour la suite du mémoire.

	Femmes	Hommes
LEE & CARTER	115 603	102 418
LOG POISSON	109 983	98 040
RENSHAW & HABERMAN	30 910	30 926

TABLE 1 – Les valeurs des BIC

L'intégration des données récentes dans la construction

Classiquement, dans les modèles prospectifs, la composante temporelle est assimilée à un processus stochastique, en particulier à une marche aléatoire avec dérive. Cette approche est un cas particulier de modèle de type *AutoRegressive Integrated Moving Average (ARIMA)*, précisément un modèle $ARIMA(0, 1, 0)$ avec une dérive. Ce modèle est particulièrement utile pour les séries temporelles qui présentent une tendance constante à long terme.

Afin de quantifier l'impact des perturbations récentes et de tester la robustesse de ce modèle face aux changements récents, deux scénarios de projection sont envisagés. Le premier scénario exclut les années marquées par cette perturbation (étude de la série de 1991 à 2019), tandis que le second scénario les inclut (étude de la série de 1991 à 2023).

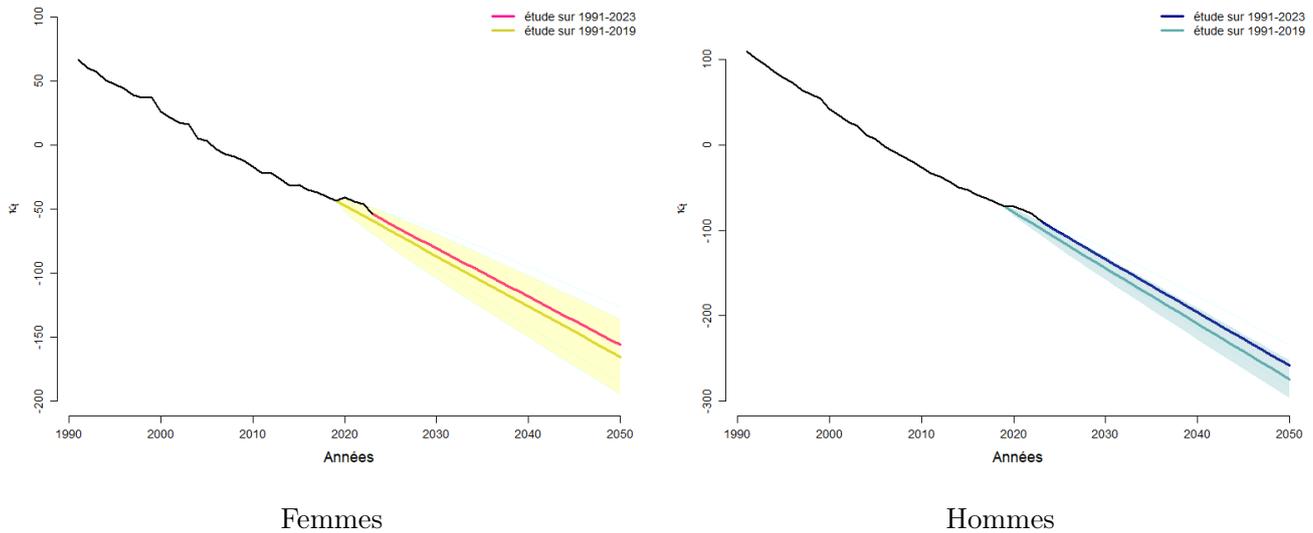


FIGURE 3 – Les projections de la série κ_t avec le modèle *ARIMA*

Les projections futures de la composante temporelle, κ_t , obtenues à l'aide du modèle *ARIMA*, continuent de suivre la tendance historique observée, avec une diminution progressive de κ_t . Cette décroissance reflète une amélioration continue des taux de mortalité, en prolongeant linéairement les dynamiques de baisse observées au cours des dernières décennies.

Toutefois, l'intégration des données les plus récentes, marquées par des perturbations, modifie cette trajectoire. Lorsque ces dernières années sont incluses dans les projections, la pente de la décroissance de κ_t devient moins marquée, indiquant un ralentissement de l'amélioration de la mortalité. Cela souligne l'importance de prendre en compte ces nouvelles données pour ajuster les modèles et refléter au mieux les dynamiques actuelles.

Les projections précises aux âges les plus avancés, sont essentielles dans le calcul des rentes viagères. Les méthodes de COALE & KISKER et de DENUIT & GODERNIAUX ont été utilisées pour extrapoler les taux de mortalité précédemment estimés et projetés. Ces deux méthodes ont montré une continuité des taux, sans cassure. La méthode de DENUIT & GODERNIAUX, qui extrapole les taux jusqu'à 130 ans, a été retenue.

Deux scénarios de mortalité sont ainsi considérés :

- ARIMA 2019 : Les taux de mortalité ont été projetés avec une méthode classique en excluant les années marquées par des perturbations.
- ARIMA 2023 : Les taux de mortalité ont été projetés avec une méthode classique en incluant les années marquées par des perturbations.



1. ESTIMATIONS

Trois méthodes de construction de tables générationnelles ont été utilisées : LEE & CARTER, LOG POISSON, RENSHAW & HABERMAN.



2. VALIDATION

La proximité des taux de mortalité prédits et observés a été examinée.



Le modèle de RENSHAW & HABERMAN est retenu.



3. PROJECTIONS

Les taux de mortalité ont été projetés avec une méthode classiquement utilisée, une marche aléatoire avec dérive. Deux scénarios ont été comparés : une table inclut les années affectées par un changement de dynamique et l'autre les exclut.



4. FERMETURE

Deux méthodes ont été comparées pour extrapoler les taux de mortalité aux très grands âges : COALE & KISKER et DENUIT & GODERNIAUX.



La méthode de DENUIT & GODERNIAUX a été retenue.



Des tables de mortalité par sexe, couvrant la période de 1991 à 2050 et les âges de 0 à 130 ans, ont été construites selon deux scénarios :

- **Les données de 1991 à 2019 ont été utilisées pour projeter les taux de mortalité de 2020 à 2050.**
- **Les données de 1991 à 2023 ont été utilisées pour projeter les taux de mortalité de 2024 à 2050.**

FIGURE 4 – Le schéma synthétique de la construction classique

La projection alternative de la composante temporelle

Dans le contexte actuel, marqué par des perturbations récentes dans les dynamiques de mortalité, il devient pertinent de réévaluer l'efficacité de la méthode de projection classiquement utilisée dans les modèles de mortalité prospectifs, à savoir une marche aléatoire avec dérive. Si les changements observés dans les tendances de mortalité sont de nature structurelle plutôt que conjoncturelle, ce modèle, efficace dans un contexte de linéarité, pourrait ne plus être en mesure de capturer correctement ces nouvelles dynamiques.

Dans cette optique, il est pertinent d'envisager des modèles de projection plus sophistiqués, tels que les modèles ETS et PROPHET, qui offrent une meilleure flexibilité pour capturer les fluctuations récentes dans les dynamiques de mortalité.

Le modèle ETS (Error, Trend, Seasonal) est un modèle de séries temporelles basé sur le lissage exponentiel, qui décompose une série en trois composantes : l'erreur (Error), la tendance (Trend) et la saisonnalité (Seasonal). Chaque composante peut être modélisée de différentes manières offrant une grande flexibilité pour capturer divers comportements de séries temporelles. Dans le cadre de ce mémoire, qui vise à élaborer une table de mortalité prospective pouvant servir de référence si les changements récents de dynamique de mortalité se révèlent être structurels, le modèle AAdN, qui combine une erreur additive, une tendance amortie et aucune composante saisonnière, constitue une approche robuste pour projeter les tendances futures.

Le modèle PROPHET est décrit comme une solution innovante pour la prévision des séries temporelles. PROPHET est une alternative aux modèles plus classiques. C'est un modèle additif qui est décomposé en plusieurs composantes, notamment la tendance, la saisonnalité et les effets d'événements particuliers. Cet outil est puissant pour la prévision des séries temporelles, offrant à la fois précision et interprétabilité. La modélisation d'une tendance linéaire avec 6 points de changement choisis automatiquement sur l'ensemble des données et qui inclut une composante de saisonnalité d'ordre de Fourier égal à 1 est retenue dans le cadre de ce mémoire.

Bien que ces projections soient globalement similaires, de légères divergences apparaissent au fur et à mesure que l'horizon de projection s'étend.

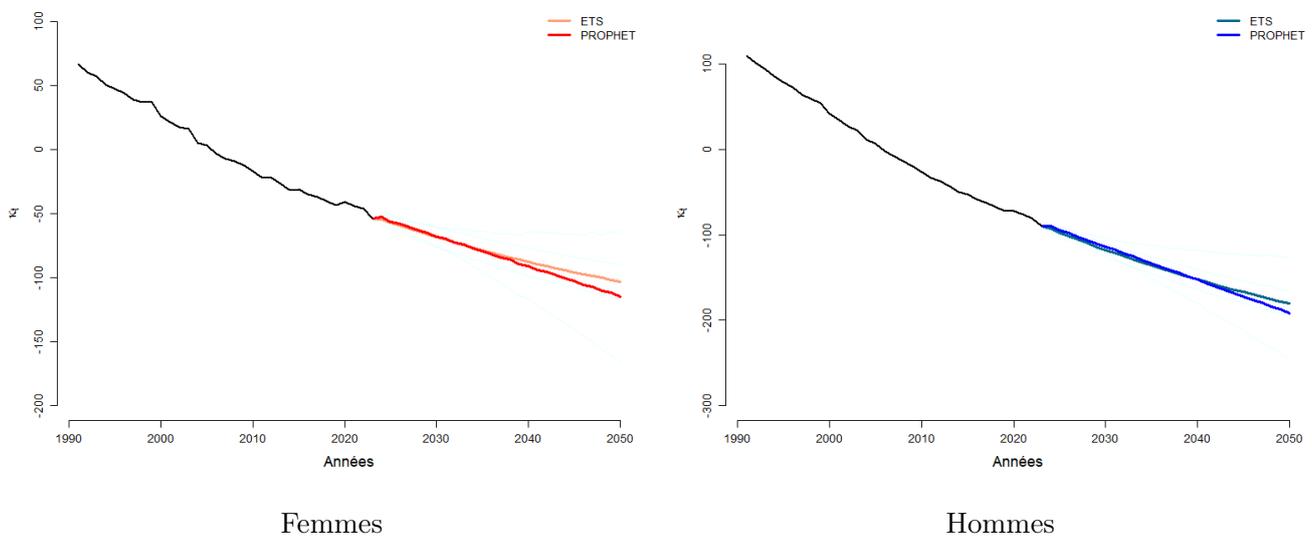


FIGURE 5 – Les projections avec les modèles ETS et PROPHET

Le modèle de PROPHET qui minimise les erreurs de prédiction par rapport au modèle ETS est ainsi retenu pour la suite de mémoire.

	Femmes	Hommes
ETS	6	3,89
PROPHET	3,45	2,17

TABLE 2 – Les valeurs du MSE des modèles ETS et PROPHET

Il est essentiel de comparer les projections obtenues avec le modèle classique (une marche aléatoire avec dérive) et le modèle PROPHET pour évaluer leur pertinence. Bien que les deux modèles, ARIMA et PROPHET, continuent d’indiquer une baisse des taux de mortalité au fil du temps, des différences notables émergent dans les résultats de ces projections. Le modèle PROPHET, en particulier, reflète plus fortement un potentiel ralentissement de l’amélioration de la mortalité, contrairement au modèle ARIMA qui tend à prolonger les tendances historiques de manière plus linéaire. En effet, le modèle PROPHET, en intégrant mieux les récents changements de tendance, suggère une décélération plus marquée des améliorations de la mortalité. À l’inverse, le modèle ARIMA maintient une perspective plus optimiste, alignée avec les tendances de baisse observées au cours des dernières décennies.

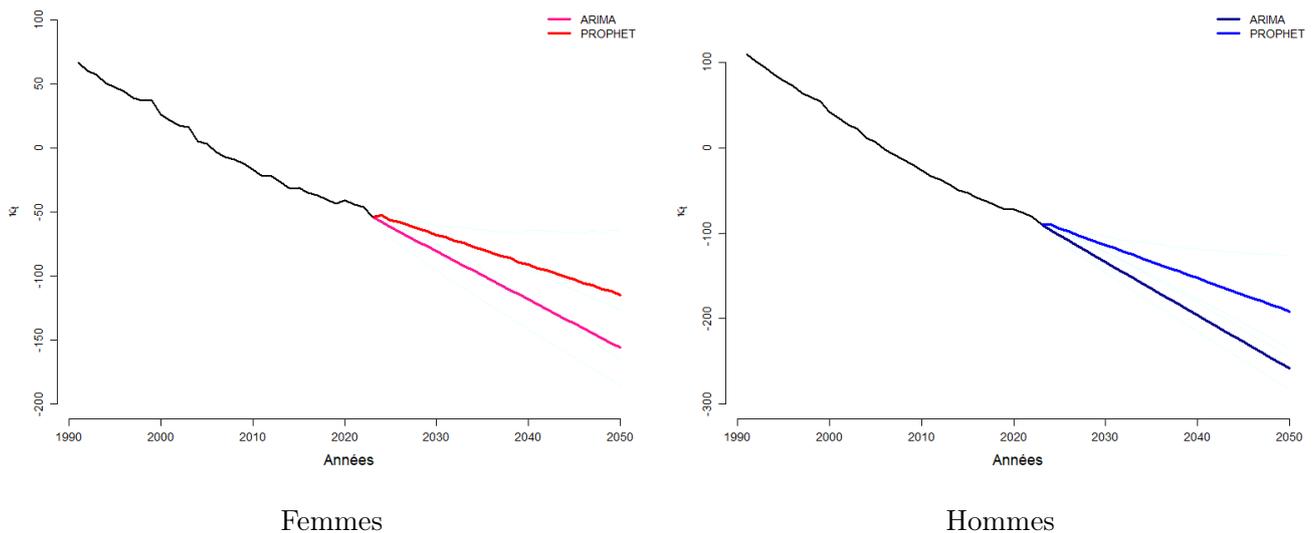


FIGURE 6 – Les projections avec les modèles ARIMA et PROPHET

Les taux de mortalité ainsi projetés ont ensuite été fermés à l’aide de la méthode de DENUIT & GODER-
NIAUX.

Un scénario de mortalité supplémentaire est ainsi considéré :

- PROPHET 2023 : Les taux de mortalité ont été projetés avec une méthode alternative en incluant les années marquées par des perturbations.



1. ESTIMATIONS

Trois méthodes de construction de tables générationnelles ont été utilisées : LEE & CARTER, LOG POISSON, RENSHAW & HABERMAN.



2. VALIDATION

La proximité des taux de mortalité prédits et observés a été examinée.



Le modèle de RENSHAW & HABERMAN est retenu.



3. PROJECTIONS

Deux méthodes alternatives ont été explorées pour la projection de la composante temporelle : ETS et PROPHET.



Le modèle PROPHET est retenu.



4. FERMETURE

La méthode de DENUIT & GODERNIAUX a été utilisée pour extrapoler les taux de mortalité aux grands âges.



Des tables de mortalité par sexe, couvrant la période de 1991 à 2050 et les âges de 0 à 130 ans, ont été construites :

- **Les données de 1991 à 2023 ont été utilisées pour projeter les taux de mortalité de 2024 à 2050 avec le modèle de PROPHET.**

FIGURE 7 – Le schéma synthétique de construction alternative

L'impact sur le provisionnement d'un régime de retraite supplémentaire

Le régime étudié est un dispositif collectif à cotisations définies, reposant sur un système de points et géré par capitalisation. Les assurés accumulent des points grâce à leurs cotisations, qui sont investies, leur donnant droit à une rente viagère réversible.

Le portefeuille d'assurés comprend 102 913 individus répartis en trois catégories :

- **Actifs ou radiés** : 87 130 individus (85%), principalement des hommes (75%), âgés de 50 à 70 ans.
- **Bénéficiaires de droits directs** : 12 501 individus (12%), majoritairement des hommes (84%), âgés de 70 à 80 ans.
- **Bénéficiaires de droits indirects** : 3 282 individus (3%), essentiellement des femmes (98%), âgées de 80 à 100 ans.

L'analyse des PMT met en évidence une sensibilité aux données utilisées. Le scénario ARIMA 2023, qui intègre des années marquées par une bosse de la mortalité, projette des montants de PMT inférieurs à ceux du scénario ARIMA 2019. Cela reflète une décélération de l'amélioration de la mortalité dans le temps, mettant en évidence l'impact des perturbations récentes sur la dynamique de mortalité et sur les engagements futurs.

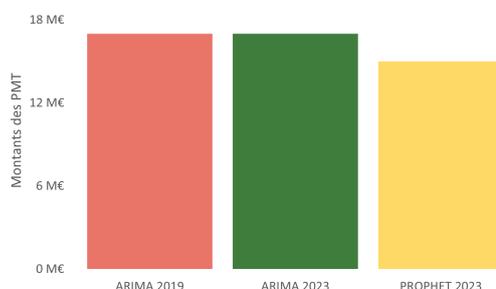
Par ailleurs, la sensibilité aux modèles de projection est fortement marquée, comme le montre le scénario PROPHET 2023, qui présente des PMT encore plus faibles. Ce scénario repose sur l'hypothèse d'un ralentissement prononcé de l'amélioration de la mortalité à long terme. Cependant, à ce stade, il est difficile de déterminer si ce scénario est représentatif ou trop extrême. Un suivi rigoureux, incluant des *backtesting* réguliers, sera nécessaire dans les années à venir pour évaluer la mortalité réelle de ce portefeuille.



Les assurés actifs ou radiés



Les bénéficiaires de droits directs



Les bénéficiaires de droits indirects

FIGURE 8 – Les sommes des PMT

En complément des hypothèses démographiques, l'actualisation des flux futurs joue un rôle clé dans l'évaluation des provisions, influençant directement les montants provisionnés. Les analyses montrent que l'application d'un taux d'actualisation réduit mécaniquement le montant des provisions nécessaires, en ramenant les engagements futurs à leur valeur actuelle. Plus le taux est élevé, plus cet effet est marqué. L'utilisation de la courbe des taux du marché entraîne une diminution significative des provisions, illustrant l'impact des paramètres financiers sur l'évaluation des engagements du régime.

Dans une optique de validation des résultats par rapport aux standards en vigueur, les PMT issues du scénario ARIMA 2023 actualisées avec la courbe EIOPA ont été comparées aux estimations obtenues avec les tables réglementaires TGH-05/TGF-05 également actualisées avec la courbe EIOPA. Le faible écart observé suggère que les projections fondées sur le scénario ARIMA 2023 sont globalement cohérentes avec les référentiels réglementaires en vigueur.

Synthesis

Construction of prospective reference life tables in the context of a potential slowdown in mortality improvement

Amélie MARTIN

Keywords : Mortality tables - Possible slowdown in mortality improvement - Prospective construction - LEE & CARTER model - RENSHAW & HABERMAN model - Temporal component projection - ARIMA model - PROPHET model

Perspectives of this thesis

Based on historical data, mortality tables capture the dynamics of mortality within a population, particularly in a prospective context where extrapolating mortality rates allows for the analysis of future trends. By reflecting the impact of living conditions, healthcare, public policies, and other socio-economic factors on longevity, life expectancy serves as a key indicator for assessing the overall health and well-being of a population. Over the centuries, life expectancy has significantly improved. However, recent years have seen a slowdown in this progression, raising questions about the underlying causes of this phenomenon : is it primarily due to temporary factors, or does it reflect a deeper structural shift? If this slowdown indeed stems from structural causes, it is critical to understand its long-term implications for population mortality.

This study aims to explore different mortality scenarios and construct reference tables to better anticipate future challenges.

Several research avenues are proposed :

- **Development of new mortality tables incorporating recent data :** Leveraging data from recent years allows for accounting for changes in mortality dynamics.
- **Alternative modeling of the temporal component for mortality rate projection :** Utilizing alternative models for projecting the temporal component, effective in non-linear contexts, enables consideration of historical trends while incorporating recent data.

In addition to analyzing the impact on mortality rates, this thesis presents an example of how these new reference tables affect the provisioning of a supplementary retirement plan.

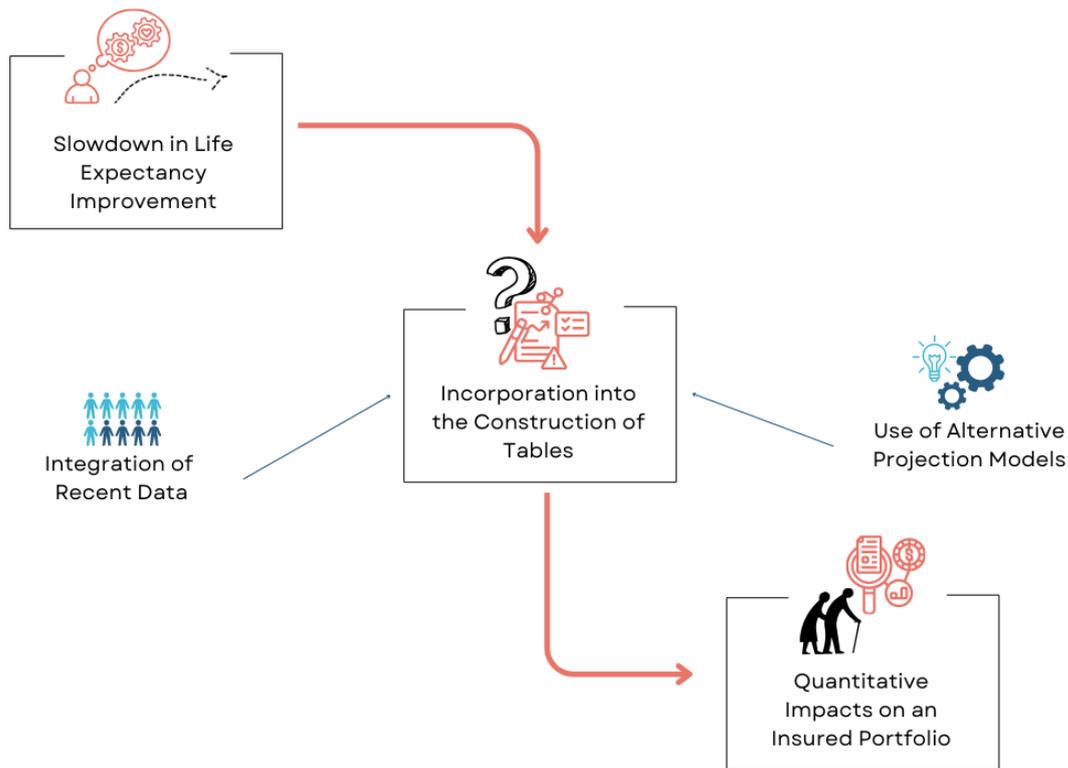


FIGURE 9 – The thesis perspectives

Study data

The study is based on population and death data provided by the National Institute of Statistics and Economic Studies. From this information, crude mortality rates by gender were calculated.

The analysis of these crude mortality rates reveals a general long-term downward trend. However, a recent disruption is observed, marked by an increase in mortality rates since 2019. This finding corroborates previous observations and highlights the importance of updating data to better understand current dynamics. Exploring other projection approaches could also enhance the ability to account for these changes if they prove to be structural.

Estimated mortality rates

Mortality rates were estimated using three methods : the LEE & CARTER model and two of its extensions, the LOG POISSON model, which models the number of deaths using a Poisson distribution, and the RENSHAW & HABERMAN model, which incorporates a cohort term. These models consist of several components, including a temporal component that represents the evolution of mortality over time.

By applying these three models, the estimated mortality rates and, consequently, the estimated number of deaths were obtained. These estimates were compared to observed deaths using several criteria, including the Standardized Mortality Ratio (SMR). For all three models, the SMR was close to 1, indicating that the estimated number of deaths closely matched the observed number of deaths.

The RENSHAW & HABERMAN model stands out from the other models by exhibiting random Pearson residuals, indicating a more accurate fit.

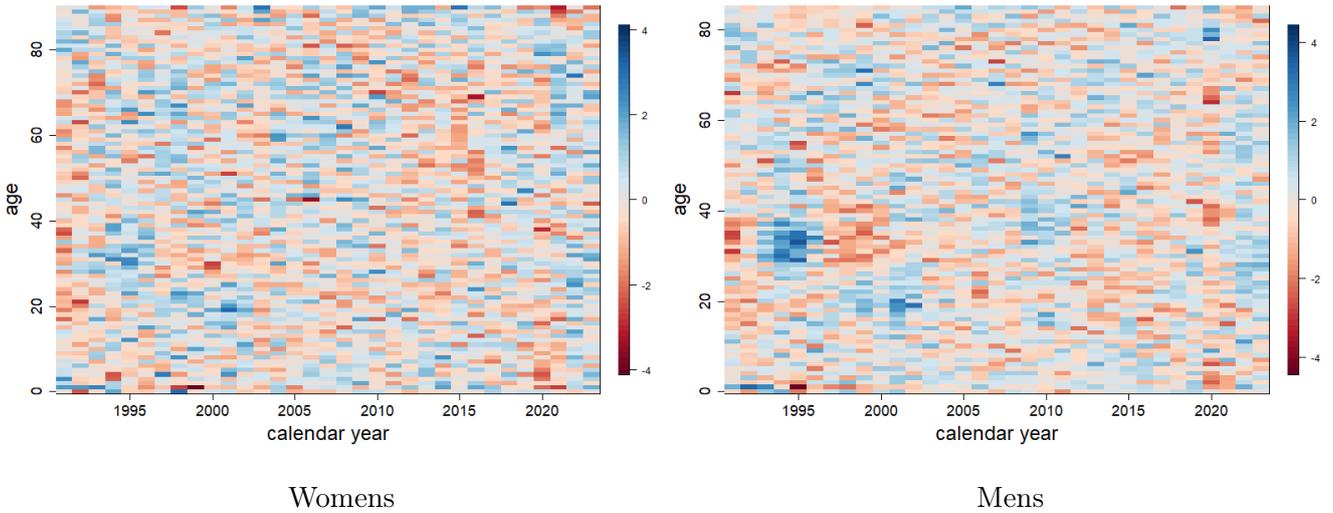


FIGURE 10 – Heatmap of Pearson residuals for the RENSHAW & HABERMAN model

Furthermore, the analysis of the Bayesian Information Criterion (BIC) highlights a better balance between complexity and predictive accuracy with this model, making it the preferred choice for the remainder of the thesis.

	Womens	Mens
LEE & CARTER	115 603	102 418
LOG POISSON	109 983	98 040
RENSHAW & HABERMAN	30 910	30 926

TABLE 3 – The values of the BIC

Integration of recent data in mortality table construction

Traditionally, in prospective models, the temporal component is modeled as a stochastic process, particularly as a random walk with drift. This approach represents a specific case of the AutoRegressive Integrated Moving Average (ARIMA) model, specifically an ARIMA(0, 1, 0) model with drift. This model is particularly useful for time series exhibiting a steady long-term trend.

To quantify the impact of recent disruptions and test the robustness of this model against recent changes, two projection scenarios are considered. The first scenario includes the years affected by this disruption (study of the series from 1991 to 2023), while the second scenario excludes them (study of the series from 1991 to 2019).

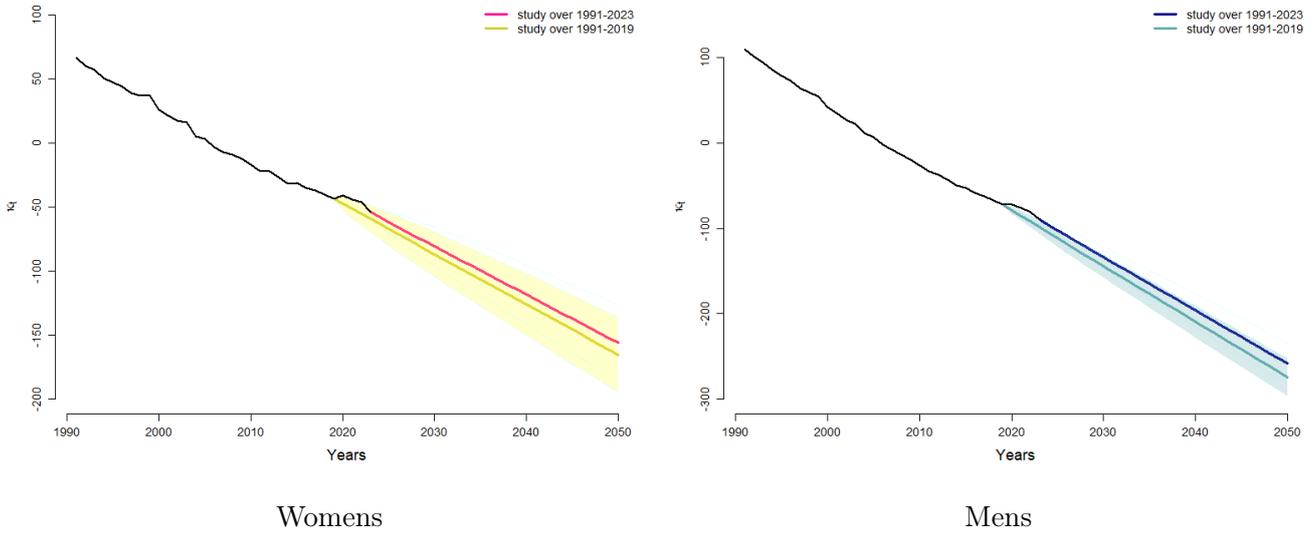


FIGURE 11 – Projections of the κ_t series using the *ARIMA* model

Future projections of the temporal component, κ_t , obtained using *ARIMA* models, continue to follow the observed historical trend, with a gradual decrease in κ_t . This decline reflects a continuous improvement in mortality rates, extending the downward dynamics observed over recent decades.

However, incorporating the most recent data, marked by disruptions, alters this trajectory. When these recent years are included in the projections, the slope of κ_t 's decline becomes less pronounced, indicating a slowdown in mortality improvement. This highlights the importance of integrating new data to adjust models and accurately reflect current dynamics.

Accurate projections for advanced ages are critical for calculating life annuities. The methods of COALE & KISKER and DENUIT & GODERNIAUX were used to extrapolate the previously estimated and projected mortality rates. Both methods demonstrated continuity in the rates, without breaks. The DENUIT & GODERNIAUX method, which extrapolates rates up to age 130, was selected.

Two mortality scenarios are thus considered :

- ARIMA 2019 : Mortality rates were projected using a traditional method, excluding years marked by disruptions.
- ARIMA 2023 : Mortality rates were projected using a traditional method, including years marked by disruptions.



1. ESTIMATIONS

Three methods for constructing generational tables were used : LEE & CARTER, LOG POISSON, and RENSHAW & HABERMAN.



2. VALIDATION

The proximity between predicted and observed mortality rates was examined.



The RENSHAW & HABERMAN model is selected.



3. PROJECTIONS

Mortality rates were projected using a commonly used method, a random walk with drift. Two scenarios were compared: one table includes the years affected by a change in dynamics, while the other excludes them.



4. CLOSURE

Two methods were compared for extrapolating mortality rates at very advanced ages : COALE & KISKER and DENUIT & GODERNIAUX.



The DENUIT & GODERNIAUX method was selected.

Mortality tables by gender, covering the period from 1991 to 2050 and ages 0 to 130, were constructed under two scenarios :

- **Data from 1991 to 2019 were used to project mortality rates from 2020 to 2050.**
- **Data from 1991 to 2023 were used to project mortality rates from 2024 to 2050.**

FIGURE 12 – The synthetic framework of the classical construction

Alternative projection of the temporal component

In the current context, marked by recent disruptions in mortality dynamics, it becomes relevant to reassess the effectiveness of the traditionally used projection method in prospective mortality models, namely the random walk with drift. If the observed changes in mortality trends are structural rather than cyclical, this model effective in a linear context—may no longer adequately capture these new dynamics.

In this light, it is worth considering more sophisticated projection models, such as ETS and PROPHET, which offer greater flexibility to capture recent fluctuations in mortality dynamics.

The ETS (Error, Trend, Seasonal) model is a time series model based on exponential smoothing, which decomposes a series into three components : error (Error), trend (Trend), and seasonality (Seasonal). Each component can be modeled in various ways, providing significant flexibility to capture diverse time series behaviors. For this thesis, which aims to develop a prospective mortality table that could serve as a reference if recent changes in mortality dynamics are structural, the AAdN model—combining additive error, damped trend, and no seasonal component—represents a robust approach for projecting future trends.

The PROPHET model is described as an innovative solution for time series forecasting. PROPHET offers an alternative to more traditional models. It is an additive model decomposed into several components, including trend, seasonality, and the effects of specific events. This tool is powerful for time series forecasting, offering both precision and interpretability. In this thesis, a linear trend model with six automatically selected changepoints across the dataset and a Fourier seasonal component of order 1 was selected.

While these projections are generally similar, slight divergences appear as the projection horizon extends.

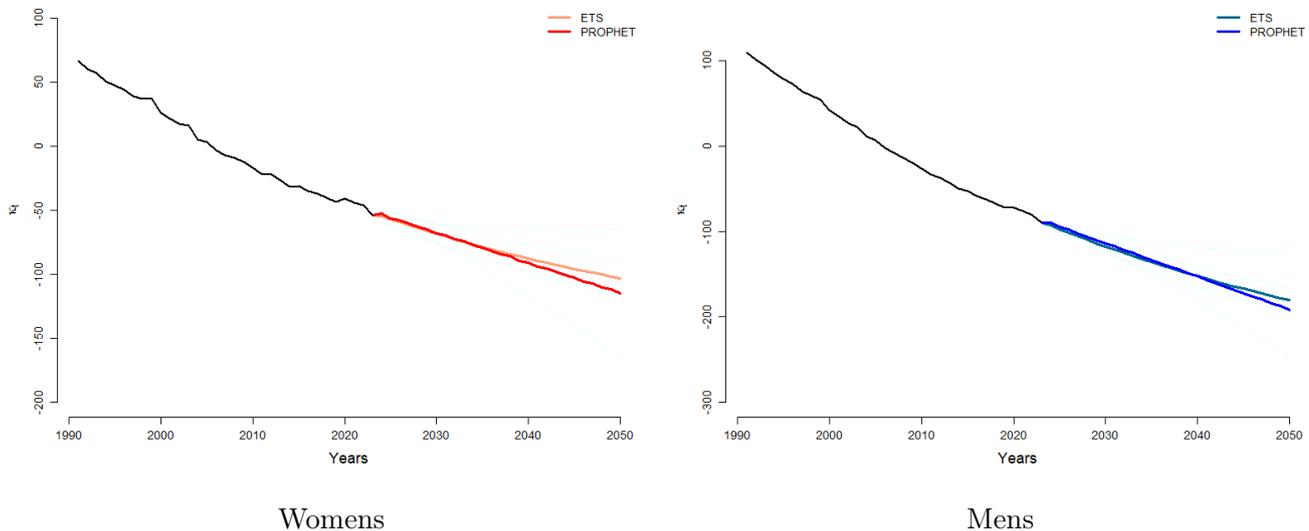


FIGURE 13 – Projections with the ETS and PROPHET models

The PROPHET model, which minimizes prediction errors compared to the ETS model, is therefore selected for the remainder of the thesis.

	Womens	Mens
ETS	6	3,89
PROPHET	3,45	2,17

TABLE 4 – The MSE values of the ETS and PROPHET models

It is essential to compare the projections obtained using the traditional model (random walk with drift) and the PROPHET model to evaluate their relevance. While both models, ARIMA and PROPHET, continue to indicate a decline in mortality rates over time, notable differences emerge in their projection results.

The PROPHET model, in particular, more strongly reflects a potential slowdown in mortality improvement, unlike the ARIMA model, which tends to extend historical trends in a more linear fashion. Indeed, by better integrating recent trend changes, the PROPHET model suggests a more pronounced deceleration in mortality improvements. Conversely, the ARIMA model maintains a more optimistic outlook, aligned with the downward trends observed over recent decades.

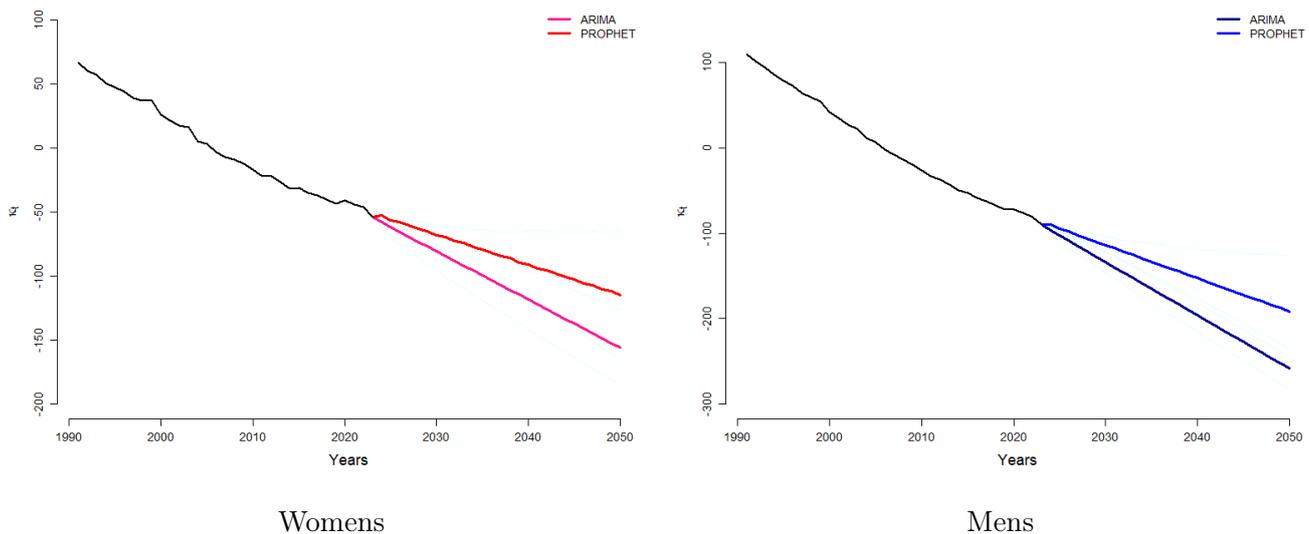


FIGURE 14 – Projections with the ARIMA and PROPHET models

The projected mortality rates were then extrapolated using the DENUIT & GODERNIAUX method.

An additional mortality scenario is thus considered :

- PROPHET 2023 : Mortality rates were projected using an alternative method, including the years marked by disruptions.



1. ESTIMATIONS

Three methods for constructing generational tables were used : LEE & CARTER, LOG POISSON, and RENSHAW & HABERMAN.



2. VALIDATION

The proximity between predicted and observed mortality rates was examined.



The RENSHAW & HABERMAN model is selected.



3. PROJECTIONS

Two alternative methods were explored for projecting the temporal component : ETS and PROPHET.



The PROPHET model is selected.



4. CLOSURE

The DENUIT & GODERNIAUX method was used to extrapolate mortality rates at advanced ages.



Mortality tables by gender, covering the period from 1991 to 2050 and ages 0 to 130, were constructed.

- **Data from 1991 to 2023 were used to project mortality rates from 2024 to 2050 using the PROPHET model.**

FIGURE 15 – The synthetic framework of alternative construction

Impact on the Provisioning of a Supplementary Retirement Plan

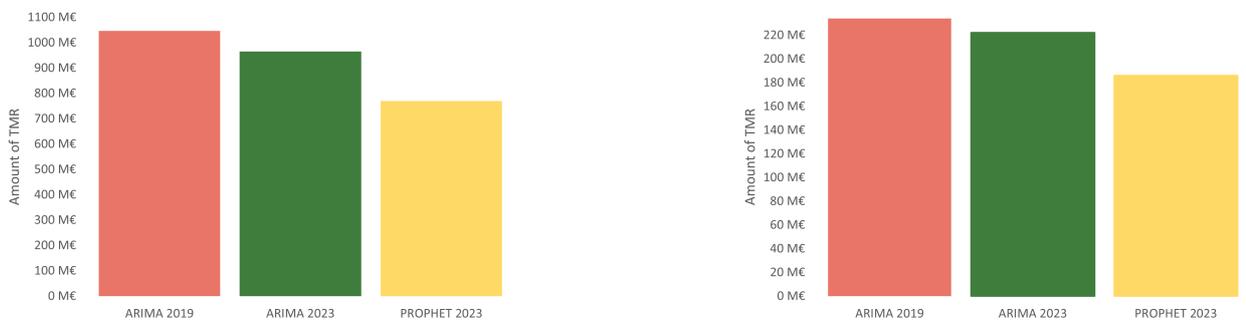
The retirement plan studied is a collective defined-contribution scheme, based on a points system and managed through capitalization. Insured individuals accumulate points through their contributions, which are invested and entitle them to a reversible life annuity.

The insured portfolio comprises 102,913 individuals divided into three categories :

- **Active or inactive members** : 87,130 individuals (85%), primarily men (75%), aged 50 to 70.
- **Beneficiaries of direct entitlements** : 12,501 individuals (12%), mostly men (84%), aged 70 to 80.
- **Beneficiaries of indirect entitlements** : 3,282 individuals (3%), predominantly women (98%), aged 80 to 100.

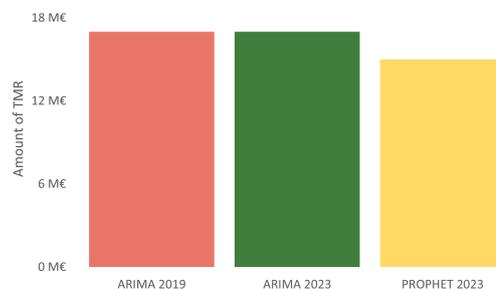
The analysis of TMR highlights sensitivity to the data used. The ARIMA 2023 scenario, which incorporates years marked by a mortality surge, projects lower TMR amounts compared to the ARIMA 2019 scenario. This reflects a deceleration in mortality improvement over time, underscoring the impact of recent disruptions on mortality dynamics and future commitments.

Additionally, sensitivity to projection models is strongly evident, as shown by the PROPHET 2023 scenario, which presents even lower TMR amounts. This scenario is based on the assumption of a pronounced slowdown in long-term mortality improvement. However, at this stage, it is difficult to determine whether this scenario is representative or overly extreme. Rigorous monitoring, including regular backtesting, will be necessary in the coming years to assess the actual mortality of this portfolio.



Active or inactive insured individuals

Beneficiaries of direct entitlements



Beneficiaries of indirect entitlements

FIGURE 16 – The totals of TMR

In addition to demographic assumptions, the discounting of future cash flows plays a key role in the assessment of provisions, directly influencing the amounts set aside. Analyses show that applying a discount rate mechanically reduces the required provisions by bringing future commitments back to their present value. The higher the rate, the more pronounced this effect. Using the market yield curve results in a significant reduction in provisions, highlighting the impact of financial parameters on the assessment of the scheme's liabilities.

To validate the results against prevailing standards, the PMT derived from the ARIMA 2023 scenario, discounted using the EIOPA curve, were compared to estimates obtained using the regulatory tables TGH-05/TGF-05, also discounted with the EIOPA curve. The small discrepancy observed suggests that projections based on the ARIMA 2023 scenario are generally consistent with the applicable regulatory benchmarks.

Introduction

Dans le secteur de l'assurance, et particulièrement en gestion des risques, comprendre et maîtriser les dynamiques de mortalité et de longévité est essentiel. Cette expertise est indispensable pour les organismes d'assurance vie, dont les produits, comme les rentes viagères, sont étroitement liés aux évolutions de la mortalité. Une connaissance approfondie des tables de mortalité prospectives permet à ces organismes de tarifer et de provisionner leurs produits avec précision, assurant ainsi leur solvabilité à long terme et leur capacité à honorer leurs engagements pris envers les assurés.

L'élaboration de projections fiables des taux de mortalité constitue un défi majeur pour les assureurs et les décideurs publics, qui s'appuient sur ces prévisions pour anticiper les besoins futurs et gérer les risques financiers associés aux produits de longévité. Historiquement, les taux de mortalité ont suivi une tendance à la baisse, soutenue par les progrès en matière de santé, l'amélioration des conditions de vie et les avancées technologiques. Cependant, les dernières années ont révélé un possible ralentissement de cette tendance. Si ce ralentissement est d'origine structurelle plutôt que conjoncturelle, il pourrait remettre en question la fiabilité des tables de référence actuelles, qui reposent principalement sur des évolutions linéaires et stables de la mortalité.

Face à cette situation, ce mémoire vise à explorer différents scénarios en matière de mortalité, afin de mieux anticiper les incertitudes liées à la longévité. D'une part, il s'agit d'intégrer les années marquées par des changements dans la construction des tables de mortalité et d'autre part, de modéliser la composante temporelle avec des modèles alternatifs conçus pour tenir compte des variations non linéaires.

Ce mémoire s'articule en quatre parties principales :

- La première partie présente le contexte de l'étude et les données utilisées, en détaillant les caractéristiques et les informations nécessaires à l'élaboration des tables de mortalité prospectives.
- La seconde partie est consacrée à la construction de tables de mortalité prospectives selon une méthode de projection classique, basée sur des tendances linéaires. Deux scénarios y sont envisagés : le premier exclut les données récentes marquées par des changements, tandis que le second les intègre pour évaluer leur impact.
- La troisième partie explore une approche alternative de construction des tables de mortalité prospective, fondée sur une méthode de projection prenant en compte les changements de tendance, afin de mieux adapter les prévisions aux dynamiques actuelles.
- La quatrième partie propose une application concrète de ces tables à travers l'exemple du provisionnement d'un régime de retraite supplémentaire, permettant d'illustrer l'impact de ces projections sur les engagements financiers de l'assureur.

Première partie

Contexte et Exploration des données

Chapitre 1

Le contexte historique et juridique

Dans le domaine de l'assurance, et plus spécifiquement en gestion des risques, il est crucial de comprendre et de maîtriser les dynamiques de mortalité et de longévité. Cette expertise est particulièrement importante pour les organismes d'assurance vie qui proposent des produits fortement liés à ces dynamiques, comme les rentes viagères. Une connaissance approfondie des tables de mortalité permet à ces organismes de gérer les risques de manière efficace, en leur permettant de tarifer et de provisionner leurs produits avec une grande précision. De plus, cela joue un rôle fondamental dans le maintien de la solvabilité à long terme des organismes, en garantissant que les engagements pris envers les assurés pourront être honorés.

L'objectif de ce chapitre est de présenter le risque de mortalité à travers son histoire, les principes fondamentaux et les aspects réglementaires.

1.1 Les risques de mortalité et de longévité

Les concepts, antonymes, de risque de mortalité et de risque de longévité sont fondamentaux dans le domaine de l'actuariat. Les organismes d'assurance, en particulier dans le cadre de produits d'assurance vie et des régimes de retraite, étudient avec précisions ces risques en projetant la durée de vie de leurs assurés pour garantir durablement leur stabilité financière, par exemple, à travers un provisionnement juste des engagements futurs. Ces engagements peuvent être revus à la hausse ou à la baisse selon les différents cas.

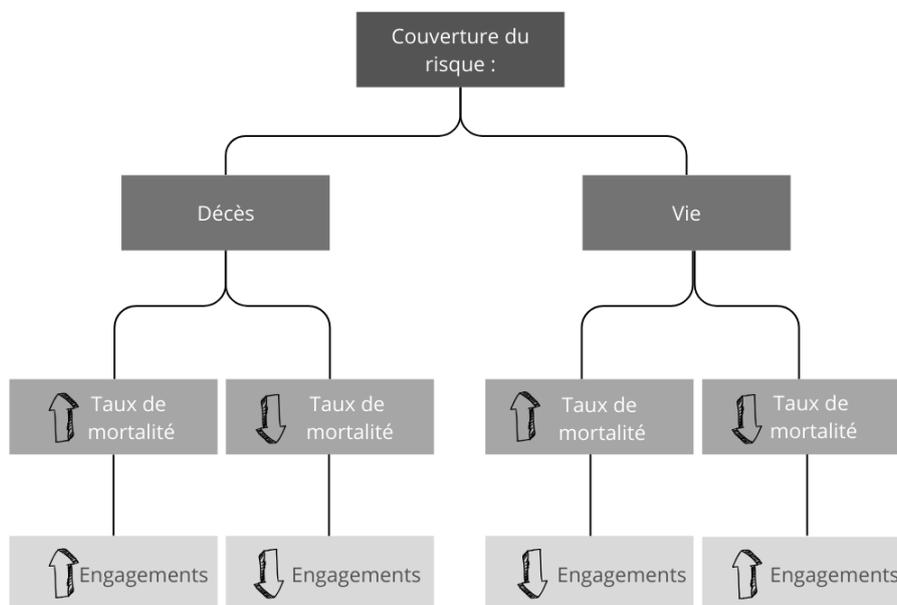


FIGURE 1.1 – Les risques de mortalité et de longévité

1.2 Les tables de mortalité

1.2.1 Les premières réflexions sur les tables de mortalité

L'intérêt porté à la quantification des probabilités de décès remonte sur plusieurs siècles, suscitant la curiosité de nombreux scientifiques. L'un des pionniers, John GRAUNT, a marqué l'utilisation des statistiques dans les études de mortalité avec son ouvrage de 1662, "*Natural and political observations upon the bills of mortality*" analysant les décès hebdomadaires de la ville de Londres. Plus connu pour ses contributions dans le domaine de l'astronomie, Edmond HALLEY, a publié en 1693 l'une des premières tables de mortalité, par âge, des temps modernes, basée sur les registres des naissances et des décès de 1687 à 1691, de la ville de Breslau en Silésie. Dans les années qui ont suivi, les mathématiciens anglais Thomas SIMPSON (1742) et James DODSON (1752) ont développé les idées énoncées précédemment. En France, l'ingénieur-mathématicien Antoine DEPARCIEUX a publié son "Essai sur les probabilités de la durée de la vie humaine" en 1746, offrant ainsi des perspectives intéressantes pour l'étude des taux de mortalité à des étapes différentes de la vie.

1.2.2 Les types de tables de mortalité

Deux types de tables de mortalité se distinguent : les tables du moment aussi appelées tables instantanées, et les tables prospectives dites de "génération".

Les tables du moment sont des tables unidimensionnelles avec comme variable explicative du décès l'âge de l'assuré, x . L'hypothèse selon laquelle les taux de décès par âge sont constants dans le temps est alors émise avec une même probabilité de décès appliquée quelle que soit la génération de la personne.

Les tables prospectives sont des tables bidimensionnelles avec comme variables explicatives du décès l'âge de l'assuré, x , mais également son année de naissance (ou, par translation, l'année calendaire, t). Ces tables permettent ainsi de visualiser les tendances d'évolution de la mortalité dans le temps. Les tables prospectives sont généralement utilisées dans le cadre de risques à horizons longs notamment pour des produits proposant des garanties viagères.

1.2.3 Les tables réglementaires ou d'expérience

Pour tarifier leurs produits ou provisionner des réserves, les organismes assureurs peuvent utiliser des tables de mortalité réglementaires, dites homologuées, ou des tables d'expérience conformément à l'Article A132-18¹ du Code des assurances.

1.2.3.1 Les tables réglementaires

Les tables de mortalité réglementaires se distinguent comme suit :

- **Les tables du moment TH 00-02 et TF 00-02** : Ces tables réglementaires sont disponibles sous la forme d'un tableau scindé en deux colonnes. Une première colonne représente l'âge de l'assuré, x , de 0 à 110 ans et une deuxième colonne représente la survie d'une cohorte de 100 000 personnes diminuant avec l'âge, noté l_x . Les tables réglementaires actuellement utilisées sont les TH 00-02 et TF 00-02. Il s'agit des tables établies à partir des données INSEE de la population française de 2000 à 2002. Elles sont construites respectivement à partir de la population masculine (TH) et de la population féminine (TF). Homologuées par l'arrêté du 20/12/2005, ces tables sont utilisées pour le provisionnement des engagements en cas de vie avec une sortie en capital et pour le provisionnement des engagements

1. L'Article A132-18 est présenté en annexe.

en cas de décès. Elles succèdent à la table TV 88-90, utilisée pour les contrats en cas de vie, et à la table TD 88-90, utilisée en cas de décès, toutes deux élaborées par l'INSEE sur la base d'observations réalisées entre 1988 et 1990. Les femmes ayant une espérance de vie moyenne plus élevée que celle des hommes, l'utilisation de tables basées sur une population féminine pour les contrats en cas de vie et sur une population masculine pour les contrats en cas de décès permet d'introduire un niveau de prudence supplémentaire dans la construction des tables. Les nouvelles tables du moment par sexe (TH 00-02 et TF 00-02), intégrant des données plus récentes, permettent de prendre en compte l'amélioration de l'espérance de vie² dans le temps. À titre d'exemple, l'espérance de vie résiduelle à 40 ans avec la table TD 88-90 était de 35,77 ans versus 37,94 ans avec la TH 00-02, soit un gain d'espérance de vie de plus de 2 ans.

Pour les contrats en cas de décès, les tables TH 00-02 et TF 00-02, s'utilisent directement, sans décalage d'âge. Concernant les contrats en cas de vie, hors rentes viagères, le texte de loi précise la nécessité d'appliquer des décalages d'âge, mais n'indique pas le mode d'emploi de ces décalages. Après une analyse, l'Institut des Actulaires indique que la méthode à recommander est celle qui fait porter les décalages d'âge sur les taux, déterminant ainsi des l_x reconstitués. Une notice d'utilisation est disponible sur le site de l'Institut des Actulaires.

- **Les tables prospectives TGH-05 et TGF-05** : Ces tables réglementaires sont disponibles sous la forme d'un tableau avec une première colonne représentant l'âge de l'assuré de 0 à 120 ans et pour chaque année de naissance de 1900 à 2005, une colonne représentant la survie d'une cohorte de 100 000 personnes diminuant avec l'âge. Homologuées par l'arrêté du 01/08/2006, les tables TGF-05 et TGH-05 succèdent à la table générationnelle TPG-93. Ces nouvelles tables ont été élaborées à partir des données de 700 000 rentes liquidées, totalisant 2 millions de lignes sur la période 1993-2005. En raison du nombre insuffisant de données sur les rentiers, notamment aux âges élevés, une méthode de positionnement par rapport à la table prospective INSEE, construite à partir des tables du moment de 1962 à 2000, a été utilisée pour la construction de ces tables.

La réglementation permet ainsi aux organismes d'assurance d'intégrer une distinction technique entre les hommes et les femmes lors du provisionnement. Toutefois, il est possible d'utiliser une unique table à condition de retenir la plus prudente : TH00-02 pour le risque décès, TF00-02 pour le risque vie avec sortie en capital et TGF05 pour le provisionnement des rentes viagères.

1.2.3.2 Les tables de mortalité d'expérience

Les tables de mortalité d'expérience sont élaborées directement par les organismes d'assurance à partir de leurs propres données d'assurés et de sinistres. Ces tables offrent une évaluation plus précise des risques en tenant compte des spécificités et caractéristiques propres à chaque portefeuille. En pratique, l'élaboration et l'autorisation d'utilisation d'une table d'expérience impliquent trois étapes :

1. La construction de la table :

La construction de tables de mortalité d'expérience est une étape essentielle pour les organismes d'assurance, permettant d'évaluer précisément les risques spécifiques à leurs portefeuilles d'assurés. Contrairement aux tables réglementaires, ces tables sont élaborées à partir des données internes des assureurs, comprenant les informations sur les assurés et les sinistres observés. Cette approche permet de capturer les particularités démographiques, socio-économiques et comportementales des assurés, ce qui se traduit par une tarification et un provisionnement plus précis.

2. Cette notion est présentée dans la suite de ce mémoire L'espérance de vie en France

- **Analyse des besoins et des caractéristiques du portefeuille** : Le processus débute par une analyse approfondie des besoins de l'assureur et des spécificités du portefeuille. Il est crucial de définir le cadre d'application des tables, en tenant compte des types de contrats (décès, rentes viagères, etc.) et des caractéristiques des assurés (âge, sexe, catégorie socioprofessionnelle, etc.). Cette phase implique également une réflexion sur le type de table à construire : tables du moment, reflétant la mortalité à un instant donné, ou tables prospectives, qui intègrent une dimension temporelle pour anticiper les évolutions futures.
- **Étude des données internes** : La qualité et la quantité des données disponibles sont ensuite évaluées. Cette étape doit permettre de s'assurer que les nombres de sinistres et d'expositions sont suffisants pour garantir la robustesse statistique des estimations. Des analyses d'hétérogénéité sont conduites pour identifier des sous-groupes au sein du portefeuille présentant des dynamiques de mortalité distinctes, comme les différences de mortalité entre les sexes ou les catégories socioprofessionnelles. Si ces différences sont significatives, il peut être pertinent de construire des tables distinctes pour chaque sous-groupe, par exemple, des tables spécifiques pour les hommes et les femmes. Deux modes de construction sont alors envisagés :
 - **Construction intrinsèque** : Cette approche consiste à exploiter directement les données du portefeuille pour estimer les taux bruts de mortalité, qui sont ensuite ajustés à l'aide de modèles statistiques avancés. L'objectif est de refléter fidèlement la réalité du portefeuille en tenant compte des spécificités des assurés et des tendances de mortalité observées.
 - **Construction par positionnement** : L'assureur utilise une table de mortalité de référence externe (généralement une table réglementaire) et applique un ajustement pour aligner les taux de mortalité avec ceux observés dans son portefeuille. Cette méthode est souvent utilisée lorsque les données internes sont insuffisantes ou lorsque le portefeuille présente des similitudes avec une population de référence bien établie.
- **Validation des tables de mortalité** : La dernière étape consiste à valider les tables élaborées. Cette validation repose sur la comparaison des taux de mortalité estimés avec ceux observés, l'évaluation de la cohérence des résultats par rapport aux hypothèses de départ, et l'application de tests statistiques pour vérifier la robustesse des estimations. Des comparaisons avec d'autres tables de mortalité, qu'elles soient réglementaires ou d'expérience, sont également effectuées pour s'assurer que les tables fournissent une évaluation fiable des risques.

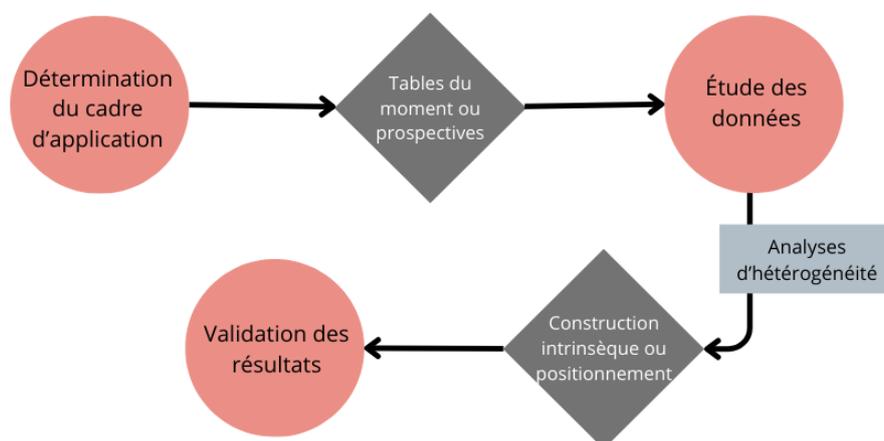


FIGURE 1.2 – Les tables d'expérience : la construction

2. La certification initiale :

La certification des tables de mortalité³, rendue obligatoire par l'Article A 132-18 du Code des assurances, assure que les tables construites reflètent fidèlement les dynamiques de mortalité des populations assurées et qu'elles permettent une gestion efficace des engagements de l'assureur, tout en anticipant les évolutions futures.

3. Le suivi annuel :

Après la certification initiale des tables de mortalité d'expérience, les organismes d'assurance doivent réaliser un suivi rigoureux pour s'assurer que ces tables restent alignées avec les caractéristiques évolutives de leur portefeuille⁴. Le rapport de suivi annuel exprime clairement le maintien ou l'abandon de la certification des tables de mortalité.

Schématisation synthétique de l'élaboration d'une table d'expérience :



FIGURE 1.3 – Synthèse sur l'élaboration d'une table d'expérience

1.2.4 Les autres tables disponibles

D'autres tables de mortalité sont également disponibles. Elles offrent des perspectives intéressantes en tenant compte de divers facteurs d'hétérogénéité.

- **Des tables différenciant le sexe :** Au-delà de ces tables réglementaires, il existe d'autres tables de mortalité qui différencient également le sexe.
 - **Les tables INSEE 2007-2060 :** L'Institut National de la Statistique et des Études Économiques (INSEE) a établi une projection démographique de la population en France métropolitaine pour la période 2007 à 2060 (BLANPAIN & CHARDON, 2010). Chaque année, le nombre de femmes et d'hommes est estimé par âge en se basant sur des hypothèses concernant trois composantes clés : la fécondité, la mortalité et les migrations (émigration et immigration). Pour chaque composante, plusieurs scénarios sont envisagés : central, haut et bas, permettant ainsi une analyse de la sensibilité des projections en fonction des scénarios retenus. Ces tables servent de référence dans le domaine de l'actuariat.

3. Le principe de certification est détaillé en annexe

4. Le principe de suivi est détaillé en annexe

- Annuellement, l'Institut National d'Études Démographiques (INED) publie une table de mortalité du moment détaillée par âge et par sexe. Les tables 2020-2022, basées sur des données arrêtées à fin 2023, sont actuellement disponibles⁵. Ces tables sont accessibles pour l'ensemble de la France depuis 1999.

— **Les tables par facteurs d'hétérogénéité autres :** Des tables sont également élaborées selon des facteurs d'hétérogénéité spécifiques qui influencent la mortalité au sein de la population générale :

- **Le lieu de résidence :** En France, la mortalité varie significativement en fonction du lieu de résidence, révélant des inégalités territoriales marquées. Ces disparités peuvent être attribuées à des facteurs socio-économiques propres à chaque région et département. L'INSEE publie des taux de mortalité par sexe et tranche d'âge pour chaque département et région, permettant ainsi une analyse fine des différences géographiques.
- **Le diplôme :** Le niveau d'éducation est un facteur déterminant de la mortalité. Les individus ayant un diplôme élevé tendent à avoir une espérance de vie plus longue que ceux ayant un niveau d'éducation inférieur. Les tables de mortalité du moment, selon l'âge et le niveau de diplôme, sont ainsi disponibles, offrant une vision plus fine des disparités en termes de longévité.
- **La catégorie socioprofessionnelle :** Des inégalités significatives sont également constatées selon la catégorie socioprofessionnelle des individus. Des tables de mortalité du moment sont disponibles par âge, sexe et catégorie socioprofessionnelle.
- **Le niveau de vie :** Le niveau de vie est un facteur déterminant de la mortalité, influençant directement l'espérance de vie des individus. Une étude de l'INSEE de 2018 montre que l'espérance de vie augmente significativement avec la richesse des individus. Des tables de mortalité du moment sont disponibles par âge, sexe, et niveau de vie. Elles permettent d'affiner l'évaluation des risques associés à ces inégalités. Elles montrent également que les femmes, malgré un niveau de vie moins élevé, ont une espérance de vie supérieure à celle des hommes les plus aisés.

5. Les tables 2020-2022 de l'INED : <https://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-e-cause-deces/table-mortalite/>



Ce qu'il faut retenir du chapitre :

- * La gestion des risques de mortalité et de longévité est fondamentale pour les assureurs, car elle influence directement la tarification des produits d'assurance et le provisionnement des engagements futurs. Une fluctuation des taux de mortalité peut entraîner des ajustements importants dans les engagements financiers des assureurs. Une bonne maîtrise de ces concepts est essentielle pour assurer la stabilité financière à long terme des organismes d'assurance.
- * Les tables de mortalité se sont développées à partir de réflexions historiques. Ces outils sont classés en deux catégories : les tables du moment, qui se concentrent sur l'âge des assurés, et les tables prospectives, qui intègrent également l'année de naissance pour prévoir les évolutions futures de la mortalité.
- * Les assureurs peuvent utiliser des tables de mortalité réglementaires, comme les TH00-02 ou TGH05, ou des tables d'expérience, basées sur leurs données internes. Les tables d'expérience offrent une évaluation plus précise, adaptée aux particularités du portefeuille. Le choix de la table dépend des besoins de tarification et de provisionnement de l'assureur.
- * En dehors des tables réglementaires, d'autres tables de mortalité existent prenant en compte divers facteurs d'hétérogénéité, tels que le sexe, le lieu de résidence, le niveau d'éducation, la catégorie socioprofessionnelle et le niveau de vie. Notamment, les tables INSEE 2007-2060 servent de référence en actuariat.

Dans la suite de l'étude, les notations mathématiques relatives aux tables de mortalité seront présentées.

Chapitre 2

Le contexte mathématique

Les notions de mortalité et de longévité reposent sur des concepts mathématiques et des outils statistiques avancés, essentiels pour modéliser les risques en assurance.

Les objectifs de ce chapitre sont les suivants :

- Définir les notations et représentations visuelles utilisées dans les études de mortalité et de longévité.
- Énoncer les défis courants liés à l'utilisation de données manquantes ou incomplètes.
- Présenter une revue des principaux modèles de mortalité, permettant de capturer les dynamiques observées dans les données historiques.

2.1 Les notations

2.1.1 La distribution de survie

La durée de vie :

La durée de vie d'un individu, représentée par la variable aléatoire T , mesure le temps écoulé depuis la naissance jusqu'au décès. Elle est définie sur l'intervalle $[0, +\infty]$ et sa fonction de répartition notée $F(t) = \mathbb{P}(T \leq t)$ est continue à droite. Si une densité existe, elle est donnée par :

$$f(t) = \frac{d}{dt}F(t) = \lim_{u \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + u)}{u}$$

La fonction de survie :

La fonction de survie de T exprime la probabilité qu'un individu vive au-delà du temps t . Notée $S(t)$, elle est définie par :

$$S(t) = 1 - F(t) = 1 - \mathbb{P}(T \leq t) = \mathbb{P}(T > t)$$

Cette fonction décroît avec le temps, partant de 1 au moment de la naissance et tendant vers 0 à mesure que l'âge augmente : $S(0) = 1$ et $\lim_{t \rightarrow \infty} S(t) = 0$.

La fonction de risque instantané :

La fonction de risque instantané, aussi appelée fonction de hasard ou fonction de taux de défaillance, mesure la probabilité de décéder instantanément sur un laps de temps $t + dt$ sachant que l'individu a survécu jusqu'à l'âge t . Elle est définie par :

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t))$$

Elle peut également s'exprimer comme :

$$h(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + dt | T > t)}{dt} = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + dt)}{dt S(t)}$$

Ainsi, la fonction de survie s'écrit en fonction de $h(t)$:

$$S(t) = e^{-\int_0^t h(s) ds}$$

La fonction de hasard cumulée :

La fonction de hasard cumulée $H(t)$ représente l'accumulation des risques de décès au fil du temps jusqu'à l'âge t . Elle est définie par :

$$H(t) = \int_0^t h(s) ds$$

Ce qui permet d'écrire la fonction de survie sous la forme :

$$S(t) = e^{-H(t)}$$

2.1.2 Les notations pour l'analyse de la mortalité

La durée de vie résiduelle :

La durée de vie résiduelle d'un individu ayant atteint l'âge x , notée T_x , représente le temps restant à vivre à partir de l'âge x . Elle se définit par :

$$T_x = {}_d[T - x | T > x]$$

La fonction de survie conditionnelle :

La fonction de survie conditionnelle représente la survie d'un individu à l'âge $x + t$, sachant qu'il a déjà vécu x années. Notée $S_x(t) = P(T > x + t | T > x)$, elle s'exprime via la fonction de survie :

$$S_x(t) = \mathbb{P}(T > x + t | T > x) = \frac{\mathbb{P}(T > x + t)}{\mathbb{P}(T > x)} = \frac{S(x + t)}{S(x)}$$

La probabilité de survie entre x et $x + t$:

La probabilité de survie entre x et $x + t$ correspond à la probabilité qu'un individu ayant déjà vécu x années vive encore t années supplémentaires. Elle est exprimée par :

$${}_t p_x = \mathbb{P}(T_x > t) = \mathbb{P}(T > x + t | T > x)$$

Elle peut se réécrire comme suit :

$${}_t p_x = \frac{S(x + t)}{S(x)}$$

Le quotient de mortalité entre x et $x + t$:

Le quotient de mortalité, représentant la probabilité qu'un individu ayant déjà vécu x années décède dans les t prochaines années, se déduit de la probabilité de survie :

$${}_tq_x = 1 - {}_tp_x = \mathbb{P}(T_x \leq t) = \mathbb{P}(T \leq x + t | T > x)$$

Le nombre d'individus vivant à l'âge x :

La population vivante à l'âge x se note :

$$l_x = S(x)$$

Ainsi, les probabilités de survie et de décès entre x et $x + t$ peuvent se réécrire comme suit :

$${}_tp_x = \frac{l_{x+t}}{l_x}$$

$${}_tq_x = 1 - {}_tp_x = 1 - \frac{l_{x+t}}{l_x} = \frac{l_x - l_{x+t}}{l_x}$$

Dans les études de mortalité, il est courant d'étudier la probabilité de décès dans l'année à un âge donné x :

$${}_1q_x = q_x = 1 - \frac{l_{x+1}}{l_x}$$

Le nombre de décès entre x et $x + t$:

Noté d_x , il est défini par :

$${}_td_x = l_x - l_{x+t}$$

Le temps vécu par les individus d'une cohorte entre x et $x + t$:

Noté ${}_tL_x$, il est défini par

$${}_tL_x = \int_0^t l_{x+u} du$$

La durée de vie résiduelle :

La durée de vie résiduelle, notée E_x , mesure l'espérance de vie restante à partir de l'âge x . Elle se définit comme suit :

$$E_x = \int_0^{\infty} l_{x+u} du = \sum_{i=x}^{\infty} L_x$$

Les taux de mortalité :

Les taux de mortalité représentent la probabilité de décès d'un individu sur une période donnée. Ils sont définis par :

$${}_tm_x = \frac{{}_td_x}{{}_tL_x}$$

Les taux de mortalité instantanés :

Les taux de mortalité instantanés mesurent le risque de décès à un instant précis. Ils se définissent comme suit :

$$\mu_{x+t} = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(t < T_x \leq t + h | T_x > t) = \frac{1}{t p_x} \frac{\partial}{\partial t} {}_t q_x$$

Extension dans le cas des études prospectives :

Ces notations peuvent être étendues au cas prospectif. Avec T la durée de vie totale de l'individu et U son année de naissance, les taux de mortalité instantanés s'écrivent :

$$\mu_{x,t} = \lim_{h \rightarrow 0} h^{-1} \mathbb{P}(t < T \leq t + h | T > t, U = t - x)$$

Le quotient de mortalité s'écrit :

$$q_{x,t} = \mathbb{P}(x \leq T < x + 1 | T \geq x, U = t - x)$$

En considérant l'hypothèse suivante vérifiée :

$$\forall (x, t) \in \mathbb{R}^2, \forall (x_0, t_0) \in [0, 1]^2, \mu_{x+x_0, t+t_0} = \mu_{x,t}$$

le quotient de mortalité peut s'écrire en fonction des taux instantanés comme suit :

$$q_{x,t} = 1 - e^{-\mu_{x,t}}$$

2.2 Le diagramme de Lexis

Lors de l'analyse des phénomènes démographiques, tels que la mortalité, les données sont représentées à l'aide d'un diagramme de Lexis. Cet outil introduit par LEXIS au XIX^{ème} siècle permet de représenter simultanément la durée de vie des individus en fonction de l'âge, du temps et de la cohorte, c'est-à-dire de leur année de naissance. Chaque individu est représenté par une ligne dans ce diagramme avec l'âge sur l'axe vertical et le temps sur l'axe horizontal.

- Les lignes verticales représentent les années.
- Les lignes horizontales représentent les âges.
- Les lignes obliques montrent l'évolution dans le temps d'une génération donnée, ou ce qu'on appelle une «cohorte».

Plusieurs visualisations peuvent être réalisées, notamment :

- Les décès à l'âge x et à l'année t .
- Les décès de la génération $t - x$ à l'âge x .
- Les décès de la génération $t - x$ au temps t .
- Les décès de la génération $t - x$ au temps t et à l'âge x .

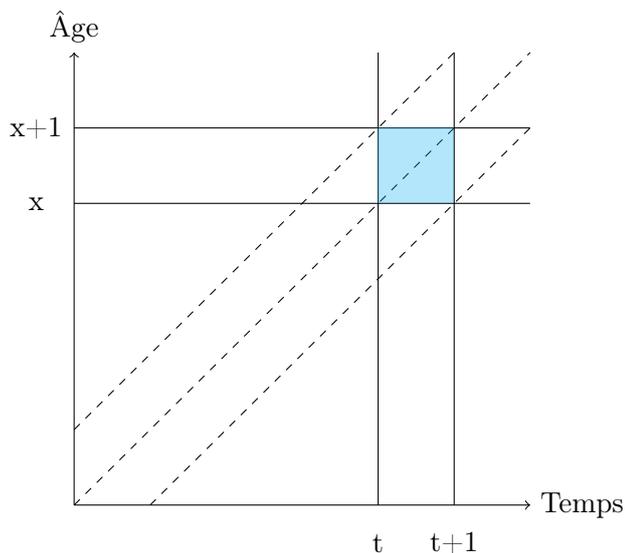


FIGURE 2.1 – Le diagramme de Lexis - Décès à l'âge x et à l'année t

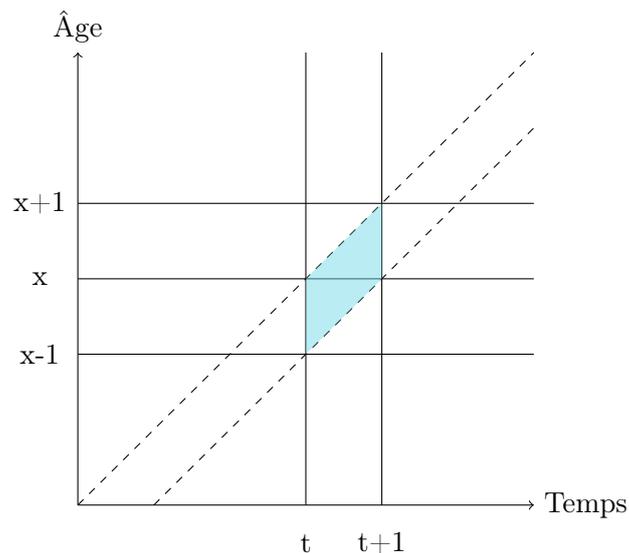


FIGURE 2.2 – Le diagramme de Lexis - Décès de la génération $t - x$ à l'âge x

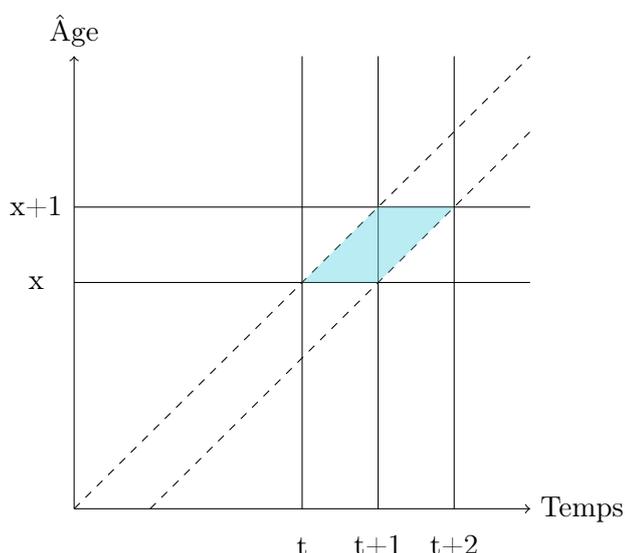


FIGURE 2.3 – Le diagramme de Lexis - Décès de la génération $t - x$ au temps t

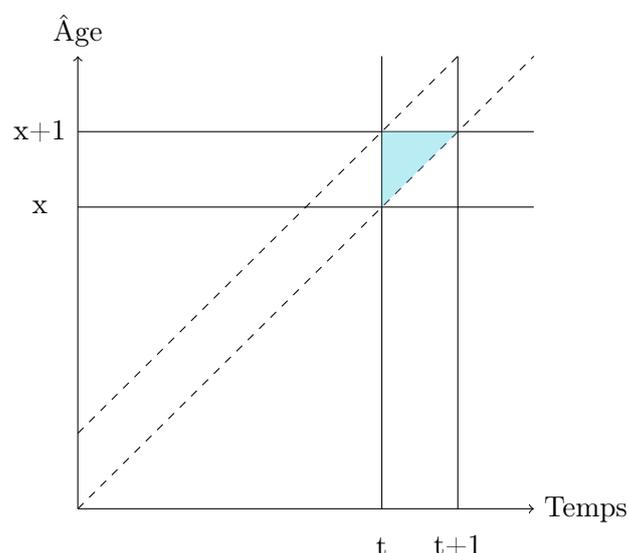


FIGURE 2.4 – Le diagramme de Lexis - Décès de la génération $t - x$ à l'âge x et au temps t

2.3 La prise en compte des données manquantes ou incomplètes

Dans les études de mortalité, deux concepts fondamentaux influencent l'interprétation et le traitement statistique des données : la censure et la troncature. Les données, souvent représentées par le temps jusqu'à un évènement, le décès dans le cas de ce mémoire, peuvent être complexes à analyser en raison de la présence de données manquantes ou incomplètes. La censure se produit lorsque l'évènement d'intérêt n'est pas observé pour certains individus, tandis que la troncature concerne l'exclusion de certaines observations. Lors de calculs actuariels, comme la construction de tables de mortalité, il est particulièrement courant d'être en présence de données censurées à droite et tronquées à gauche :

- **La censure à droite** : Les données sont dites censurées à droite si la variable d'intérêt X est telle que $X \geq C$, avec $C > 0$ l'observation de la censure fixée. En d'autres termes, il y a censure à droite si les individus ne subissent pas l'évènement sur la période d'observation pour les raisons suivantes :

- Appelés "exclus vivants", les individus sont toujours vivants au moment de l'extraction de la base. Par exemple, si une étude de mortalité suit des patients pendant 5 ans, mais que certains patients sont encore en vie à la fin de l'étude, leur temps de survie est censuré à droite.
- Les individus sont perdus de vue au cours de la période d'observation sans que l'évènement n'ait lieu. Par exemple, si les individus ont résilié leurs contrats.

Mathématiquement, les variables suivantes sont observées :

$$T_i = X_i \wedge C \quad \text{et} \quad D_i = \begin{cases} 1 & \text{si } X_i \leq C \\ 0 & \text{si } X_i > C \end{cases} \quad \forall i$$

- **La troncature à gauche :** Les données sont tronquées à gauche lorsque la variable d'intérêt n'est pas observable en dessous d'un certain seuil $c > 0$. Cela signifie que l'inclusion d'un individu dans l'étude est retardée, et les données sont collectées uniquement si l'évènement d'intérêt ne s'est pas produit avant le début du suivi. Mathématiquement, la troncature à gauche peut être formulée ainsi : Soit Z une variable aléatoire indépendante de X , il y a troncature à gauche lorsque X n'est observable que si $X > Z$. Dans ce cas, le couple (X, Z) est observé sous la condition $X > Z$. Par exemple, dans une étude de mortalité où l'âge minimum d'inclusion est de 18 ans, les individus qui décèdent avant cet âge ne sont pas pris en compte.

Exemple : Une étude sur la survie des patients diabétiques dans un hôpital suit 2 709 patients diagnostiqués entre 2000 et 2018. L'évènement d'intérêt est le décès des patients, mesuré depuis le diagnostic jusqu'à la fin de l'étude, le 31 décembre 2020. Parmi ces patients, 707 sont décédés, soit 26% de l'échantillon.

- Les 2 002 patients restants sont censurés à droite, soit parce qu'ils ont quitté l'étude, soit parce qu'ils étaient encore en vie au 31 décembre 2020.
- Les patients n'ont pas tous été suivis immédiatement après leur diagnostic de diabète, ce qui introduit une troncature à gauche, car seuls ceux ayant survécu jusqu'à leur première consultation sont inclus dans l'étude.

2.4 La revue de la littérature des modèles de construction

Après une exploration de l'histoire des tables de mortalité dans le chapitre précédent, il est intéressant d'examiner les méthodologies de modélisation des taux de mortalité. Divers modèles ont été développés afin de capturer les dynamiques complexes liées à la mortalité :

- **Les modèles paramétriques :** Ces modèles supposent que les taux de mortalité suivent une distribution spécifique. En assurance, les modèles tels que EXPONENTIEL, WEIBULL, GAMMA, GOMPERTZ-MAKEHAM et THATCHER sont utilisés. Certains modèles paramétriques intègrent également une dimension temporelle. Ces modèles sont dits prospectifs. Ils permettent de projeter l'évolution future de la mortalité à partir des tendances observées dans les données historiques. Parmi ces modèles figurent les modèles de LEE & CARTER, LOG POISSON et RENSHAW & HABERMAN.
- **Les modèles non paramétriques :** Ne font aucune hypothèse préalable sur la forme de la fonction de survie. Les plus utilisés sont les modèles de NELSON-AALEN et KAPLAN-MEIER.
- **Les modèles semi-paramétriques :** Ces modèles, tels que COX et BRASS, combinent des composantes paramétriques et non paramétriques, offrant un bon équilibre entre flexibilité et précision.



Ce qu'il faut retenir du chapitre :

- * La durée de vie humaine est décrite à l'aide de plusieurs outils mathématiques, tels que la fonction de survie, le quotient de mortalité et la fonction de risque instantané. Ces concepts fondamentaux permettent de quantifier la probabilité de survie et les risques associés à différents âges.
- * Pour mieux visualiser et analyser ces dynamiques, le diagramme de Lexis est introduit comme un outil visuel clé, illustrant la relation entre l'âge, le temps, et les générations.
- * La gestion des données manquantes ou incomplètes est particulièrement cruciale dans les études de mortalité, car elle permet d'obtenir des estimations plus précises et fiables. Pour cela, des concepts, tels que la censure et la troncature, sont explorés.
- * Ce chapitre propose une revue des principaux modèles de construction des taux de mortalité, mettant en lumière la diversité des approches disponibles pour modéliser ces taux en fonction des données disponibles et des objectifs de l'analyse.

Ce cadre théorique permet de mieux appréhender les outils et les méthodes utilisés pour l'analyse de la mortalité. Le prochain chapitre explorera les motivations et les enjeux de ce mémoire.

Chapitre 3

Les enjeux de l'étude

Basées sur des données historiques, les tables de mortalité capturent les dynamiques de mortalité d'une population, particulièrement dans un cadre prospectif où l'extrapolation des taux de mortalité permet d'analyser les tendances futures. Ces dernières années, des changements ont été observés dans la composante temporelle des modèles prospectifs. Ceux-ci seront explorés en détail dans les parties suivantes. L'objectif de ce chapitre est de réfléchir aux facteurs influençant la mortalité, de se demander si les changements observés dans ces facteurs sont structurels, et s'ils impactent ainsi durablement la dynamique de l'espérance de vie.

3.1 L'espérance de vie en France

La définition

L'espérance de vie est un indicateur essentiel pour évaluer la santé globale et le bien-être d'une population. Elle reflète l'impact des conditions de vie, des soins de santé, des politiques publiques et d'autres facteurs socio-économiques sur la longévité des individus. Ce concept, central en démographie et en actuariat, permet d'évaluer le nombre moyen d'années qu'une personne peut espérer vivre à partir d'un âge donné. Dans un cadre prospectif, l'espérance de vie est particulièrement importante car elle permet de projeter les dynamiques de mortalité actuelles dans le futur pour les générations présentes et à venir. Elle est liée à la notion de survie dans une population et est calculée à partir des probabilités de mortalité associées à chaque âge. L'espérance de vie se décline sous plusieurs formes, selon le contexte.

- **L'espérance de vie périodique** : Cet indicateur repose sur l'hypothèse que ces taux de mortalité resteront constants dans le futur, permettant d'évaluer la longévité d'une population à un moment précis dans le temps. L'espérance de vie à un âge donné, souvent appelée espérance de vie résiduelle (sur ω années), représente le nombre moyen d'années qu'une personne ayant atteint un certain âge peut encore espérer vivre (sur ω années), sous les mêmes conditions de mortalité. Elle s'exprime par :

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

avec l'indice \uparrow signifiant le déplacement le long d'une ligne verticale du diagramme de Lexis.

L'espérance de vie à la naissance, représentant le nombre moyen d'années qu'une personne peut espérer vivre, sous les mêmes conditions de mortalité, s'en déduit :

$$e_{0,t}^{\uparrow} = \sum_{k=1}^{+\infty} \prod_{j=0}^{k-1} (1 - q_{j,t})$$

Cette expression est l'espérance mathématique de la durée de vie T :

$$\begin{aligned}
 e_{0,t}^{\uparrow} &= \sum_{k=1}^{+\infty} \prod_{j=0}^{k-1} (1 - q_{j,t}) = \sum_{k=1}^{+\infty} \prod_{j=0}^{k-1} p_{j,t} \\
 &= \sum_{k=1}^{+\infty} \mathbb{P}(T \geq k) \\
 &= \sum_{k=1}^{+\infty} \sum_{i=k}^{+\infty} \mathbb{P}(T = i) \\
 &= \sum_{i=1}^{+\infty} i \mathbb{P}(T = i) \\
 &= E[T]
 \end{aligned}$$

- **L'espérance de vie générationnelle** : Cet indicateur mesure la durée de vie moyenne d'une cohorte d'individus, en tenant compte des variations de mortalité tout au long de leur existence. Contrairement à l'espérance de vie périodique, qui se base sur les taux de mortalité observés à un moment précis, l'espérance de vie générationnelle prend en compte les changements futurs des conditions de mortalité avec des taux de mortalité extrapolés. En supposant que les forces de mortalité restent constantes entre les âges entiers x et les années t , l'espérance de vie générationnelle partielle (sur ω années) s'exprime par :

$${}^{\nearrow} e_{x,t}^{\omega} = \sum_{k=1}^{\omega} \prod_{j=0}^{k-1} (1 - q_{x+j,t+j})$$

avec l'indice \nearrow signifiant le déplacement le long d'une diagonale du diagramme de Lexis.

L'évolution historique

Marquée par des fluctuations et un progrès constant, l'espérance de vie a évolué significativement au cours des siècles :

- Au milieu du $XVIII^{\text{ème}}$ siècle, l'espérance de vie à la naissance était extrêmement basse, ne dépassant pas 25 ans, principalement en raison des taux élevés de mortalité infantile et des maladies infectieuses. À cette époque, près d'un enfant sur deux mourait avant l'âge de 10 ans. Ce chiffre a légèrement augmenté pour atteindre 30 ans à la fin du siècle, dû notamment à des avancées comme la vaccination contre la variole.
- Le $XIX^{\text{ème}}$ siècle a vu une augmentation plus lente de l'espérance de vie, passant à 45 ans en 1900. Cependant, des périodes de régression ont été observées, notamment pendant les guerres napoléoniennes et la guerre franco-prussienne de 1870, où l'espérance de vie est retombée en dessous de 30 ans. Les progrès médicaux et sanitaires, bien que limités à cette époque, ont commencé à influencer positivement sur la durée de vie moyenne.
- Le $XX^{\text{ème}}$ siècle a marqué un tournant majeur avec des gains rapides en espérance de vie, malgré les perturbations liées aux deux guerres mondiales. La mortalité infantile a drastiquement diminué : alors qu'en 1900, 15% des enfants mouraient avant leur premier anniversaire, ce chiffre est passé à 5% en 1950, et à seulement 0,4% en 2015. De plus, pour les âges plus avancés, les progrès dans la lutte contre les maladies cardiovasculaires et les cancers ont joué un rôle crucial dans l'amélioration continue de l'espérance de vie.

L'évolution de l'espérance de vie à la naissance en France de 1740 à 2020 est présentée dans le graphique ci-dessous. L'augmentation générale est observée ainsi que les pics de décroissance affectés aux évènements particuliers.

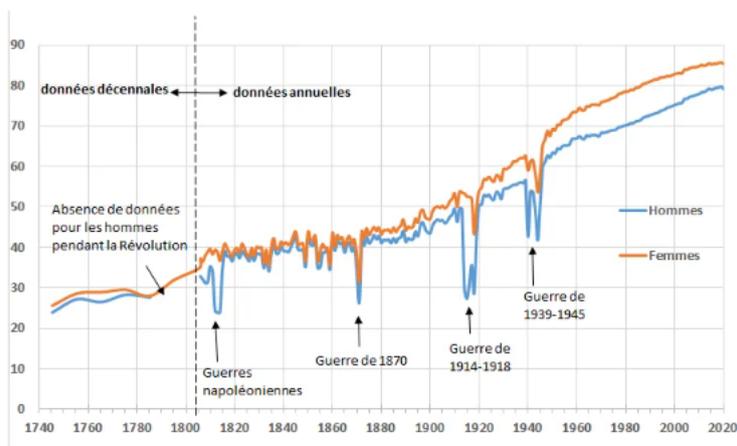


FIGURE 3.1 – L'évolution de l'espérance de vie à la naissance en France de 1740 à 2020

Source : Institut National d'Études Démographiques (INED) - L'espérance de vie en France

Le ralentissement de l'amélioration de l'espérance de vie

Les dernières décennies ont été marquées par une augmentation constante de l'espérance de vie, même si le rythme de cette augmentation a quelque peu diminué. En 2019, Gilles PISON s'interrogeait sur un possible plafonnement de l'espérance de vie. Dans son article "L'espérance de vie a-t-elle atteint ses limites en France?", il met en lumière plusieurs indicateurs inquiétants. Il note que l'espérance de vie à la naissance avait progressé de manière régulière tout au long du $XX^{\text{ème}}$ siècle, avec un gain moyen de trois mois par an. Cependant, depuis 2014, cette progression s'est ralentie, avec une augmentation de seulement un mois et demi par an pour les hommes et un mois pour les femmes. Ce ralentissement soulève des questions sur les causes sous-jacentes. Est-il temporaire ou représente-t-il une nouvelle tendance de fond? A-t-on atteint un plafond? L'espérance de vie peut-elle diminuer?

Depuis 2019, ces préoccupations se sont intensifiées avec un ralentissement marqué de l'espérance de vie, mis en avant par les graphiques suivants. Que ce soit pour les hommes ou pour les femmes, l'espérance de vie à la naissance a brusquement chuté en 2019, puis a stagné entre 2020 et 2022. Le gain d'espérance de vie à chaque âge est également moins important entre 2010 et 2023, qu'entre 1997 et 2010.

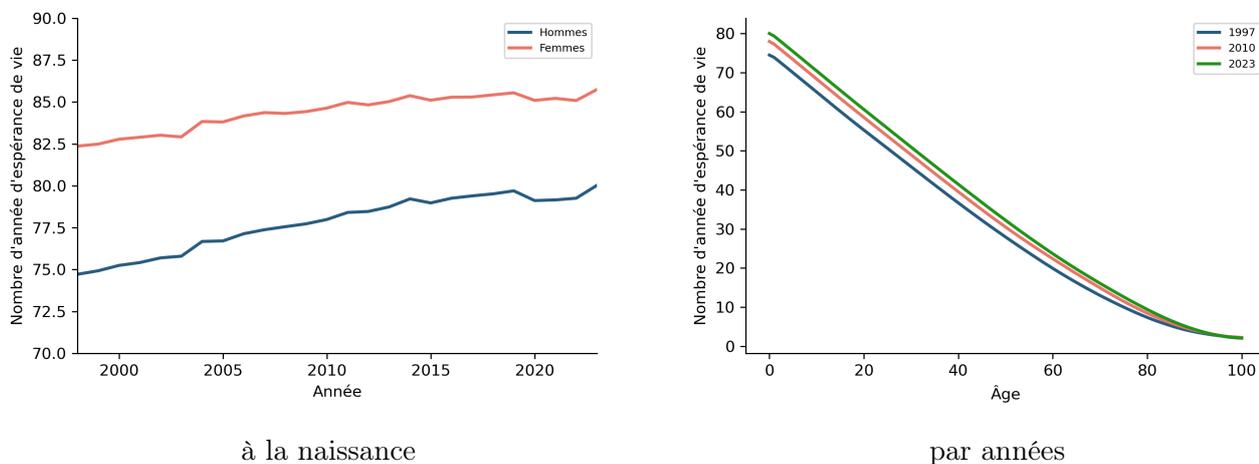


FIGURE 3.2 – L'espérance de vie

Source : Institut National de la Statistique et des Études Économiques (INSEE)

Un autre point à considérer est la convergence des espérances de vie entre les sexes. Historiquement, les femmes ont eu une espérance de vie plus élevée que les hommes, mais cet écart se réduit progressivement. En 1982, les femmes avaient plus de 8 ans d'espérance de vie comparativement aux hommes. Cet écart s'est réduit à près de 7 ans en 2002 pour atteindre en 2018 moins de 6 ans d'écart. Depuis 2019, une bosse est observée, ce changement amène à des questionnements sur un possible changement de dynamique de mortalité.

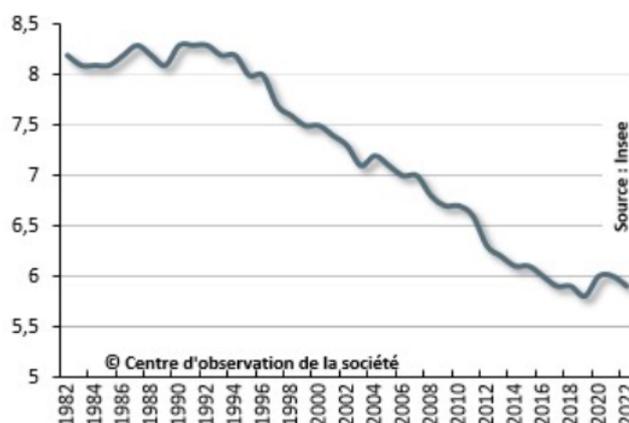


FIGURE 3.3 – L'écart d'espérance de vie entre les femmes et les hommes en années

Source : Institut National de la Statistique et des Études Économiques (INSEE)

3.2 Les facteurs à prendre en compte dans l'analyse des tendances de mortalité dans le temps

Une multitude de facteurs influencent la mortalité d'une population. Ils peuvent être de nature différentes : démographiques, socio-économiques, environnementales, comportementales, etc. Chacun de ces éléments influence de manière complexe la mortalité.

Les facteurs environnementaux :

— Les conditions climatiques extrêmes :

Les vagues de chaleur et de froid extrêmes provoquent une surmortalité importante, notamment chez les populations vulnérables. Depuis la canicule de 2003, qui a causé 15 000 décès en France, ces événements climatiques se sont intensifiés et répétés, comme en 2019, 2022, et 2023. Ces années ont été marquées par des canicules prolongées et des températures records, accentuant les risques sanitaires. La tendance à la hausse des épisodes de chaleur se poursuit en 2024, impactant la mortalité.

— La qualité de l'air :

La qualité de l'air est un facteur environnemental majeur influençant la santé des populations, en particulier dans les zones urbaines où les niveaux de pollution dépassent souvent les seuils recommandés par les normes européennes et l'Organisation Mondiale de la Santé (OMS). La pollution atmosphérique, notamment les particules fines (PM2.5) et le dioxyde d'azote (NO2), a des effets néfastes sur la santé humaine, entraînant notamment des maladies cardiovasculaires, des maladies respiratoires, et un risque accru de cancer du poumon.

Les facteurs sanitaires et épidémiologiques :

— Les maladies infectieuses et pandémies :

La pandémie de COVID-19 a eu un impact majeur sur la mortalité en France, avec un pic important en 2020 et une surmortalité prolongée dans les années suivantes. En 2020, une augmentation de 11% des décès a été observée dans l'Union européenne. Entre 2020 et 2021, les décès ont largement dépassé les attentes, notamment chez les personnes de plus de 75 ans. En 2022, bien que l'épidémie ait été moins meurtrière, les décès sont restés au-dessus des niveaux anticipés.

Les épidémies de grippe hivernale contribuent également à la surmortalité. Les hivers 2016-2017 et 2017-2018 ont causé environ 14 000 décès par an, et l'épidémie de 2018-2019 a entraîné environ 12 000 décès supplémentaires.

— Les maladies chroniques :

Les maladies chroniques, telles que les cancers, les maladies cardiovasculaires, l'insuffisance cardiaque et les maladies respiratoires, sont parmi les principales causes de décès en France. En 2023, les cancers demeurent la première cause de décès chez les hommes et la deuxième chez les femmes. Les maladies cardiovasculaires, premières causes de décès chez les femmes, ainsi que les maladies respiratoires, continuent d'avoir un impact important. Bien que mieux maîtrisée, l'insuffisance cardiaque reste également une cause importante de décès. Ces maladies sont souvent aggravées par des facteurs tels que l'obésité et la sédentarité.

Les facteurs liés aux habitudes de vie :

Les habitudes de vie jouent un rôle déterminant dans la mortalité, en particulier en aggravant les maladies chroniques. Le tabagisme reste une des principales causes de décès évitables, responsable de nombreux cancers et maladies cardiovasculaires. La consommation de drogues illicites, bien que moins fréquente, augmente les risques de maladies graves et de décès, souvent aggravés par l'obésité et la sédentarité. Ces comportements affectent également la mortalité infantile, avec un risque accru de décès néonatal lorsque des toxiques sont consommés pendant la grossesse. Depuis 2015, la mortalité infantile en France stagne à un niveau supérieur à la moyenne européenne.

Schématisation synthétique des facteurs influençant la mortalité :

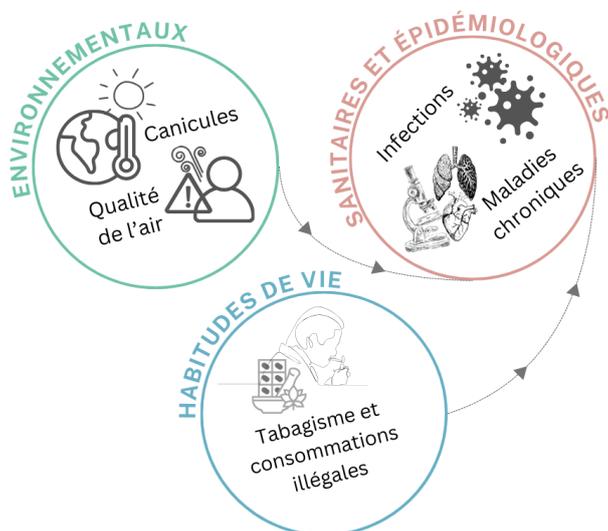


FIGURE 3.4 – Le schéma synthétique des facteurs

3.3 La problématique et les perspectives de ce mémoire

Depuis plusieurs années, l'espérance de vie en France montre des signes de ralentissement après des décennies de progression régulière. Ce phénomène suscite des interrogations quant à ses causes profondes : est-il principalement dû à des facteurs conjoncturels, ou révèle-t-il un changement structurel plus profond ? Si ce ralentissement est effectivement d'origine structurelle, il est crucial de comprendre quelles en seraient les implications à long terme pour la mortalité des populations.

Cette étude se propose d'explorer différents scénarios en matière de mortalité et ainsi construire des tables de référence permettant de mieux anticiper les défis futurs.

Plusieurs axes de recherche sont proposés :

- **La construction de nouvelles tables de mortalité intégrant des données récentes :** L'utilisation de données relatives aux dernières années permettra de tenir compte du changement dans les dynamiques de mortalité.
- **La modélisation alternative de la composante temporelle pour la projection des taux de mortalité :** L'utilisation de modèles alternatifs dans la projection de la composante temporelle, efficaces dans un contexte de non-linéarité, permet de tenir compte des tendances historiques tout en intégrant les dernières années.

En plus d'analyser l'impact sur les taux de mortalité, ce mémoire examinera les conséquences de l'utilisation de ces nouvelles tables de référence sur le provisionnement d'un régime de retraite supplémentaire.

Schématisation synthétique des objectifs du mémoire :

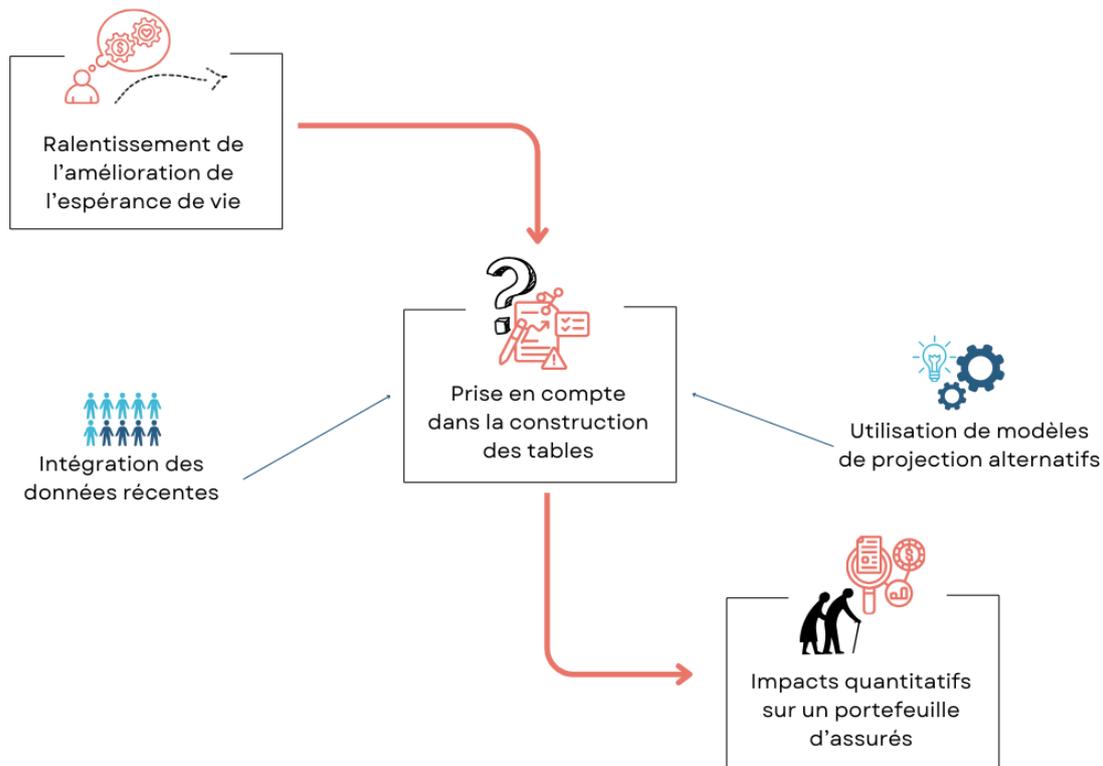


FIGURE 3.5 – Le schéma synthétique des objectifs du mémoire



Ce qu'il faut retenir du chapitre :

- * L'espérance de vie en France a connu une évolution significative au fil des siècles. Cependant, ces dernières années sont marquées par un changement sur les tendances de mortalité.
- * Les dynamiques de mortalité en France sont influencées par une combinaison de facteurs conjoncturels et structurels. Ceux-ci peuvent modifier de manière complexe les tendances de mortalité.
- * La problématique centrale de ce mémoire consiste à élaborer différents scénarios de mortalité pour prendre en compte les points "atypiques" observés ces dernières années et d'en évaluer les impacts dans les projections futures.

Le prochain chapitre présentera les données nécessaires à cette étude ainsi qu'une première analyse de celles-ci.

Chapitre 4

L'étude des données de mortalité utilisées en vue d'une construction prospective

L'analyse des données de mortalité est une étape cruciale pour la construction de tables de mortalité prospectives, afin qu'elles soient précises et pertinentes. Une compréhension approfondie et une préparation rigoureuse de ces données sont essentielles pour garantir la cohérence et la fiabilité des tables produites. Ce chapitre vise à présenter les données utilisées dans ce mémoire, à décrire leur traitement, et à réaliser les premières analyses qui en découlent.

4.1 La présentation des bases de données

4.1.1 Les données de population française

Les données de population utilisées dans cette étude proviennent de l'Institut National de la Statistique et des Études Économiques (INSEE)¹. Elles contiennent le nombre d'individus au 1^{er} janvier de chaque année, par âge de 0 à 99 ans et plus et par sexe. Les données couvrant la période de 1991 à 2020 sont issues des recensements de la population. Les données des années 2021, 2022 et 2023 sont des estimations de l'INSEE. Ces estimations reposent sur les statistiques de l'état civil et des projections du solde migratoire.

Ces données fournissent une base essentielle pour la construction de tables de mortalité précises, reflétant les évolutions récentes de la population française. Le nombre d'individus a considérablement augmenté au fil des années, pour être proche de 68 millions en 2023. Cette population a tendance à vieillir avec un âge moyen légèrement supérieur à 36 ans en 1991 passant à près de 42 ans en 2023.

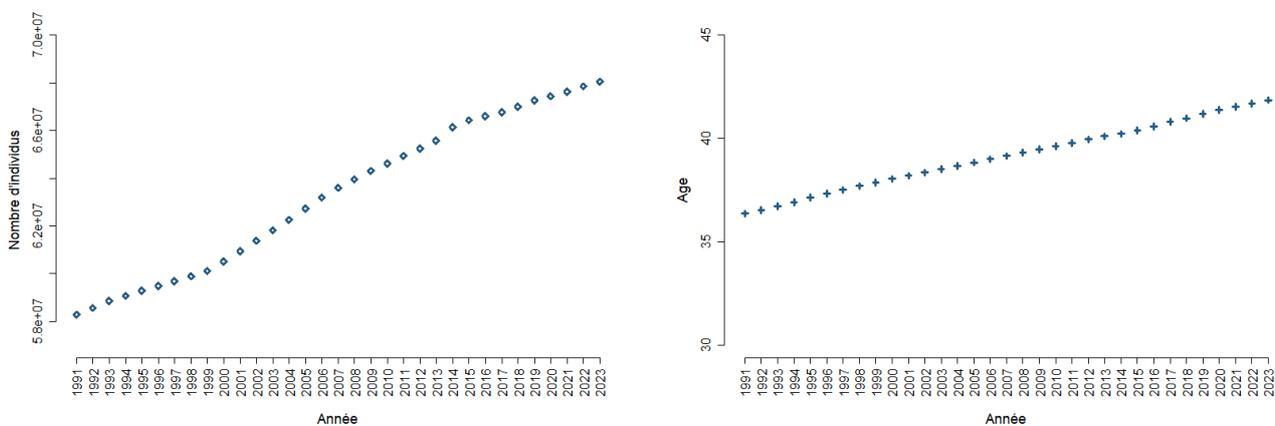


FIGURE 4.1 – Le nombre d'individus et l'âge moyen des individus par année

1. Certaines années sont illustrées en annexe.

4.1.2 Les données de décès

Les données de décès utilisées dans cette étude proviennent des fichiers décennaux des personnes décédées, fournis par l'INSEE. Ces fichiers couvrent la période de 1970 à 2023 et sont établis à partir des informations reçues des communes dans le cadre de leur mission de service public. Pour chaque décès enregistré, les informations suivantes sont fournies : nom, prénoms, sexe, date de naissance, code et libellé du lieu de naissance, date du décès, code du lieu de décès et numéro de l'acte de décès.

De 1991 à 2010, le nombre de décès était proche de 555 000 tous âges confondus. Ce nombre a eu tendance à augmenter depuis 2010. L'âge moyen de ces décès est de plus en plus élevé, passant de près de 73 ans en 1991 à plus de 79 ans en 2023.

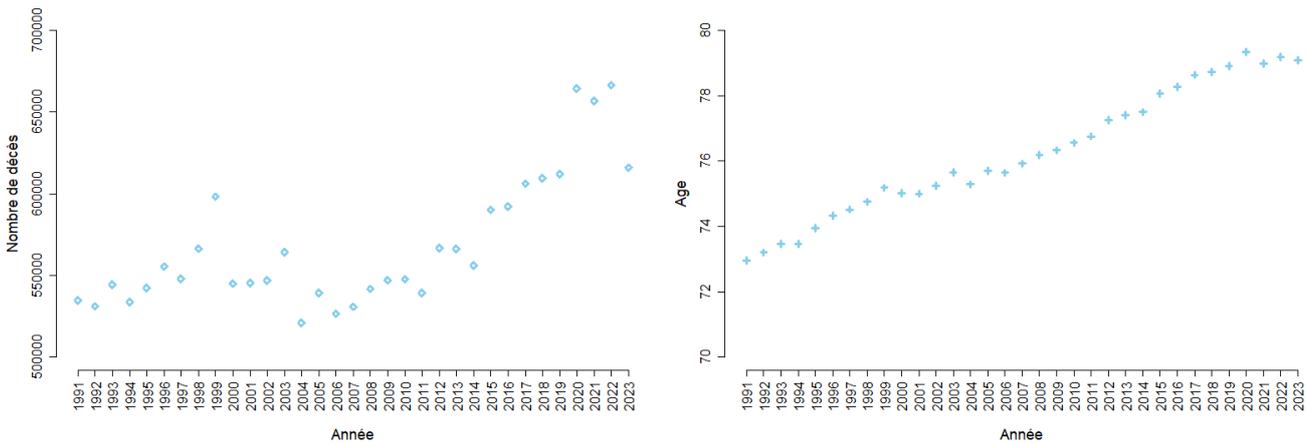


FIGURE 4.2 – Le nombre de décès et l'âge moyen des décès par année

4.2 Le traitement des données

4.2.1 La mise en forme des données

Afin de mener à bien l'analyse de la mortalité et la construction de tables prospectives, une structuration des données pour les rendre cohérentes et directement exploitables a été effectuée.

- **Les données de population** : Les données de population initialement disponibles étaient structurées sous forme d'un tableau comportant quatre colonnes : l'année, le sexe, l'âge et le nombre d'individus. Elles ont été réorganisées pour créer deux matrices distinctes, une pour les femmes et une pour les hommes. Chaque matrice a pour colonnes les années calendaires (de 1991 à 2023) et pour lignes les âges de 0 à 99 ans.
- **Les données de décès** : Les données de décès ont nécessité un traitement spécifique pour être alignées sur la structure des données de population. Ces données étaient initialement fournies sous forme de fichiers annuels, chaque fichier listant les décès ligne par ligne. Pour chaque année, les décès ont été comptabilisés par sexe et par âge de décès. Afin de garantir une cohérence avec les données de population, tous les individus décédés au-delà de 99 ans ont été regroupés dans la catégorie des décès à 99 ans.

Ces traitements ont ainsi permis de disposer de quatre matrices distinctes :

- Une matrice contenant le nombre de femmes avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).
- Une matrice contenant le nombre d'hommes avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).
- Une matrice contenant le nombre de femmes décédées avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).
- Une matrice contenant le nombre d'hommes décédés avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).

4.2.2 Le calcul des taux de mortalité bruts

Maintenant que les données ont été uniformisées et structurées, il est possible de procéder au calcul des taux de mortalité bruts.

L'estimateur binomial

L'estimateur binomial est un outil d'estimation des taux de mortalité bruts. Cette méthode repose sur l'hypothèse que les décès peuvent être modélisés par une loi binomiale.

- $N_{x,t}$ est le nombre d'individus vivants à l'âge x de l'année t .
- La variable aléatoire $D_{x,t}$ modélise le nombre de décès dans l'intervalle $[x, x + 1]$ de l'année t . La valeur observée de $D_{x,t}$ est notée $d_{x,t}$.
- Tous les décès sont supposés indépendants les uns des autres et chaque décès à l'âge x et à l'année t , a une probabilité $q_{x,t}$ de se produire.
- Deux issues sont possibles sur l'intervalle $[x, x + 1]$: vivre ou décéder. La variable aléatoire $X_{i,t}$ qui suit une loi de Bernoulli de paramètre $q_{x,t}$ est alors posée.

Les individus étant indépendants, le nombre de décès sur $[x, x + 1]$ de l'année t , $D_{x,t} = \sum_{i=1}^{N_{x,t}} X_{i,t}$ suit une loi Binomiale, $\mathcal{B}(N_{x,t}, q_{x,t})$. Autrement dit,

$$P(D_{x,t} = d_{x,t}) = \binom{N_{x,t}}{d_{x,t}} q_{x,t}^{d_{x,t}} (1 - q_{x,t})^{N_{x,t} - d_{x,t}}$$

L'estimation de $q_{x,t}$ s'effectue par la méthode du maximum de vraisemblance, en dérivant la log-vraisemblance par rapport à $q_{x,t}$. L'estimateur binomial est donc le suivant :

$$\hat{q}_{x,t} = \frac{d_{x,t}}{N_{x,t}}$$

Cet estimateur suppose que les individus sont entièrement observables sur l'intervalle $[x, x + 1]$. Dans le cadre de ce mémoire, la seule sortie possible est le décès. Ainsi, l'estimateur binomial est compétent. Dans les situations où les individus ne sont pas observés durant l'ensemble de la période, c'est le cas par exemple lors d'une construction sur un portefeuille d'assurés où une des sorties possibles est le rachat, des méthodes alternatives existent. L'estimateur de HOEM², par exemple, ajuste l'exposition des individus pour tenir compte des périodes où ils ne sont pas observables.

2. Cet estimateur est présenté en annexe.

L'application de l'estimateur binomial sur les données d'étude

Dans ce mémoire, la seule sortie possible étant le décès, les individus sont entièrement observables sur l'intervalle $[x, x + 1]$. Ainsi, il est pertinent d'utiliser l'estimateur binomial pour le calcul des taux de mortalité bruts. Ces traitements ont ainsi permis de disposer de deux matrices supplémentaires, analysées dans la section suivante :

- Une matrice des taux de mortalité bruts des femmes avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).
- Une matrice des taux de mortalité bruts des hommes avec en colonnes les années calendaires (1991 à 2023) et en lignes les âges (0 à 99 ans).

Ces taux serviront par la suite à distinguer les effets sur la mortalité liés à l'âge et ceux liés à l'année calendaire, via des modèles prospectifs, permettant ainsi leur extrapolation.

4.3 L'analyse des données

4.3.1 L'examen des facteurs d'hétérogénéité

Les chapitres précédents ont mis en évidence que le sexe est un déterminant clé de la mortalité. Les données de population étant différenciées uniquement par sexe, l'analyse se focalise exclusivement sur ce facteur.

L'analyse de la mortalité révèle des différences significatives entre les hommes et les femmes. Les graphiques des logarithmes des taux de mortalité pour deux années distinctes montrent que les hommes présentent des taux de mortalité plus élevés que les femmes pour la majorité des âges.

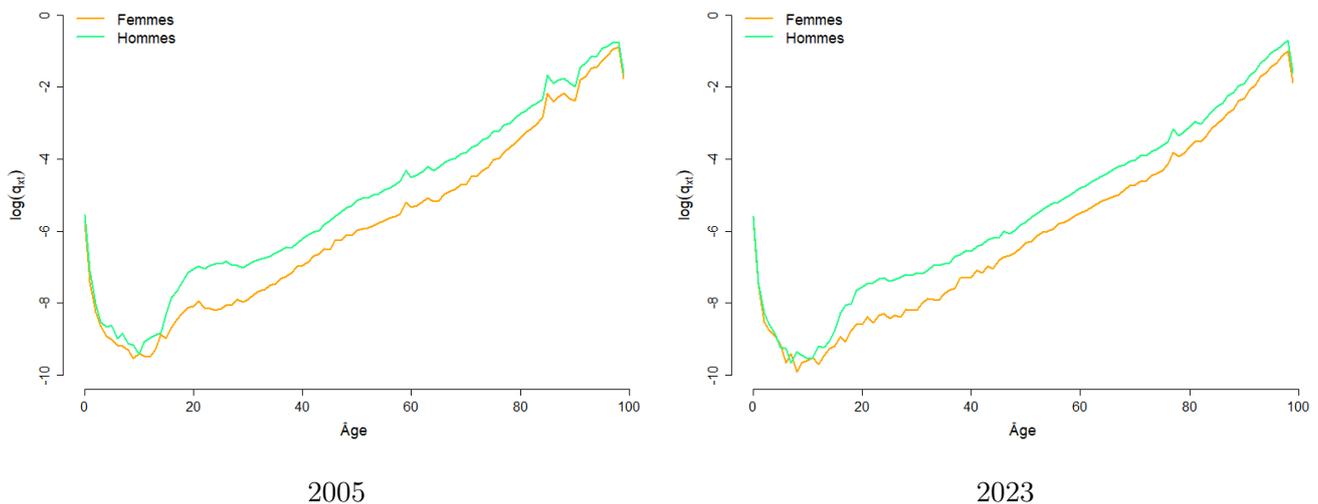


FIGURE 4.3 – Les logarithmes des taux de mortalité bruts selon le sexe

En plus de l'analyse des taux de mortalité, l'examen des sex-ratios, dans la population et parmi les décès, fournit des informations complémentaires essentielles. Globalement, les femmes sont légèrement plus nombreuses que les hommes dans la population, représentant 51,53% du total. Cependant, lorsqu'on examine les décès, la tendance s'inverse légèrement, les femmes représentant 49,46% des décès.

Bien que l'écart de mortalité entre les sexes se soit réduit au fil des années, des différences persistent. Il est donc justifié de continuer à construire des tables de mortalité distinctes pour les hommes et les femmes afin de refléter au mieux les réalités démographiques actuelles et les comportements différenciés en matière de mortalité.

4.3.2 L'analyse des tendances de mortalité

Pour explorer les tendances de mortalité et évaluer si elles ont significativement évolué au cours des années étudiées, les taux de mortalité bruts par âge et par année sont visualisés sur un graphique en 3 dimensions. Cette analyse couvre la période de 1991 à 2023 pour les âges de 0 à 99 ans.

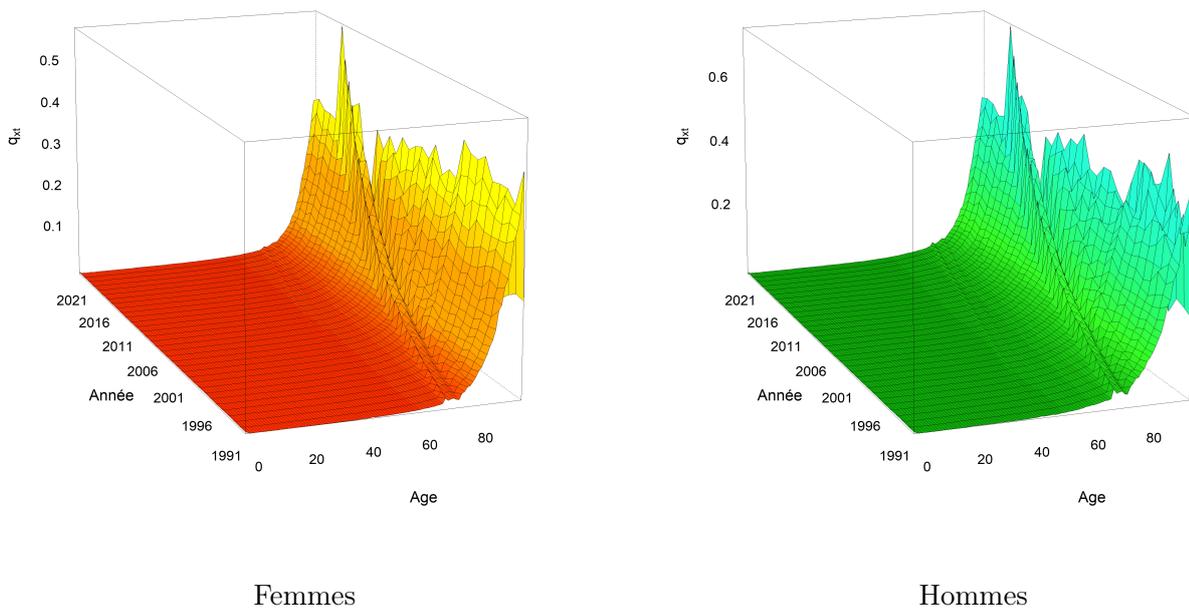


FIGURE 4.4 – Les taux de mortalité bruts

Sur ces surfaces, certaines tendances émergent. Que ce soit pour les hommes ou pour les femmes, les taux de mortalité augmentent avec l'âge et diminuent avec les temps. Des sauts sont cependant observés en partie en raison de l'utilisation des taux bruts non ajustés. Ces sauts peuvent être problématiques, car ils ne reflètent pas nécessairement une véritable variation de la mortalité, mais plutôt des fluctuations dues à la variabilité des données ou à des limitations du modèle. En principe, les taux de mortalité devraient suivre une progression quasi linéaire en fonction de l'âge, sans rupture de tendance dans le temps.

Des graphiques en deux dimensions, avec le temps fixé, permettent une analyse plus approfondie des taux de mortalité et de leurs logarithmes en fonction de l'âge selon les années t .

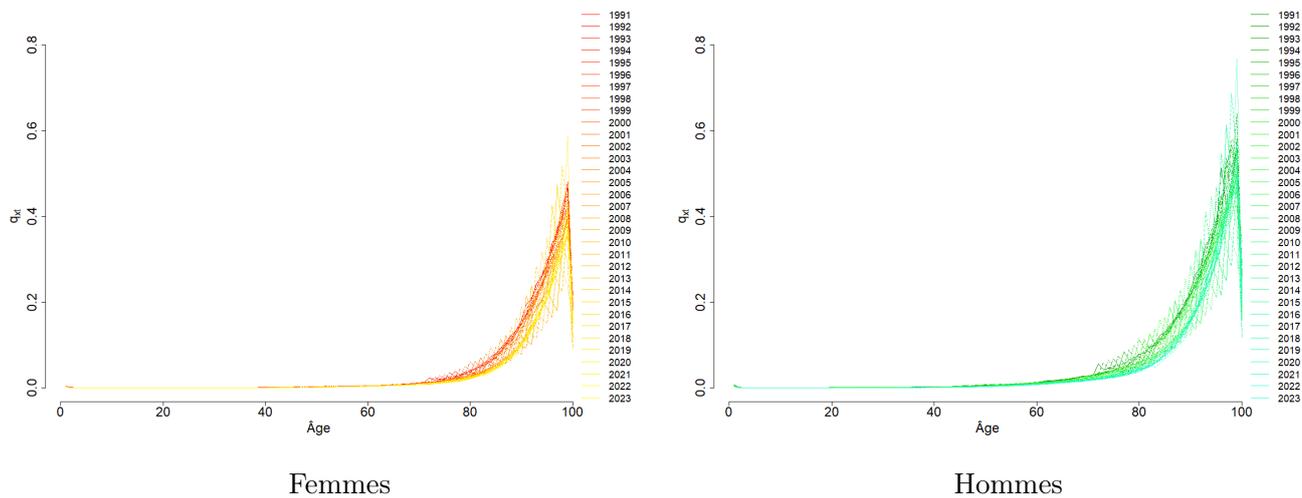


FIGURE 4.5 – Les taux de mortalité bruts selon l'âge par année

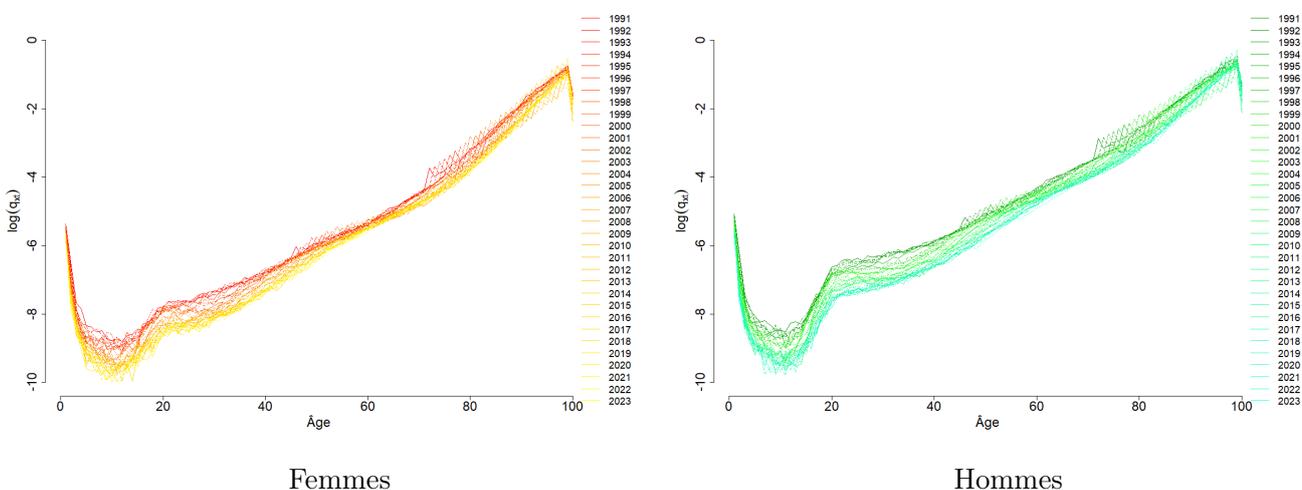


FIGURE 4.6 – Les logarithmes des taux de mortalité bruts selon l'âge par année

L'effet du temps

Les graphiques révèlent une dynamique à la baisse des taux de mortalité au fil du temps. Un décalage des courbes d'une année à l'autre est observé, avec des courbes plus récentes présentant des taux de mortalité inférieurs pour les mêmes âges, aussi bien chez les femmes que chez les hommes.

L'effet de l'âge

Les graphiques révèlent plusieurs tendances distinctes dans les profils de mortalité, tant pour les hommes que pour les femmes. En particulier, l'analyse met en évidence des phases importantes comme la mortalité infantile, la bosse des accidents, et l'augmentation exponentielle de la mortalité aux âges avancés.

- **Mortalité infantile** : Les taux de mortalité sont élevés dans la première année de vie, reflétant la vulnérabilité des nouveau-nés aux maladies, aux complications de la naissance et aux conditions environnementales. Cette phase est critique, car les nouveau-nés sont sensibles à divers risques qui diminuent significativement après les premiers mois de vie.

- **Creux chez les enfants et les jeunes adolescents :** Un creux est observé dans les taux de mortalité chez les enfants et les jeunes adolescents, entre 5 et 15 ans. Cette période est caractérisée par des taux de mortalité très bas, reflétant une phase de la vie où les risques de décès sont les plus faibles.
- **Bosse des accidents :** Ce phénomène se manifeste par une augmentation significative des taux de mortalité des jeunes adultes, entre 15 et 30 ans. Cette augmentation est principalement attribuée aux décès dus à des causes externes telles que des accidents de la route ou des overdoses. Cette hausse est plus marquée pour les hommes que pour les femmes.
- **Augmentation progressive après 30 ans :** Passé 30 ans, les taux de mortalité commencent à augmenter quasi linéairement.
- **Augmentation exponentielle aux âges avancés :** Pour les personnes âgées de 80 ans et plus, les taux de mortalité augmentent de manière exponentielle mettant en lumière les défis concernant le vieillissement de la population. Les projections démographiques récentes de l'INSEE anticipent une poursuite du phénomène de vieillissement de la population française au cours des 50 prochaines années. D'après le scénario central, jugé le plus probable et basé sur le prolongement des tendances démographiques actuelles, la population française atteindrait 68,1 millions d'habitants au 1er janvier 2070, contre 67,4 millions au 1er janvier 2021, représentant une augmentation de 700 000 personnes. Un des aspects les plus marquants de ces projections est le doublement prévu du nombre de personnes âgées de 75 ans et plus d'ici 2070. La pyramide des âges présentée ci-après illustre cet aspect.

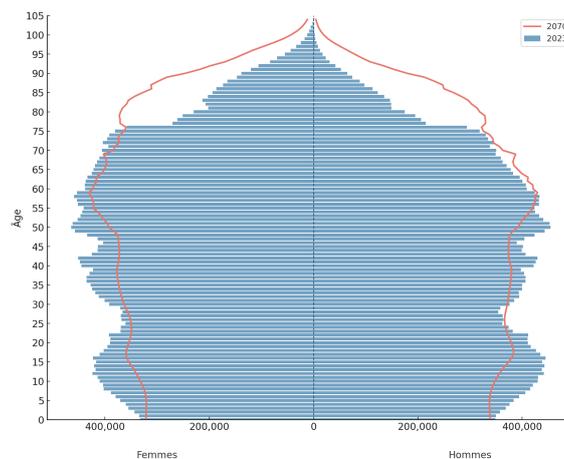


FIGURE 4.7 – Les pyramides âges - Population au 1^{er} janvier en 2023 et en 2070

Source : INSEE - Pyramides des âges - Projections de population 2021-2070 - Scénarios

Par ailleurs, une variabilité relativement élevée des taux de mortalité est notée pour certains âges avancés. Cette observation soulève la question de la pertinence des analyses aux âges les plus élevés. Dans la section suivante, il sera donc nécessaire de définir le périmètre d'étude en déterminant jusqu'à quel âge il est judicieux de poursuivre l'analyse pour garantir des résultats fiables.

L'effet du temps combiné à l'âge

L'évolution des taux de mortalité est un aspect clé de l'analyse démographique. Ce mémoire se concentre sur la projection de ces taux et l'impact des variations temporelles sur le risque de décès. L'étude des tendances par âge met en évidence les changements récents, notamment les perturbations observées depuis 2019. Comprendre ces dynamiques est essentiel pour élaborer des modèles prédictifs robustes.

- **Une tendance à la baisse** : Depuis 1991, les taux de mortalité montrent une tendance générale à la baisse, particulièrement marquée chez les enfants et les jeunes adultes. Pour les femmes, une amélioration de 58% est observée chez les 5-14 ans et de 56,5% chez les 15-30 ans entre 1991 et 2023. De même, les hommes ont vu une baisse de 63,2% pour les 5-14 ans et de 61,4% pour les 15-30 ans sur la même période. Ces chiffres illustrent les progrès significatifs dans la réduction de la mortalité, comme le montrent les graphiques suivants pour la tranche d'âge 15-30 ans chez les hommes et les femmes.



FIGURE 4.8 – Les logarithmes des taux de mortalité bruts selon l'année pour les 15-30 ans

- **Une perturbation récente** : Malgré cette tendance à la baisse des taux de mortalité observée depuis plusieurs années, les données récentes, en particulier depuis 2019, montrent une perturbation qui se traduit par une augmentation des taux de mortalité dans certaines tranches d'âge. Cette augmentation est particulièrement visible pour les femmes et les hommes dans la classe d'âge 60-75 ans. Les graphiques suivants mettent en lumière ces perturbations récentes. Ils montrent comment, après une période de baisse régulière, les taux de mortalité ont soudainement augmenté dans certaines tranches d'âge, mettant en évidence l'importance de surveiller les dynamiques de mortalité.

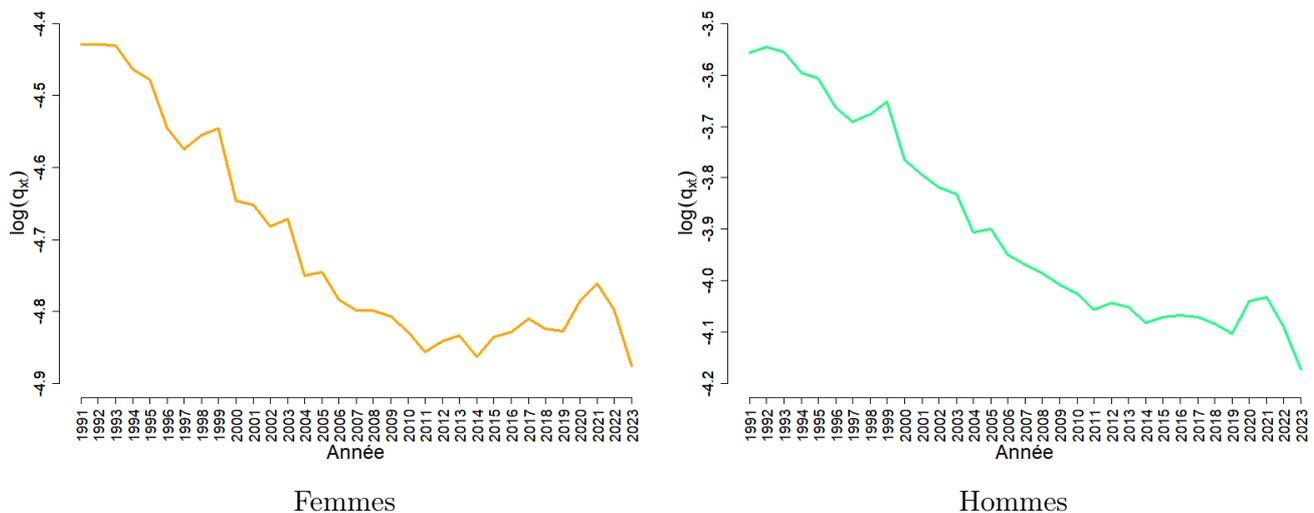


FIGURE 4.9 – Les logarithmes des taux de mortalité bruts selon l'année pour les 60-75 ans

La dynamique temporelle de la mortalité, mise en évidence par les analyses précédentes, souligne l'importance cruciale de construire des tables de mortalité prospectives. Il est essentiel de souligner que les effets du temps sur la mortalité ne sont pas homogènes à travers les différentes tranches d'âge. Cette hétérogénéité renforce l'importance d'analyser la combinaison des effets liés à l'âge et au temps dans les modèles prospectifs, afin de mieux appréhender les dynamiques de mortalité.

Pour capter avec précision les variations les plus récentes dans les tendances de la mortalité, il est important d'utiliser des données récentes. L'intégration des variations actuelles dans les modèles de projection améliore la pertinence et la fiabilité des prévisions, en tenant compte des dynamiques les plus actuelles et des nouvelles influences sur la mortalité.

Si les tendances récentes se maintiennent, il est légitime de se questionner sur l'adéquation des méthodes de projection actuelles et d'envisager de nouvelles approches capables de tenir compte de ces changements. Les modèles actuels, bien qu'efficaces dans un contexte de linéarité, pourraient ne pas capter suffisamment les nuances et les variations soudaines observées dans les données récentes.

4.3.3 Le périmètre d'étude

Précédemment, l'analyse a révélé une variabilité importante des taux de mortalité aux âges élevés, ce qui amène à se demander jusqu'à quel âge il est pertinent de construire les tables de mortalité. Pour effectuer une analyse approfondie du choix des âges pour la construction des tables de mortalité, il est nécessaire d'examiner les taux de décès et les expositions par âge. Les taux de mortalité sont fiables lorsqu'il y a un nombre suffisant d'individus exposés et de décès enregistrés. La variabilité précédemment observée aux âges élevés nécessite de fixer une limite d'âge pour garantir la robustesse des calculs. Cette analyse est effectuée séparément pour les hommes et les femmes, conformément aux différences déjà constatées.

Les graphiques suivants montrent une diminution de la population aux âges avancés, associée à une augmentation des taux de mortalité. Ainsi, pour les deux sexes, la proportion de personnes vivant jusqu'à un âge avancé diminue, tandis que les taux de mortalité augmentent significativement avec l'âge.

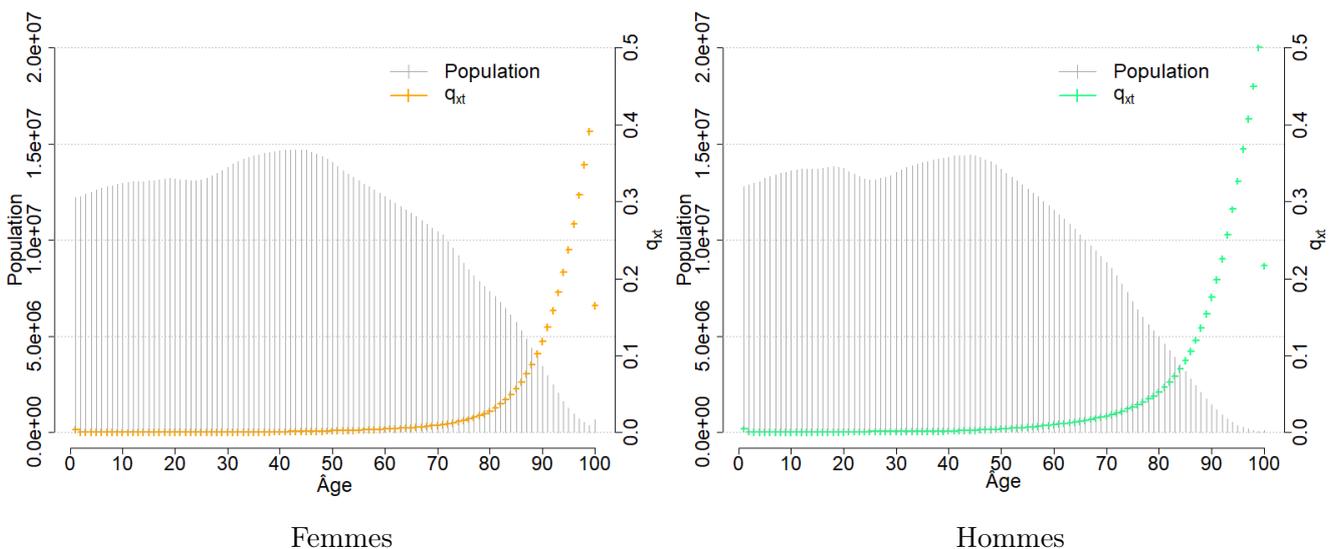


FIGURE 4.10 – Les taux de mortalité et population

En outre, les graphiques ci-dessous montrent que le nombre de décès chute brusquement après 90 ans pour les femmes et 85 ans pour les hommes. Les taux de mortalité bruts obtenus au-delà de ces âges sont très volatils et ne disposent pas de réel sens statistique. Ainsi, ces âges plafonds sont retenus pour la construction des tables prospectives, afin d'assurer des estimations robustes.

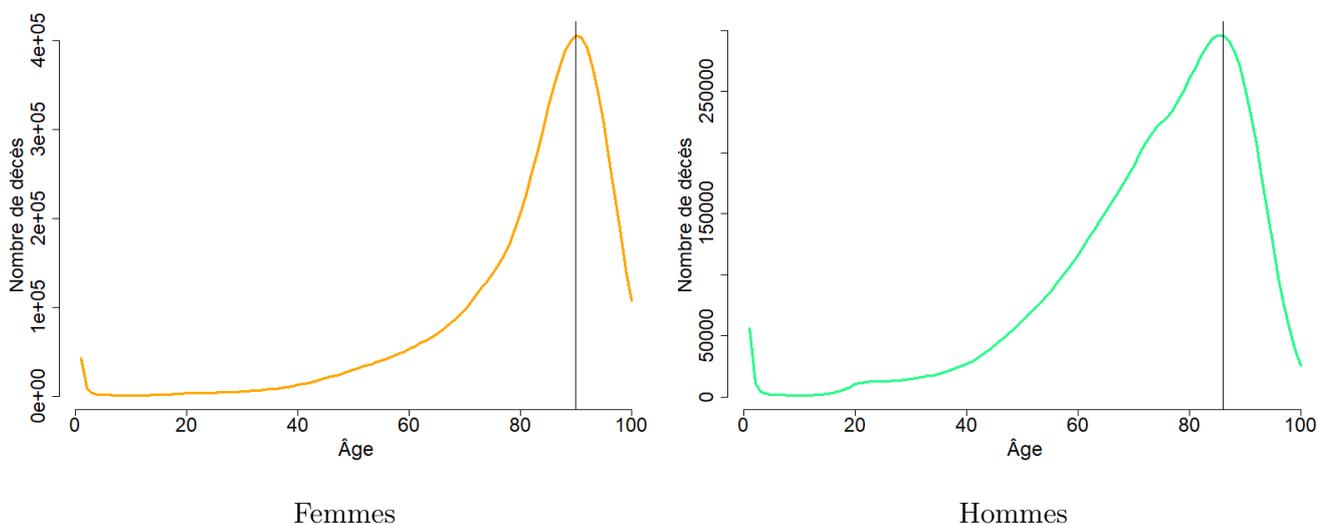


FIGURE 4.11 – Les nombres de décès



Ce qu'il faut retenir du chapitre :

- * Les données de population et de décès en France, par âge et par sexe, proviennent des bases de l'Institut National de la Statistique et des Études Économiques (INSEE).
- * Après une mise en forme des données, les taux de mortalité bruts par âge ont été calculés à l'aide de l'estimateur binomial.
- * Le sexe est identifié comme un déterminant majeur de la mortalité, justifiant l'analyse distincte pour les hommes et les femmes. Les différences de mortalité entre les sexes persistent malgré une réduction progressive de l'écart, ce qui soutient la nécessité de construire des tables de mortalité distinctes.
- * Des tendances de mortalité selon l'âge ont émergé, notamment aux âges très jeunes avec la mortalité infantile, entre 15 et 30 ans avec la "bosse des accidents" et aux âges élevés avec une mortalité exponentielle. L'évolution temporelle observée (tendance à la baisse et perturbation récente) justifie l'approche prospective dans la construction des tables de mortalité, afin de mieux anticiper et s'adapter aux changements démographiques actuels.
- * La volatilité des taux de mortalité au-delà de certains âges a conduit à fixer un âge limite pour la construction des tables de mortalité : 90 ans pour les femmes et 85 ans pour les hommes.

Schématisation synthétique du chapitre :

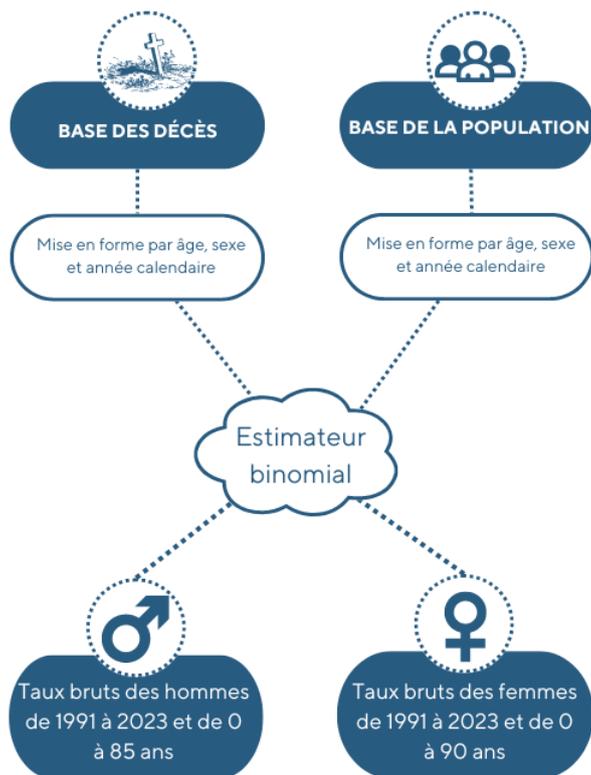


FIGURE 4.12 – Le schéma synthétique des données

Deuxième partie

Construction de tables de mortalité
prospectives

Chapitre 1

L'estimation des taux

Depuis plusieurs décennies, de nombreux chercheurs ont développé divers modèles de mortalité, répondant à des exigences d'estimation précises en termes de caractéristiques liées à l'âge et à l'année calendaire. L'obtention d'estimations fidèles à la réalité est cruciale dans le domaine de l'actuariat, influençant directement les stratégies de gestion des risques.

L'objectif de ce chapitre est d'estimer au plus près les taux de mortalité par âge, sexe et année calendaire. Pour ce faire, le modèle de LEE & CARTER sera d'abord examiné, suivi de l'exploration de deux de ses extensions : le modèle LOG POISSON et le modèle de RENSHAW & HABERMAN.

1.1 Le modèle de LEE & CARTER

La méthode développée par LEE & CARTER (1992) repose sur une étude démographique combinée à des méthodes statistiques de séries temporelles fondées sur des tendances historiques. Ce modèle repose sur l'idée que les changements dans les taux de mortalité peuvent être décomposés selon l'âge et le temps.

La définition

Ce modèle spécifie que la force de mortalité, à l'âge x et à l'année t , prend la forme suivante :

$$\ln \hat{\mu}_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

Les paramètres se définissent comme suit :

- α_x : un vecteur décrivant la forme générale des taux de mortalité en fonction de l'âge, invariant dans le temps.
- β_x : un vecteur, invariant dans le temps, qui quantifie la sensibilité du taux de mortalité à l'âge x .
- κ_t : un vecteur représentant l'évolution de la mortalité en fonction du temps. Une tendance à la baisse de la série des κ_t révèle une amélioration de la mortalité dans le temps.
- $\epsilon_{x,t}$: un terme d'erreur de moyenne nulle et de variance σ^2 invariante à l'âge et au temps.

Pour garantir des estimations uniques et cohérentes des paramètres, les conditions d'identifiabilité suivantes sont posées :

$$\sum_{x=x_m}^{x_M} \beta_x = 1 \text{ et } \sum_{t=t_m}^{t_M} \kappa_t = 0$$

L'estimation des paramètres

Les paramètres sont estimés en utilisant la méthode des moindres carrés ordinaires, résolvant ainsi le problème d'optimisation suivant :

$$(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t) = \underset{\alpha_x, \beta_x, \kappa_t}{\operatorname{arg\,min}} \sum_{x=x_m}^{x_M} \sum_{t=t_m}^{t_M} (\ln \hat{\mu}_{x,t} - \alpha_x - \beta_x \kappa_t)^2$$

L'estimation de α_x se déduit directement de la condition d'identifiabilité sur κ_t .

$$\begin{aligned} \frac{\partial}{\partial \alpha_x} \sum_{x=x_m}^{x_M} \sum_{t=t_m}^{t_M} (\ln \hat{\mu}_{x,t} - \alpha_x - \beta_x \kappa_t)^2 = 0 &\Leftrightarrow -2 \sum_{t=t_m}^{t_M} (\ln \hat{\mu}_{x,t} - \alpha_x - \beta_x \kappa_t) = 0 \\ &\Leftrightarrow \sum_{t=t_m}^{t_M} \alpha_x = \sum_{t=t_m}^{t_M} \ln \hat{\mu}_{x,t} - \sum_{t=t_m}^{t_M} \beta_x \kappa_t \\ &\Leftrightarrow (t_M - t_m + 1) \alpha_x = \sum_{t=t_m}^{t_M} \ln \hat{\mu}_{x,t} \end{aligned}$$

D'où :

$$\hat{\alpha}_x = \frac{1}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \ln \hat{\mu}_{x,t}$$

Les estimations de β_x et κ_t s'obtiennent par décomposition en valeurs singulières de la matrice $A = \ln \hat{\mu}_{x,t} - \hat{\alpha}_x$, de taille $m \times n$. Ainsi, $A = U \Sigma V^T$, avec :

- $U \in \mathcal{R}^{m \times m}$, la matrice orthogonale des vecteurs singuliers à gauche de A .
- $V \in \mathcal{R}^{n \times n}$, la matrice orthogonale des vecteurs singuliers à droite de A .
- $\Sigma \in \mathcal{R}^{m \times n}$, la matrice diagonale contenant les valeurs singulières $\sigma_1, \sigma_2, \dots, \sigma_m$.

Il est alors possible d'écrire :

$$A = \begin{bmatrix} U_{1,1} & \cdots & U_{1,m} \\ \vdots & \ddots & \vdots \\ U_{1,m} & \cdots & U_{m,m} \end{bmatrix} \times \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{bmatrix} \times \begin{bmatrix} V_{1,1} & \cdots & V_{1,m} & \cdots & V_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ V_{m,1} & \cdots & \cdots & \cdots & V_{m,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ V_{n,1} & \cdots & \cdots & \cdots & V_{n,n} \end{bmatrix}$$

En considérant $A = \ln \hat{\mu}_{x,t} - \hat{\alpha}_x$ avec une approximation de rang 1, il est possible d'écrire :

$$A_{x,t}^{(1)} = \sigma_1 U_{x,1} V_{1,t} = \beta_x^{(1)} \kappa_t^{(1)}$$

Il en découle que β_x correspondant au vecteur propre normalisé de AA^T , est égal à la première colonne de U normalisée. De plus, κ_t correspondant au premier vecteur propre normalisé de $A^T A$ est égal à la première colonne normalisée de V multipliée par la plus grande valeur propre de Σ , σ_1 .

Ainsi,

$$\hat{\beta}_x = \frac{1}{\sum_x U_{x,1}} U_1 \quad \text{et} \quad \hat{\kappa}_t = \sigma_1 \sum_x U_{x,1} V_1$$



Une idée d'exploration : Des articles de recherche ont identifié de nouvelles méthodes d'estimation de ces paramètres, notamment l'article, paru en 2022, intitulé "*Robust Parameter Estimation for the Lee-Carter Family : A Probabilistic Principal Component Approach*" par Yiping Guo et Johnny Siu-Hang Li. Cet article explore l'utilisation d'un modèle probabiliste d'analyse en composantes principales avec des distributions multivariées (*t*-PPCA) permettant une estimation robuste des paramètres. Des études démontrent la résilience de ce modèle contre les valeurs aberrantes. Cette approche pourrait être intéressante à considérer, car elle permettrait une prise en compte modérée des points atypiques.

Les limites du modèle

Le modèle de LEE & CARTER, largement utilisé et apprécié pour sa simplicité dans la modélisation de la mortalité, présente toutefois certaines limites qui peuvent encourager l'exploration d'autres modèles.

- **L'hypothèse d'homoscédasticité des taux de mortalité :** Le modèle suppose que la variance des erreurs est constante à travers tous les âges et les périodes. Cette hypothèse d'homoscédasticité, nécessaire pour l'estimation par moindres carrés ordinaires, peut ne pas être vérifiée. Par exemple, en raison de la diminution de la population vivante observée aux âges les plus avancés, la variance du logarithme des taux de décès augmente avec l'âge.
- **La sensibilité de l'âge, β_x , ne dépend pas du temps :** La sensibilité des taux de mortalité, représentée par β_x , est supposée constante dans le temps. Cette hypothèse limite le modèle dans des situations où la sensibilité pourrait varier en fonction de circonstances spécifiques. Par exemple :
 - En cas de conflit armé entre deux temps, t_1 et t_2 , il y aurait une augmentation momentanée des taux de mortalité, $q_{x,t}$ pour les jeunes âges durant cette période, ce qui entraînerait un changement sur β_x . Le modèle classique de LEE & CARTER ne capture pas cette variation temporelle de la sensibilité.
 - Une découverte médicale majeure, telle qu'un traitement efficace contre Alzheimer, pourrait augmenter l'espérance de vie en réduisant significativement les taux de mortalité, $q_{x,t}$ pour les âges avancés. La sensibilité à des âges élevés pourrait ainsi devenir plus importante à partir de ce moment, mais le modèle ne permet pas de refléter ce changement dans le temps.
- **Un unique paramètre temporel :** Le modèle ne fait intervenir qu'un seul paramètre temporel, κ_t , pour capturer l'évolution de la mortalité dans le temps. Cette simplification peut limiter la capacité du modèle à représenter des dynamiques plus complexes, telles que les changements structurels soudains ou les effets temporels multiples.
- **Un besoin important de données :** Le modèle de LEE & CARTER nécessite une quantité importante de données pour fournir des estimations robustes des paramètres. En cas de données limitées ou de faible granularité, les résultats peuvent manquer de précision et de fiabilité.

L'application du modèle sur les données

En appliquant la méthode initiale sur les données, les paramètres sont estimés comme suit :

- **Le paramètre $\hat{\alpha}_x$** traduit une croissance globale des taux de mortalité en fonction de l'âge, tout en mettant en évidence une mortalité infantile marquée et une "bosse des accidents", indiquant un pic de décès chez les 18-25 ans, plus prononcée chez les hommes que chez les femmes. De plus, une différence de niveau de mortalité entre les hommes et les femmes est observée. Ces constats sont en adéquation avec les études précédemment menées.

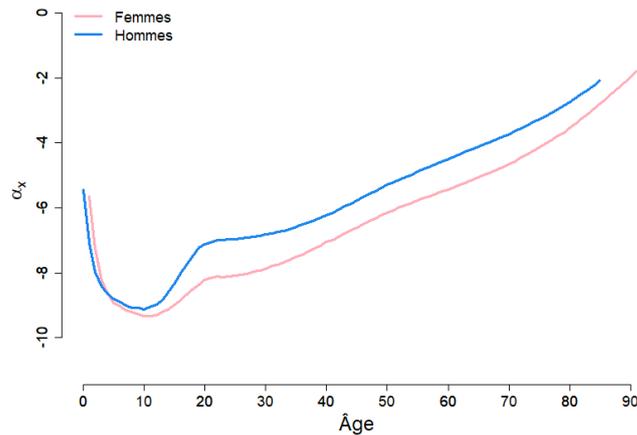


FIGURE 1.1 – Le paramètre $\hat{\alpha}_x$ du modèle de LEE & CARTER

- **Le paramètre $\hat{\beta}_x$** révèle une sensibilité variable de la mortalité selon l'âge. Une diminution significative suivie d'une augmentation est observée autour de 60 ans, avec des fluctuations moins prononcées chez les hommes que chez les femmes.

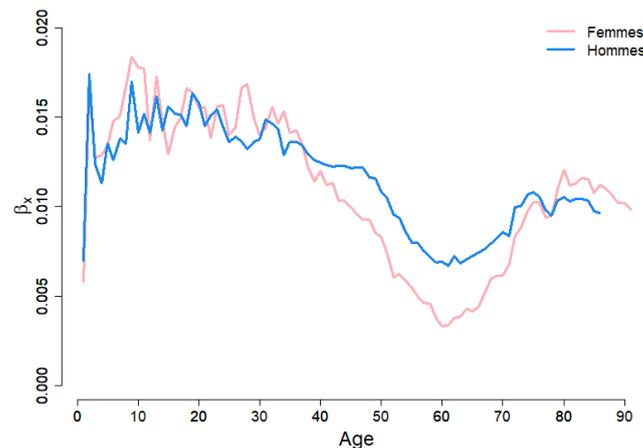


FIGURE 1.2 – Le paramètre $\hat{\beta}_x$ du modèle de LEE & CARTER

- **Le paramètre $\hat{\kappa}_t$** montre une diminution globale de la mortalité dans le temps, suggérant que les taux de mortalité se sont améliorés. Entre 2005 et 2019, les hommes ont bénéficié d'une amélioration de la mortalité plus conséquente que les femmes. La bosse observée depuis 2019 pourrait indiquer un changement de dynamique. Cette analyse est en cohérence avec l'hypothèse soulevée dans les chapitres précédents concernant un possible ralentissement de l'amélioration de la mortalité dans le temps.

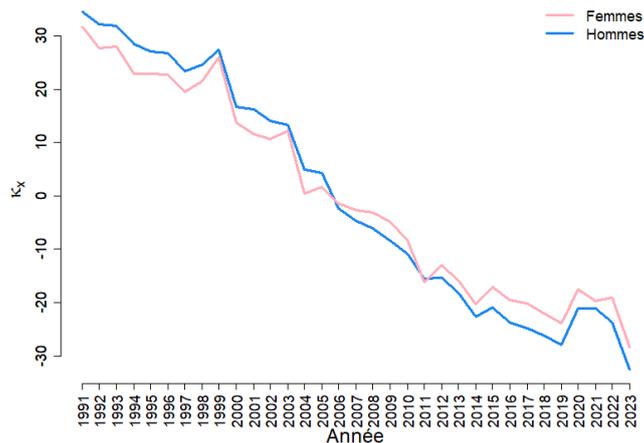


FIGURE 1.3 – Le paramètre \hat{k}_t du modèle de LEE & CARTER

Les estimations des paramètres permettent de déduire les taux de mortalité par âge et par année. Les graphiques ci-dessus illustrent ces taux estimés pour les hommes et pour les femmes. Ces taux de mortalité ne suivent pas une progression parfaitement lisse, des sauts et des irrégularités sont visibles. Ces irrégularités peuvent rappeler les sauts observés précédemment dans les taux de mortalité bruts, où des changements brusques et non anticipés compliquent l'analyse et l'interprétation des données.

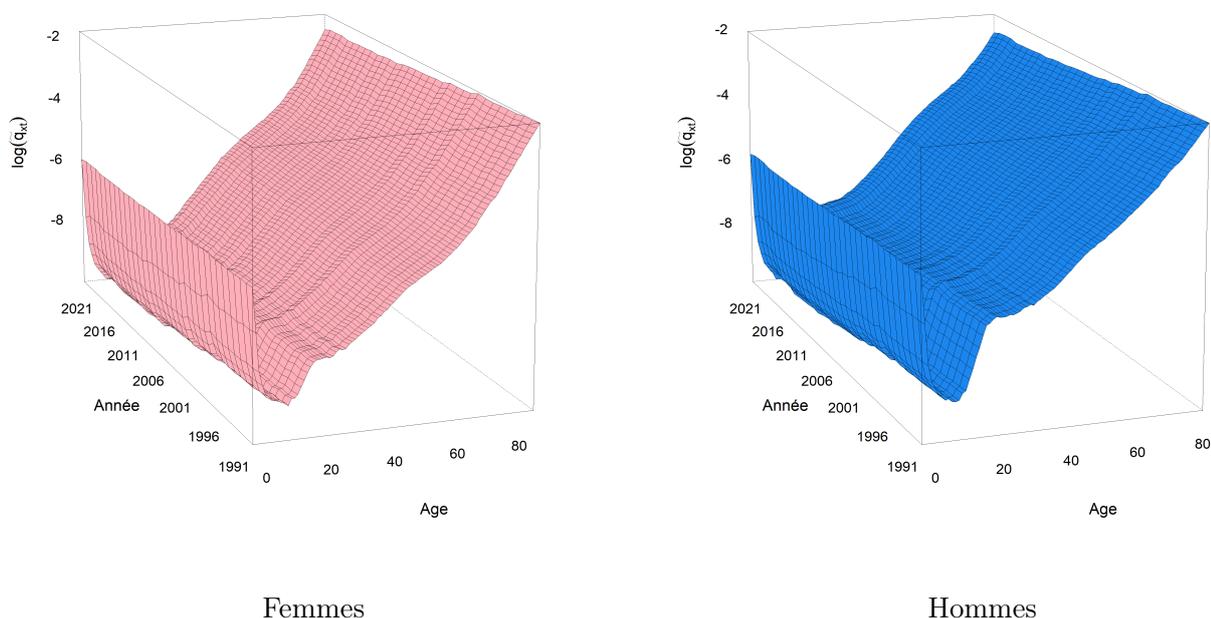


FIGURE 1.4 – Le logarithme des taux de mortalité estimés par le modèle de LEE & CARTER

Ces sauts peuvent être problématiques, car ils ne reflètent pas nécessairement une véritable variation de la mortalité, mais plutôt des fluctuations dues à la variabilité des données. Dans le contexte des projections et de la construction de tables de mortalité, il est crucial d'assurer la continuité et la régularité des taux, car des sauts non justifiés peuvent entraîner des biais ou des erreurs dans les projections futures.

Pour cette raison, il est nécessaire de lisser les taux de mortalité. Le lissage vise à éliminer les fluctuations non significatives, en fournissant une représentation plus cohérente et fidèle de l'évolution des taux de mortalité.

Dans ce mémoire, les paramètres estimés ont directement été lissés via deux méthodes différentes :

- **Une régression polynomiale** avec une optimisation du degré permettant de modéliser une relation avec des équations polynomiales¹.
- **Un lissage par splines** consiste à diviser la plage de la fonction à ajuster en plusieurs segments. Une fonction polynomiale est ajustée sur chaque segment, en veillant à ce que les fonctions soient continues aux points de jonction, appelés nœuds².



Une idée d'exploration : Un lissage des taux de mortalité aurait pu être envisagé avec des méthodes de lissage non paramétriques comme la méthode de WHITTAKER - HENDERSON. Cette méthode consiste à équilibrer un critère de fidélité aux données observées et un critère de régularité de la courbe, cherchant un compromis optimal pour minimiser la somme de ces deux critères. Celle-ci est présentée en annexe.

L'application des techniques de lissage sur les paramètres du modèle de LEE & CARTER, a conduit aux résultats suivants :

- **Le paramètre α_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre α_x des femmes et des hommes.

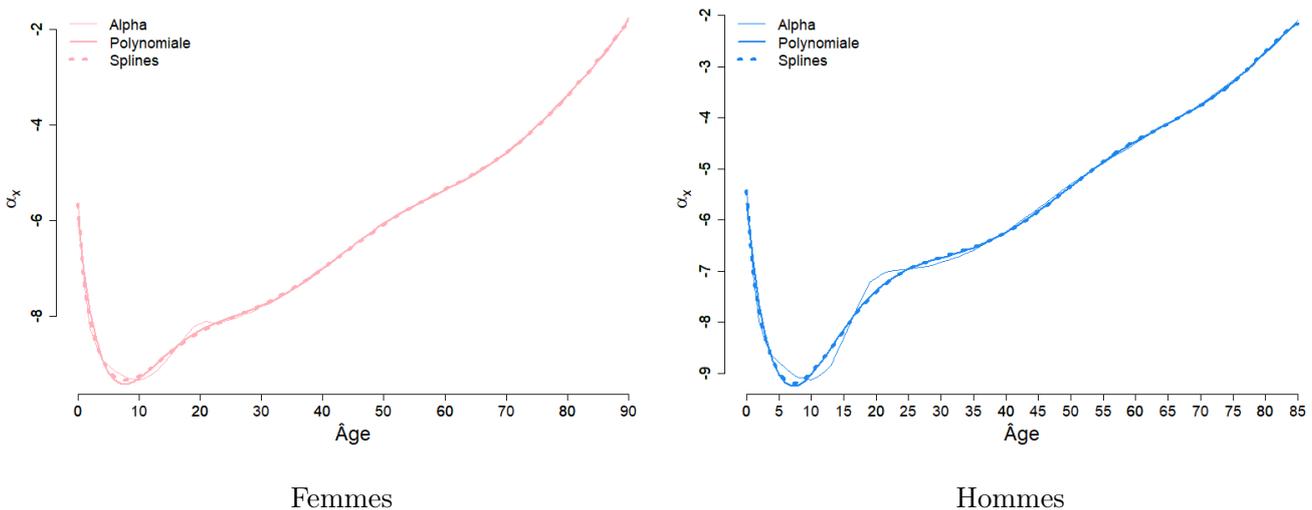


FIGURE 1.5 – Le paramètre $\hat{\alpha}_x$ du modèle de LEE & CARTER lissé

Les lissages par régression polynomiale et par splines montrent de très bonnes adéquations aux données, notamment à partir de 30 ans. Les valeurs des coefficients de détermination R^2 , très proches de 1, confirment cette analyse.

1. Cette méthode est présentée en annexe
 2. Cette méthode est présentée en annexe

	Femmes	Hommes
Polynomiale	0,9985	0,9964
Splines	0,9992	0,9969

TABLE 1.1 – Le lissage du paramètre α_x de la méthode de LEE & CARTER : Les valeurs de R^2

Ces résultats indiquent que les modèles lissés capturent parfaitement, pour les femmes et pour les hommes, la tendance générale de la mortalité par âge, tout en lissant les irrégularités observées dans les données. Le lissage par la méthode des splines est retenu dans la suite de ce mémoire.

— **Le paramètre β_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre β_x des femmes et des hommes.

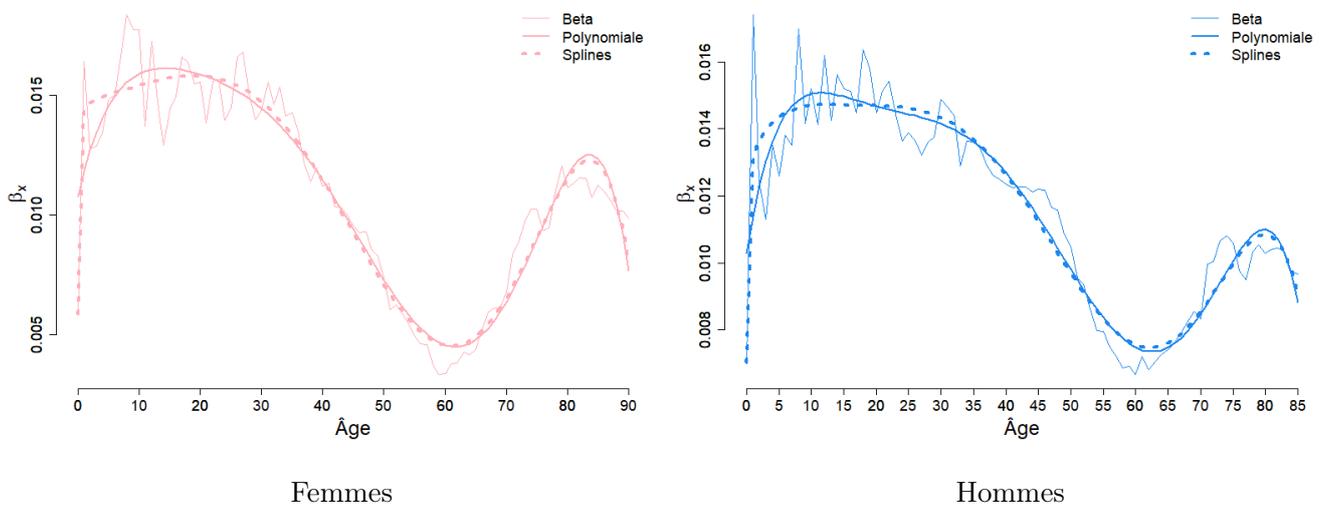


FIGURE 1.6 – Le paramètre $\hat{\beta}_x$ du modèle de LEE & CARTER lissé

Les courbes lissées par régression polynomiale et par splines capturent efficacement la tendance générale de la série brute, en atténuant les variations soudaines tout en restant fidèle aux principales inflexions.

	Femmes	Hommes
Polynomiale	0,9052	0,8686
Splines	0,9322	0,8952

TABLE 1.2 – Le lissage du paramètre β_x de la méthode de LEE & CARTER : Les valeurs de R^2

Les valeurs élevées de R^2 indiquent que les modèles expliquent bien la variance de la série β_x . Le lissage par la méthode des splines est retenu dans la suite de ce mémoire.

— **Le paramètre κ_t** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre κ_t des femmes et des hommes.

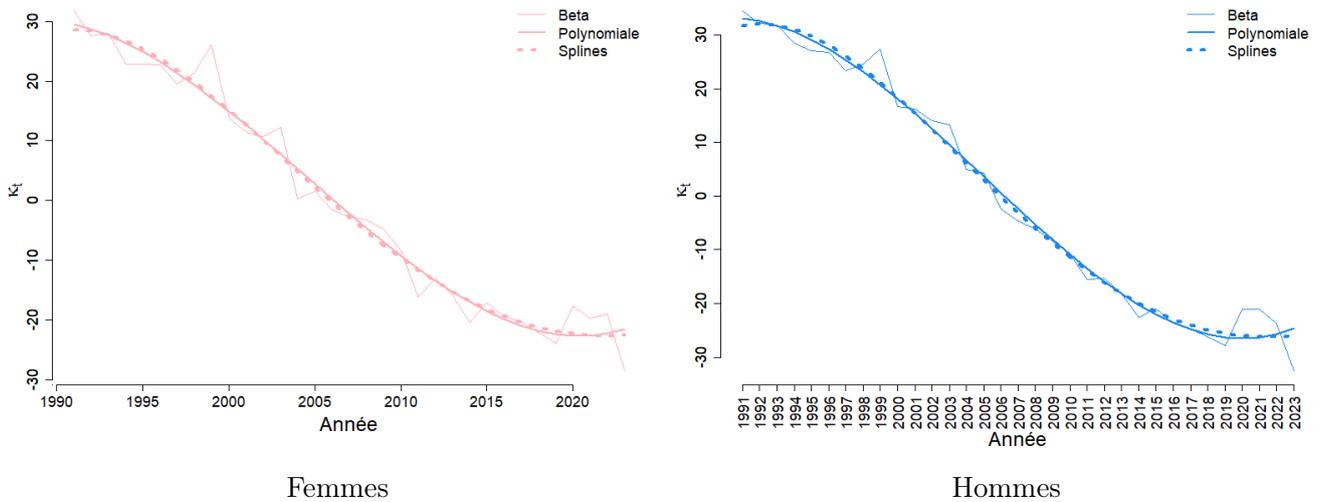


FIGURE 1.7 – Le paramètre $\hat{\kappa}_t$ du modèle de LEE & CARTER lissé

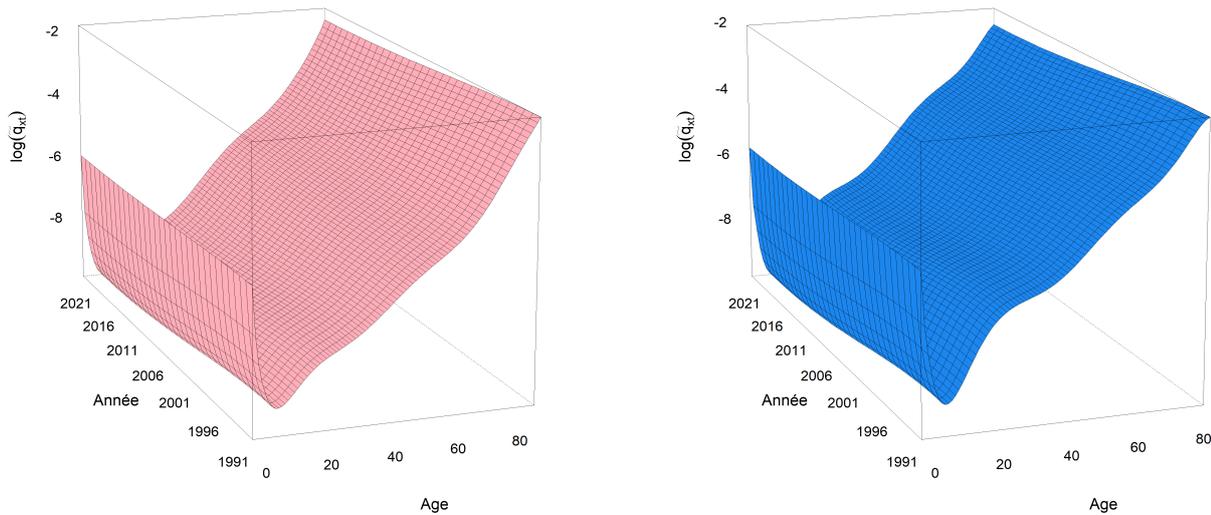
Les courbes obtenues par lissage polynomial et lissage par splines atténuent les fluctuations tout en capturant la tendance générale de la diminution de la mortalité au fil du temps.

	Femmes	Hommes
Polynomiale	0,9739	0,9848
Splines	0,9745	0,9857

TABLE 1.3 – Le lissage du paramètre κ_t de la méthode de LEE & CARTER : Les valeurs de R^2

Les valeurs de R^2 montrent que le lissage est efficace pour obtenir une courbe de κ_t plus régulière. La méthode des splines est retenue afin d'obtenir une surface de mortalité lisse, sans variation soudaine. Il est important de noter que, pour la projection des taux de mortalité, la série brute de κ_t sera utilisée. En effet, la "bosse" observée en 2019 pourrait indiquer un changement structurel dans la dynamique de mortalité. Capturer cette caractéristique est essentiel pour anticiper de manière précise l'évolution future de la mortalité, un point fondamental de ce mémoire.

Les graphiques 3D lissés montrent une amélioration significative par rapport aux versions non lissées. Le lissage des paramètres du modèle de LEE & CARTER a permis de supprimer les sauts indésirables, offrant ainsi une représentation plus exploitable des taux de mortalité par âge et par année.



Femmes

Hommes

FIGURE 1.8 – Le logarithme des taux de mortalité estimés (lissés) par le modèle de LEE & CARTER

1.2 Le modèle LOG POISSON

La méthodologie introduite par BROUHNS, DENUIT et VERMUNT en 2002, est une extension du modèle de LEE & CARTER. Ce modèle vise à surmonter certaines limitations inhérentes au modèle original, en particulier l'hypothèse d'homoscédasticité des résidus. À la différence du modèle de LEE & CARTER, le modèle LOG POISSON ne suppose pas que la variance des erreurs reste constante quel que soit l'âge ou la période. Ceci est particulièrement pertinent dans les contextes démographiques où la variance des taux de mortalité peut fluctuer avec l'âge et le temps en raison de la variabilité de la taille des populations à risque. Ainsi, le modèle LOG POISSON constitue un avantage par rapport au modèle de LEE & CARTER dans un contexte de petits échantillons. Dans notre cas, les données provenant de l'INSEE réduisent le risque d'échantillonnage avec le modèle de LEE & CARTER, rendant cet effet peu conséquent. Les résultats obtenus avec les deux modèles devront être proches.

La définition

Le modèle LOG POISSON propose de modéliser le nombre de décès à l'âge x et à l'année t , $D_{x,t}$, en utilisant une approche basée sur la loi de Poisson de paramètre $E_{x,t}\mu_{x,t}$, où $\mu_{x,t} = \exp(\alpha_x + \beta_x \kappa_t)$ et $E_{x,t}$ est l'exposition au risque à l'âge x et à l'année t . Plus précisément, bien que les décès observés ne suivent pas rigoureusement une loi de Poisson, cette hypothèse est introduite afin de rendre possible l'estimation des paramètres du modèle via la méthode du maximum de vraisemblance.

La distribution du nombre de décès peut alors être formulée comme suit :

$$D_{x,t} \sim \mathcal{P}(E_{x,t}\mu_{x,t})$$

La probabilité associée à cette distribution est exprimée par :

$$P(D_{x,t} = d) = \frac{(E_{x,t}\mu_{x,t})^d}{d!} \exp(-E_{x,t}\mu_{x,t})$$

La relation suivante résulte de la distribution de Poisson :

$$E[D_{x,t}] = Var(D_{x,t}) = E_{x,t}\mu_{x,t}$$

Comme pour le modèle de LEE & CARTER, certaines contraintes d'identifiabilité sont nécessaires pour garantir des estimations uniques et cohérentes des paramètres. Elles sont identiques à celles du modèle de LEE & CARTER.

$$\sum_{x=x_m}^{x_M} \beta_x = 1 \text{ et } \sum_{t=t_m}^{t_M} \kappa_t = 0$$

Les paramètres du modèle LOG POISSON conservent une interprétation similaire à ceux du modèle de LEE & CARTER :

- α_x : représente le niveau de mortalité de base spécifique à chaque âge, invariant dans le temps.
- β_x : mesure la sensibilité du taux de mortalité à l'âge x , c'est-à-dire l'effet d'une modification de la tendance générale de la mortalité sur un âge particulier.
- κ_t : capture l'évolution temporelle de la mortalité. Une diminution de κ_t au fil du temps traduit une amélioration générale de la mortalité dans la population étudiée.

L'estimation des paramètres

Les paramètres du modèle $(\alpha_x, \beta_x, \kappa_t)$ sont généralement estimés par la méthode du maximum de vraisemblance, consistant à maximiser la fonction de log-vraisemblance suivante (à une constante additive près) :

$$\ln(L(\alpha, \beta, \kappa)) = \sum_{x,t} (D_{x,t}(\alpha_x + \beta_x \kappa_t) - E_{x,t} \exp(\alpha_x + \beta_x \kappa_t))$$

La présence du terme bilinéaire $\beta_x \kappa_t$, ne permet pas d'obtenir une estimation directe des paramètres. Une résolution numérique, comme la méthode de NEWTON RAPHSON, doit être implémentée. Celle-ci suit plusieurs étapes :

- L'initialisation des paramètres s'effectue librement (excepté pour β_x^0 qui doit être non nul).
- Les itérations suivantes sont employées :

$$\hat{\alpha}_x^{i+1} = \hat{\alpha}_x^i - \frac{\sum_t (D_{x,t} - E_{x,t} \exp(\hat{\alpha}_x^i + \hat{\beta}_x^i \hat{\kappa}_t^i))}{-\sum_t E_{x,t} \exp(\hat{\alpha}_x^i + \hat{\beta}_x^i \hat{\kappa}_t^i)}$$

$$\hat{\kappa}_t^{i+1} = \hat{\kappa}_t^i - \frac{\sum_t (D_{x,t} - E_{x,t} \exp(\hat{\alpha}_x^{i+1} + \hat{\beta}_x^i \hat{\kappa}_t^i)) \hat{\beta}_x^i}{-\sum_t E_{x,t} \exp(\hat{\alpha}_x^{i+1} + \hat{\beta}_x^i \hat{\kappa}_t^i) (\hat{\beta}_x^i)^2}$$

$$\hat{\beta}_x^{i+1} = \hat{\beta}_x^i - \frac{\sum_t (D_{x,t} - E_{x,t} \exp(\hat{\alpha}_x^{i+1} + \hat{\beta}_x^i \hat{\kappa}_t^{i+1})) \hat{\beta}_t^i}{-\sum_t E_{x,t} \exp(\hat{\alpha}_x^{i+1} + \hat{\beta}_x^i \hat{\kappa}_t^{i+1}) (\hat{\kappa}_t^{i+1})^2}$$

— L'ajustement des paramètres pour tenir compte des contraintes d'identifiabilité du modèle :

$$\alpha_x^* = \hat{\alpha}_x + \frac{\hat{\beta}_x}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \hat{\kappa}_t$$

$$\beta_x^* = \frac{\hat{\beta}_x}{\sum_x \hat{\beta}_x}$$

$$\kappa_t^* = (\hat{\kappa}_t - \frac{1}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \hat{\kappa}_t) \sum_x \hat{\beta}_x$$

Les limites du modèle

L'expression de taux de mortalité étant similaire à celle du modèle de LEE & CARTER, certaines limites sont également présentes avec le modèle LOG POISSON : la sensibilité de l'âge à l'évolution globale (β_x) ne dépend pas du temps, le modèle repose sur un unique paramètre temporel et nécessite une grande quantité de données.

L'application du modèle sur les données

En appliquant la méthode initiale sur les données, les paramètres sont estimés comme suit :

- **Le paramètre $\hat{\alpha}_x$** révèle une mortalité plus élevée aux âges avancés, avec un pic notable chez les jeunes adultes, suggérant une "bosse des accidents".

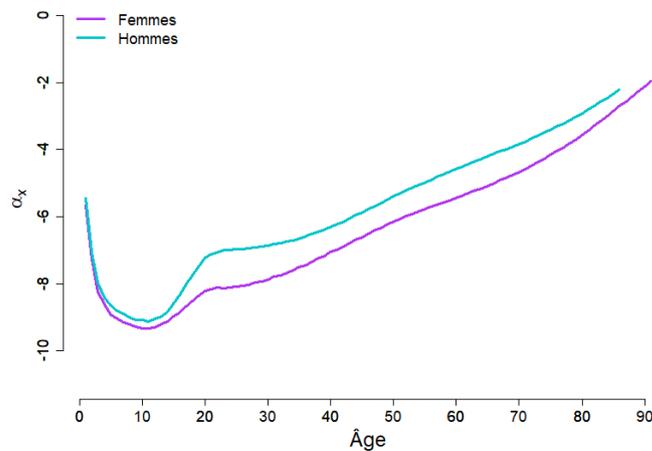


FIGURE 1.9 – Le paramètre $\hat{\alpha}_x$ du modèle LOG POISSON

- **Le paramètre $\hat{\beta}_x$** révèle des fluctuations attribuables à des périodes spécifiques où certains âges sont plus ou moins affectés par des changements de tendance.

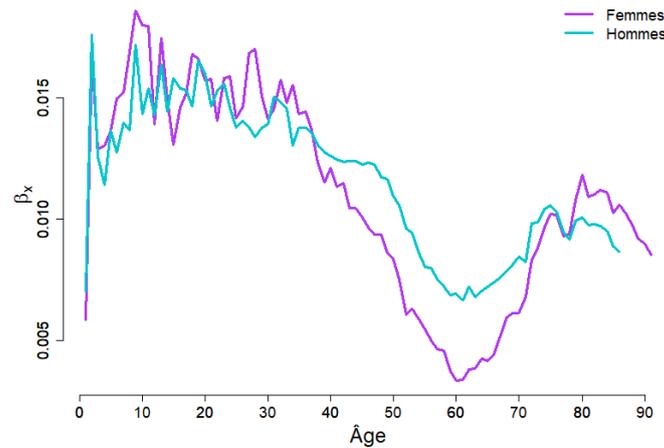


FIGURE 1.10 – Le paramètre $\hat{\beta}_x$ du modèle de LOG POISSON

- **Le paramètre $\hat{\kappa}_t$** montre une tendance à la baisse, indiquant une amélioration générale de la mortalité au fil du temps. Cependant, des fluctuations importantes, notamment autour de l'année 2019, pourraient signaler un événement ou un changement structural ayant impacté la mortalité.



FIGURE 1.11 – Le paramètre $\hat{\kappa}_t$ du modèle de LOG POISSON

Un lissage par splines a été appliqué aux paramètres. La méthode étant la même que celle présentée pour le modèle de LEE & CARTER, les résultats ne sont pas détaillés ici mais en annexe. Cette étape a permis de supprimer les sauts indésirables, offrant ainsi une représentation plus exploitable des taux de mortalité par âge et par année.

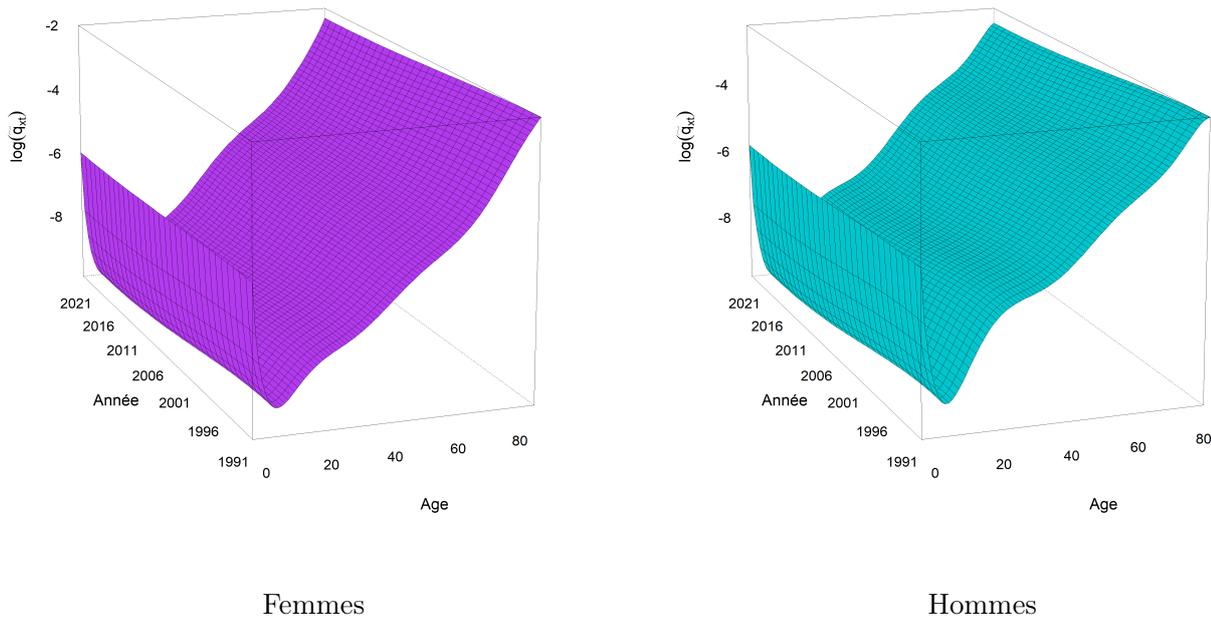


FIGURE 1.12 – Le logarithme des taux de mortalité estimés (lissés) par le modèle LOG POISSON

1.3 Le modèle de RENSHAW & HABERMAN

La modélisation proposée par RENSHAW & HABERMAN est une extension des modèles précédents. Il intègre des paramètres supplémentaires permettant de prendre en compte un effet générationnel dans l'estimation des taux de mortalité et ainsi capturer les influences historiques et sociétales partagées par une cohorte (individus nés la même année).

La définition

Le modèle de RENSHAW & HABERMAN peut être exprimé par l'équation suivante :

$$\ln \hat{\mu}_{x,t} = \alpha_x + \beta_x \kappa_t + \beta_x^{(2)} \gamma_{t-x} + \epsilon_{x,t}$$

Où :

- α_x capture l'effet de l'âge sur les taux de mortalité, invariant dans le temps.
- β_x quantifie la sensibilité de la mortalité à l'âge x , invariant dans le temps.
- κ_t capture l'évolution temporelle de la mortalité.
- γ_{t-x} représente l'effet cohorte des individus âgés de x ans et nés dans l'année $t - x$.
- $\beta_x^{(2)}$ représente la sensibilité à chaque âge de l'effet cohorte.

Les paramètres communs avec le modèle de LEE & CARTER sont soumis aux mêmes contraintes d'identifiabilité, et deux autres sont également imposées :

$$\sum_{x=x_m}^{x_M} \beta_x = 1 \text{ et } \sum_{t=t_m}^{t_M} \kappa_t = 0 \text{ et } \sum_{x=x_m}^{x_M} \beta_x^{(2)} = 1 \text{ et } \sum_{t=t_m}^{t_M} \gamma_{t-x} = 0$$

Ce modèle propose également de modéliser le nombre de décès à l'âge x et à l'année t , $D_{x,t}$, par une loi de Poisson de paramètre $E_{x,t}\mu_{x,t}$.

L'estimation des paramètres

L'estimation des paramètres s'établit en plusieurs parties :

- L'effet de l'âge est estimé par la moyenne des logarithmes des taux.

$$\hat{\alpha}_x = \frac{1}{t_M - t_m + 1} \sum_{t=t_m}^{t_M} \ln \hat{\mu}_{xt}$$

- Le processus itératif suivant est utilisé :

- Les paramètres initiaux sont fixés.
- Les itérations suivantes sont effectuées :

$$\hat{\kappa}_t^{i+1} = \hat{\kappa}_t^i + \frac{\sum_x \omega_{xt} \hat{\beta}_x^i (D_{x,t} - \hat{y}_{x,t})}{\sum_x \omega_{x,t} \hat{\beta}_x^{i2} \hat{y}_{x,t}}$$

$$\hat{\gamma}_{t-x}^{i+1} = \hat{\gamma}_{t-x}^i + \frac{\sum_x \omega_{xt} \hat{\beta}_x^{(2)i} (D_{x,t} - \hat{y}_{x,t})}{\sum_x \omega_{x,t} \hat{\beta}_x^{(2)i2} \hat{y}_{x,t}}$$

$$\hat{\beta}_x^{i+1} = \hat{\beta}_x^i + \frac{\sum_t \omega_{xt} \hat{\kappa}_t^{i+1} (D_{x,t} - \hat{y}_{x,t})}{\sum_t \omega_{x,t} \hat{\kappa}_t^{i+12} \hat{y}_{x,t}}$$

$$\hat{\beta}_x^{(2)(i+1)} = \hat{\beta}_x^{(2)i} + \frac{\sum_t \omega_{xt} \hat{\gamma}_{t-x}^{i+1} (D_{x,t} - \hat{y}_{x,t})}{\sum_t \omega_{x,t} \hat{\gamma}_{t-x}^{i+12} \hat{y}_{x,t}}$$

avec $y_{x,t} = \ln(\mu_{x,t})$

- L'ajustement des paramètres pour tenir compte des contraintes d'identifiabilité.

Une des simplifications possibles est de poser $\beta_x^{(2)} = 1$. Cette hypothèse est faite dans la suite de ce mémoire. Le modèle se définit donc comme suit :

$$\ln(\hat{\mu}_{x,t}) = \alpha_x + \beta_x \kappa_t + \gamma_{t-x} + \epsilon_{x,t}$$

L'application du modèle sur les données

En appliquant la méthode de RENSHAW & HABERMAN sur les données, les paramètres sont estimés comme suit :

- **Le paramètre $\hat{\alpha}_x$** : La courbe α_x reflète une mortalité infantile élevée, suivie d'une baisse significative de la mortalité au début de l'âge adulte, avant une remontée progressive à mesure que l'âge augmente. Les hommes tendent à avoir des taux de mortalité plus élevés que les femmes, notamment à partir de la "bosse des accidents".

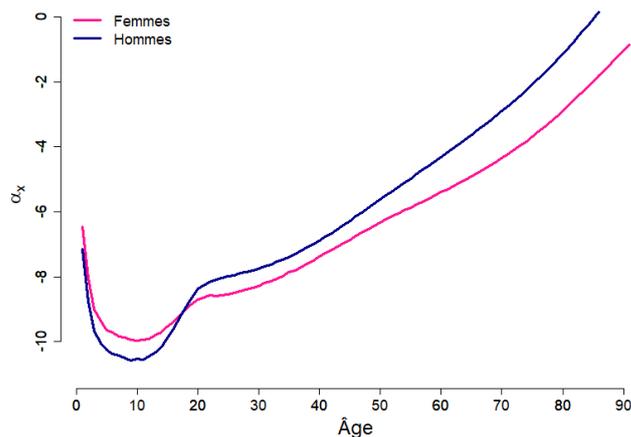


FIGURE 1.13 – Le paramètre $\hat{\alpha}_x$ du modèle de RENSRAW & HABERMAN

- **Le paramètre $\hat{\beta}_x$** : La courbe β_x montre une variabilité considérable dans la sensibilité de la mortalité à l'âge, notamment aux âges avancés.

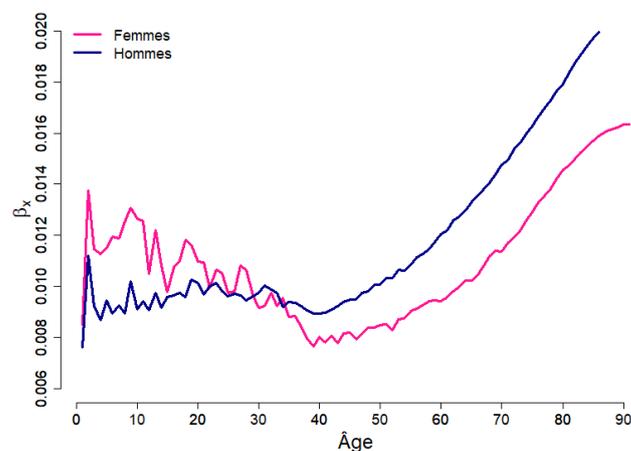


FIGURE 1.14 – Le paramètre $\hat{\beta}_x$ du modèle de RENSRAW & HABERMAN

- **Le paramètre $\hat{\kappa}_t$** : La courbe κ_t montre une tendance générale à la baisse, avec une décroissance plus marquée chez les hommes que chez les femmes. Cet écart entre les sexes apparaît plus accentué que dans les autres modèles étudiés auparavant, suggérant que le modèle de RENSRAW & HABERMAN capte de manière différente les spécificités d'évolution de la mortalité dans le temps.

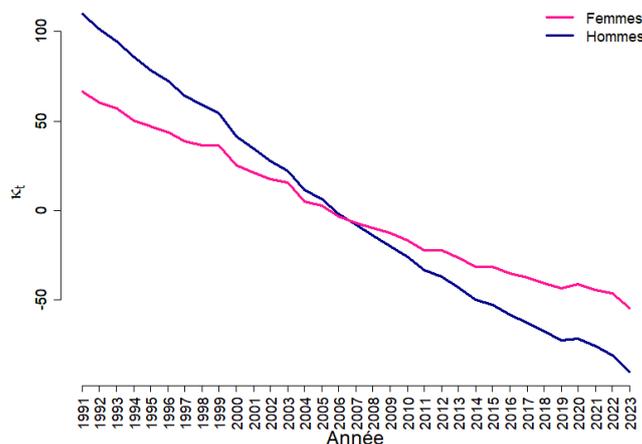


FIGURE 1.15 – Le paramètre $\hat{\kappa}_t$ du modèle de RENSRAW & HABERMAN

- **Le paramètre $\hat{\gamma}_{t-x}$** : Introduit dans le modèle de RENSCHAW HABERMAN, le paramètre γ_{t-x} capture l'effet cohorte, reflétant les influences spécifiques à une génération qui ne sont pas expliquées par l'âge ou la période. Les courbes associées à γ_{t-x} offrent des informations précieuses sur les dynamiques intergénérationnelles de la mortalité.

Pour les hommes et les femmes, les courbes de γ_{t-x} montrent une amélioration générale de la mortalité dans les cohortes récentes. Cette tendance est particulièrement marquée chez les hommes à partir de 1955, où l'amélioration s'accroît par rapport aux femmes. Cependant, des pics dans les courbes indiquent l'influence de facteurs spécifiques ou d'événements historiques, comme les guerres mondiales.

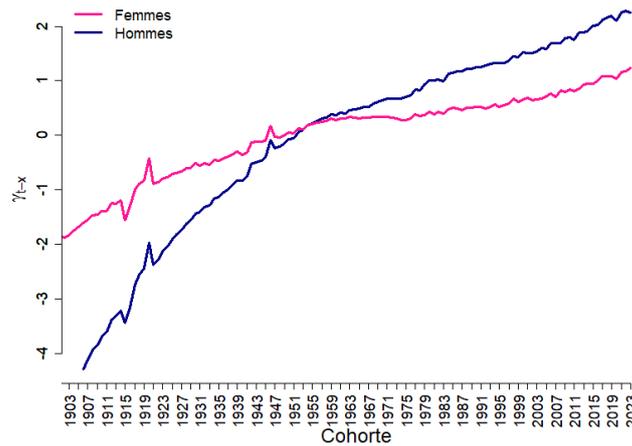


FIGURE 1.16 – Paramètres $\hat{\gamma}_{t-x}$ du modèle de RENSCHAW & HABERMAN

Un lissage par splines a été appliqué aux paramètres. La méthode étant la même que celle présentée pour le modèle de LEE & CARTER, les résultats ne sont pas détaillés ici mais en annexe. Cette étape a permis de supprimer les sauts indésirables, offrant ainsi une représentation plus exploitable des taux de mortalité par âge et par année.

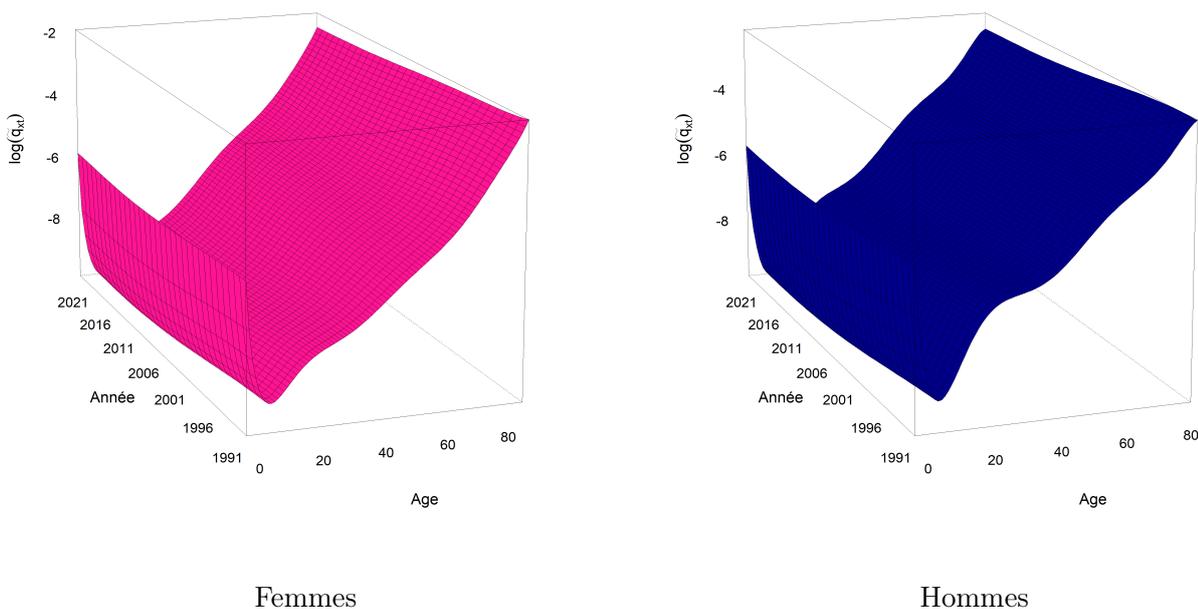


FIGURE 1.17 – Le logarithme des taux de mortalité estimés (lissés) par le modèle de RENSCHAW & HABERMAN

1.4 La comparaison des paramètres

Suite à l'estimation et au lissage des divers paramètres des modèles, une analyse comparative des paramètres communs est effectuée. Cette approche permet d'examiner dans quelle mesure le changement de la loi qui régit le nombre de décès affecte les estimations des paramètres, et d'évaluer l'impact de l'ajout d'un terme de cohorte sur les autres paramètres des modèles.

- **Le paramètre $\hat{\alpha}_x$** : La courbe du paramètre, capturant la structure de la mortalité par âge, montre des similitudes remarquables entre les trois modèles. Cependant, le modèle de RENSHAW & HABERMAN tend à estimer des valeurs plus faibles pour les âges de 0 à 60 ans et plus élevées pour les âges avancés (au-delà de 60 ans) par rapport aux modèles de LEE & CARTER et LOG POISSON. Des effets spécifiques à des générations particulières sont captés par le paramètre γ_{t-x} , modifiant la structure de mortalité par âge. Les estimations effectuées avec les modèles de LEE & CARTER et LOG POISSON se superposent quasi totalement.

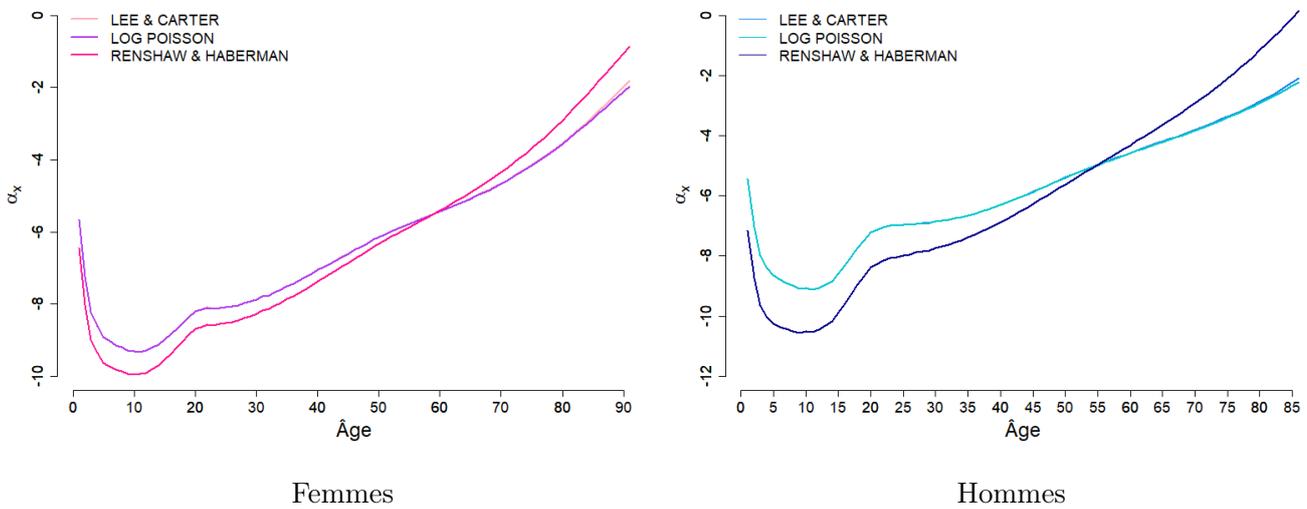


FIGURE 1.18 – La comparaison du paramètre α_x

- **Le paramètre $\hat{\beta}_x$** : Les graphiques montrent une forte sensibilité pour les âges de 0 à 50 ans dans les modèles de LEE & CARTER et LOG POISSON, qui diminue ensuite pour les âges avancés. En revanche, le modèle de RENSHAW & HABERMAN présente une sensibilité plus constante, mais avec une accentuation pour les âges avancés. Cette différence peut s'expliquer par le fait que l'effet de cohorte introduit dans le modèle de RENSHAW & HABERMAN modifie la relation entre l'âge et les changements temporels de la mortalité, en particulier pour les générations qui ont connu des événements historiques marquants. Les estimations effectuées avec les modèles de LEE & CARTER et LOG POISSON se superposent quasi totalement.

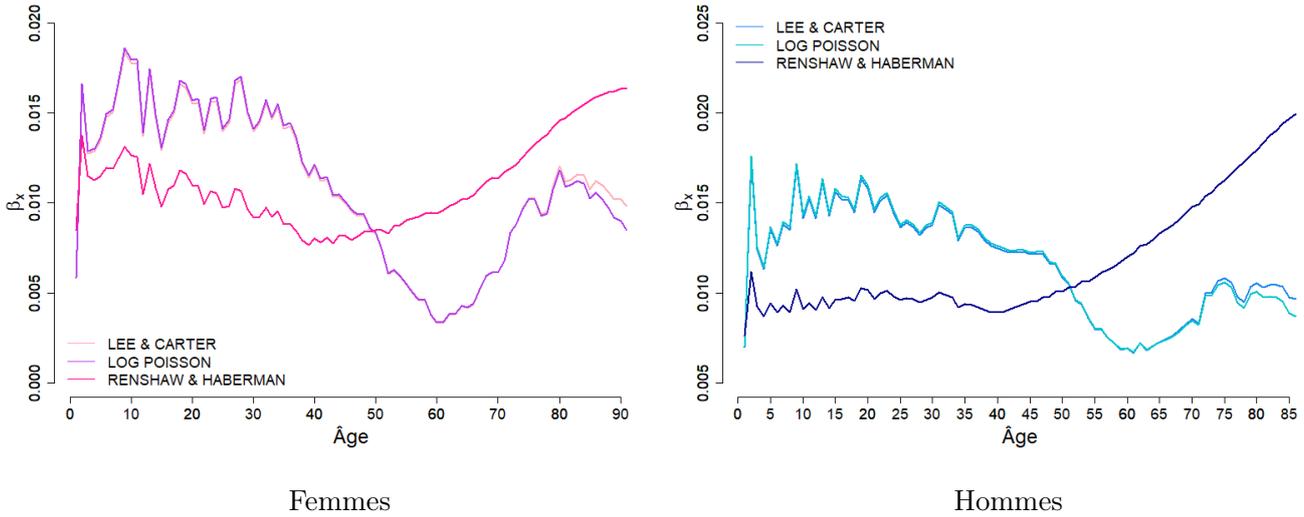


FIGURE 1.19 – La comparaison du paramètre β_x

- **Le paramètre $\hat{\kappa}_t$** : présente une diminution nettement plus raide, de la mortalité dans le temps, avec le modèle de RENSRAW & HABERMAN que pour les modèles de LEE & CARTER et LOG POISSON, que ce soit pour les femmes ou pour les hommes. La bosse observée depuis 2019 paraît moins prononcée avec le modèle de RENSRAW & HABERMAN, notamment pour les hommes. Cela pourrait être interprété comme une représentation plus fine des améliorations de la mortalité, rendue possible par l'inclusion du terme de cohorte. Les courbes des modèles de LEE & CARTER et LOG POISSON se superposent quasi totalement.

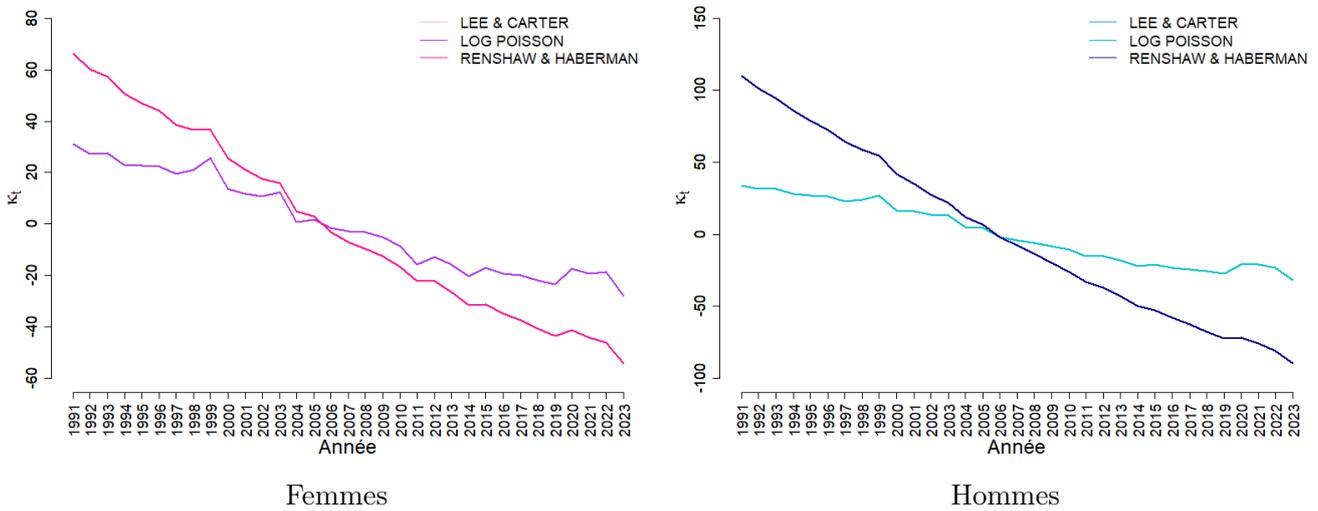


FIGURE 1.20 – La comparaison du paramètre κ_t

L'analyse comparative montre que les modèles de LEE & CARTER et LOG POISSON produisent des estimations très similaires pour les paramètres α_x , β_x , κ_t , indiquant que le changement de distribution des décès n'affecte pas significativement les résultats. Cependant, le modèle de RENSRAW & HABERMAN, intégrant un terme de cohorte, influence de manière significative les estimations, suggérant une sensibilité accrue de ce modèle aux variations associées à des effets générationnels spécifiques.



Ce qu'il faut retenir du chapitre :

- * Le modèle de LEE & CARTER capture les tendances de mortalité à travers le temps et les âges, mais présente des limites, en particulier liées à l'hypothèse d'homoscédasticité.
- * Cette contrainte est levée par le modèle LOG POISSON, qui suppose que le nombre de décès suit une loi de Poisson. Il affine l'estimation des taux de mortalité, particulièrement dans des contextes où les taux de mortalité sont volatils avec l'âge et le temps.
- * Le modèle de RENSHAW & HABERMAN introduit un terme de cohorte permettant de capturer des effets générationnels spécifiques. Il offre ainsi une représentation plus nuancée et précise des dynamiques de mortalité intergénérationnelles.
- * Les modèles de LEE & CARTER et LOG POISSON produisent des estimations de paramètres très similaires, tandis que le modèle de RENSHAW & HABERMAN, en incluant un terme de cohorte, entraîne des différences significatives dans les estimations des paramètres.

Le chapitre suivant validera le bon ajustement des taux de mortalité et fixera le modèle à retenir pour la projection.

Chapitre 2

La validation des résultats

Une validation rigoureuse permet d'assurer la fiabilité et la robustesse des estimations. Cette étape est cruciale dans la construction de tables de mortalité, permettant de vérifier la cohérence des résultats obtenus avec les données empiriques et ainsi de minimiser les risques associés aux erreurs de modélisation.

2.1 L'étude des taux de mortalité observés et prédits

Cette étude graphique évalue l'ajustement entre les taux de décès bruts, $q_{x,t}$, et ceux prédits par les différents modèles, $\hat{q}_{x,t}$. Pour chaque année calendaire et chaque sexe, les résultats montrent que les taux prédits par les divers modèles sont relativement proches les uns des autres et s'ajustent correctement aux taux observés. Les graphiques ci-dessous illustrent les taux pour les femmes et les hommes pour l'année 2015. L'alignement des prédictions avec les données observées atteste de la robustesse de ces modèles pour capturer, de manière fidèle, les tendances de mortalité.

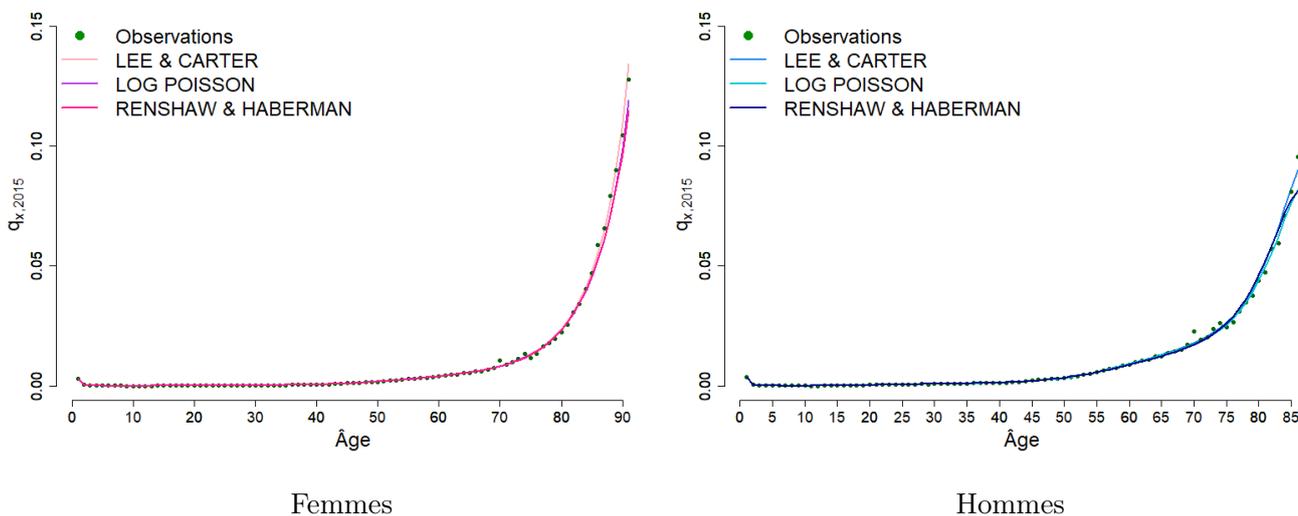


FIGURE 2.1 – Les taux de décès prédits et observés en 2015

2.2 Les intervalles de confiance du nombre de décès

À présent, l'analyse portera sur les intervalles de confiance du nombre de décès par classe d'âge, pour évaluer la précision des ajustements effectués par les différents modèles et mieux comprendre les incertitudes associées aux prédictions des taux de mortalité.

— De 0 à 40 ans :

Les ajustements sont globalement cohérents pour les trois modèles. Pour les femmes, le nombre de décès prédits reste constamment proche de l'intervalle de confiance sur l'ensemble de la période, indiquant que le modèle capture avec précision les tendances observées dans cette tranche d'âge. Pour les hommes, bien que l'ajustement dans la zone de la "bosse des accidents" entre 1991 et 2010 ait tendance à légèrement sous-estimer le nombre de décès, les prédictions sont globalement satisfaisantes. De même, malgré une surestimation autour de l'âge de 30 ans, les ajustements depuis 2013 montrent une nette amélioration, offrant une plus grande précision à tous les âges compris entre 0 et 40 ans. Les graphiques ci-dessous, présentés pour le modèle de LEE & CARTER, illustrent ces résultats positifs pour 2015.

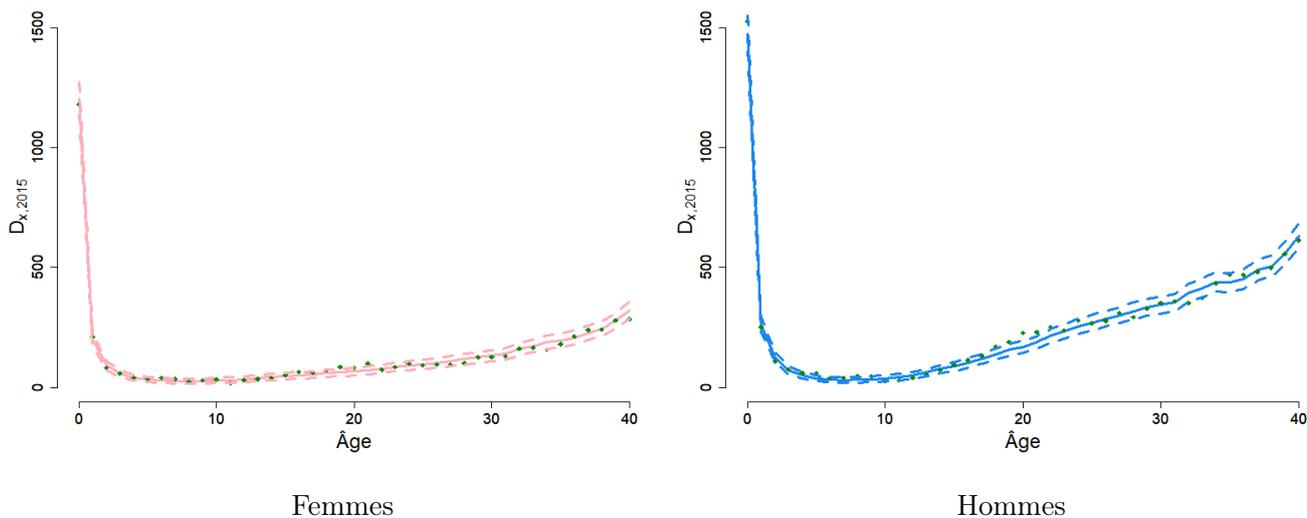
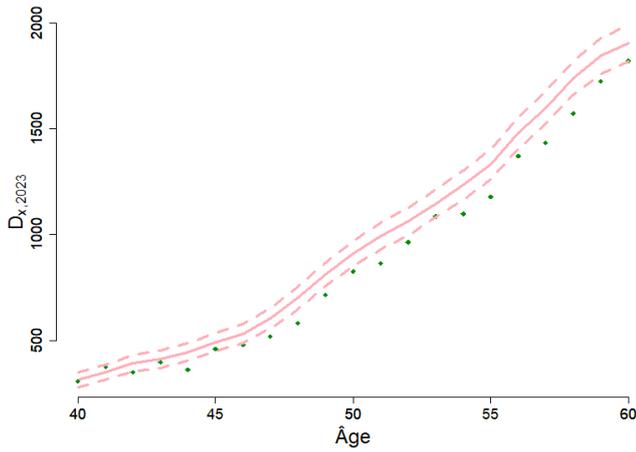


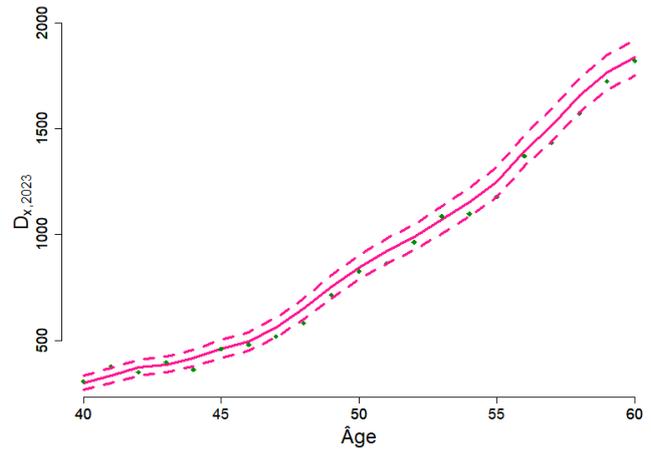
FIGURE 2.2 – Les intervalles de confiance de 0 à 40 ans en 2015 du modèle de LEE & CARTER

— De 40 à 60 ans :

Les modèles de LEE & CARTER et LOG POISSON présentent des résultats similaires. Pour les femmes, ces deux modèles prédisent les taux de mortalité de manière cohérente avec les intervalles de confiance, bien qu'une surestimation soit observée en 2023. Chez les hommes, de 1991 à 2013, plusieurs prédictions se situent en dehors de l'intervalle de confiance, mais l'ajustement s'améliore après 2013, avec cependant une surestimation des décès en 2023. Le modèle de RENSHAW & HABERMAN se distingue en offrant un ajustement fiable pour les femmes, y compris en 2023, tandis que pour les hommes, il présente des résultats similaires à ceux des modèles de LEE & CARTER et LOG POISSON. Les graphiques ci-dessous présentent les intervalles de confiance en 2023 pour les femmes avec le modèle de LEE & CARTER et le modèle de RENSHAW & HABERMAN.



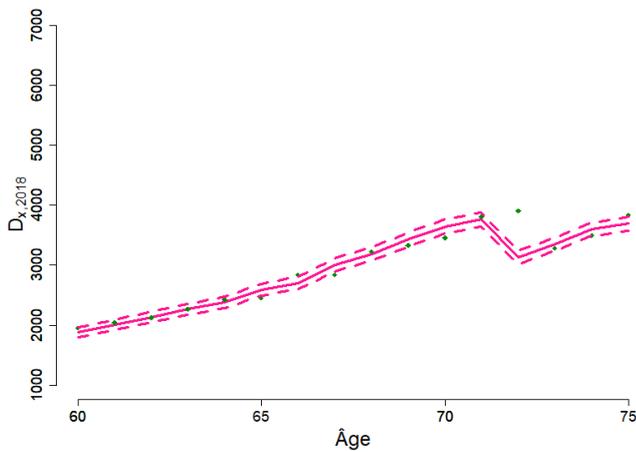
LEE & CARTER



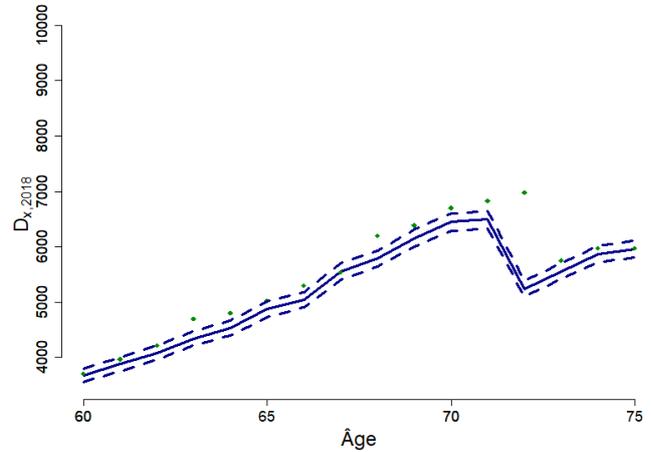
RENSHAW & HABERMAN

FIGURE 2.3 – Les intervalles de confiance de 40 à 60 ans en 2023 pour les femmes

— **De 60 à 75 ans** : Entre 60 et 75 ans, les trois modèles montrent des comportements similaires. Pour les femmes, les prédictions sont cohérentes avec l'intervalle de confiance de 1991 à 2002 et s'alignent de nouveau étroitement avec les observations entre 2013 et 2020. Pour les hommes, une nette amélioration des prédictions est visible à partir de 2013, avec un alignement satisfaisant entre 2013 et 2020. Les graphiques ci-dessous illustrent ces résultats avec les intervalles de confiance pour l'année 2018, selon le modèle de RENSCHAW & HABERMAN.



Femmes



Hommes

FIGURE 2.4 – Les intervalles de confiance de 60 à 75 ans en 2018 pour le modèle de RENSCHAW & HABERMAN

— **De 75 à 90 ans (85 ans pour les hommes)** : Entre 75 et 90 ans pour les femmes, et entre 75 et 85 ans pour les hommes, les trois modèles montrent des prédictions généralement très proches des intervalles de confiance, bien qu'il existe quelques points isolés en dehors de ces intervalles pour certains âges et années.

2.3 L'étude des résidus

L'objectif de cette étude est d'identifier si certaines caractéristiques ont été mal évaluées localement. En d'autres termes, il s'agit de détecter la présence de grands résidus. Les résidus suivants sont examinés :

— **Les résidus de la réponse :**

$$r_{x,t} = q_{x,t} - \hat{q}_{x,t}$$

Avec :

- $q_{x,t}$, les taux de mortalité bruts à l'âge x et au temps t .
- $\hat{q}_{x,t}$, les taux de mortalité prédits à l'âge x et au temps t .

L'analyse des résidus vise à identifier toute tendance systématique qui pourrait indiquer que le modèle ne s'ajuste pas de manière optimale à certaines plages d'âge ou périodes spécifiques.

L'absence de tendance dans les résidus est un signe que le modèle est bien ajusté et qu'il n'y a pas de biais significatif dans les prédictions. En revanche, la présence de tendances dans les résidus, notamment sur certaines plages d'âge ou périodes, peut indiquer des problèmes d'ajustement, suggérant que le modèle pourrait sous-estimer ou surestimer les taux de mortalité pour ces groupes.

Tout au long de la période d'étude, des fluctuations aléatoires sont observées chez les femmes et les hommes aux âges les plus élevés pour les trois modèles. Cela démontre l'efficacité de ces approches pour capturer les dynamiques de mortalité et les incertitudes associées. Les graphiques ci-dessous illustrent ces observations pour le modèle de LEE & CARTER.

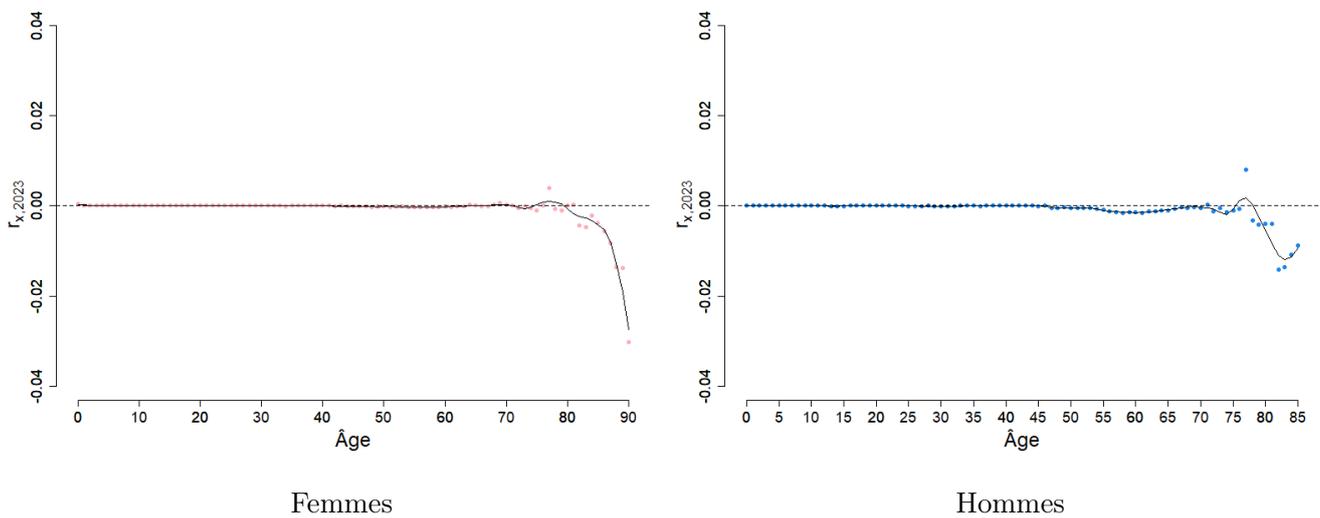


FIGURE 2.5 – Les résidus de la réponse par âge en 2023 pour le modèle de LEE & CARTER

— **Les résidus de Pearson :**

$$r_{x,t} = \frac{D_{x,t} - \hat{D}_{x,t}}{\sqrt{\text{Var}(\hat{D}_{x,t})}}$$

Avec :

- $D_{x,t}$, le nombre de décès observés à l'âge x , au temps t .
- $\hat{D}_{x,t}$, le nombre de décès prédits à l'âge x , au temps t .

L'analyse des résidus de Pearson permet d'identifier les observations qui s'écartent de manière significative des prédictions du modèle, en standardisant ces écarts pour faciliter une comparaison uniforme entre les différentes observations.

L'objectif principal de cette analyse est de détecter les tendances ou structures non capturées par le modèle. Si les résidus sont distribués de manière aléatoire autour de zéro, sans tendance discernable, cela suggère que le modèle ajuste bien les données. En revanche, la présence de tendances systématiques dans les résidus, en particulier s'ils sortent de l'intervalle $[-2, 2]$, peut indiquer des problèmes d'ajustement, tels que des biais dans le modèle ou des variables explicatives manquantes.

Pour l'ensemble des modèles, au début de l'étude, les résidus des femmes montrent quelques dépassements de l'intervalle de confiance, similaires aux observations faites avec les autres modèles. Avec le temps, les résidus se rapprochent davantage de l'intervalle et y restent principalement, sauf pour les âges au-delà de 70 ans, où quelques dépassements sont encore présents. Pour les hommes, les résidus se situent fréquemment en dehors de l'intervalle de confiance au début de l'étude, mais ils deviennent globalement plus conformes à l'intervalle au fil des années, même si certains dépassements persistent, notamment après 50 ans. Les graphiques suivants illustrent cette notion pour le modèle de RENSHAW & HABERMAN.

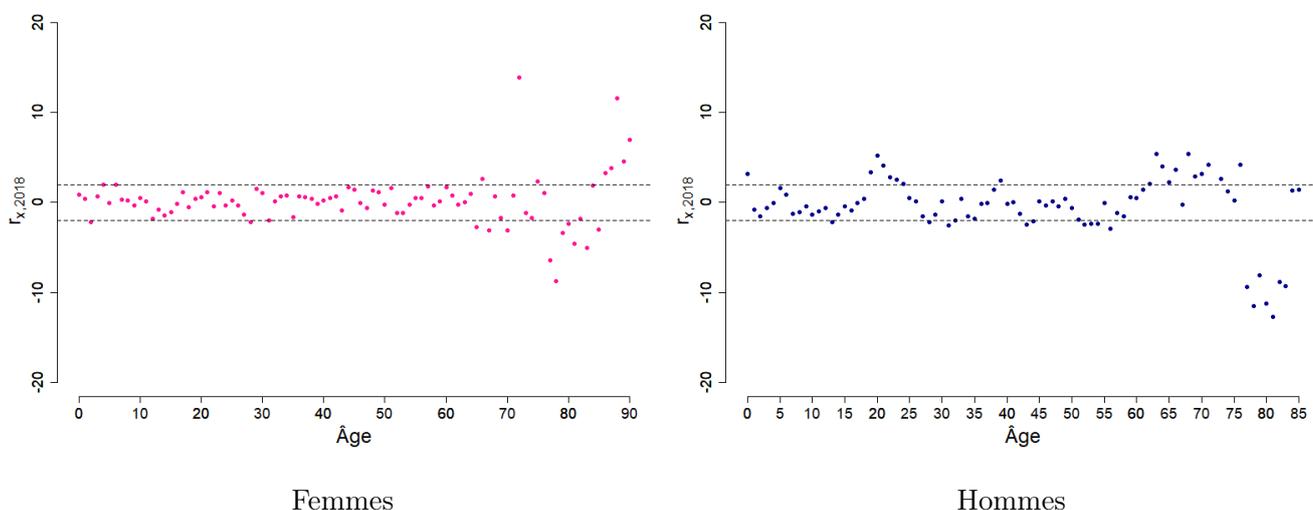


FIGURE 2.6 – Les résidus de la réponse par âge en 2018 pour le modèle de RENSHAW & HABERMAN

À présent, pour mieux visualiser la distribution et les variations des résidus, ceux-ci sont représentés sous forme de heatmaps. Cette méthode permet de capturer et de mettre en évidence les zones où les écarts entre les valeurs observées et prédites sont les plus marqués. En outre, les heatmaps aident à détecter la présence potentielle d'effets de cohorte, en observant la répartition des résidus à travers les générations. Idéalement, les résidus devraient être aléatoires, indiquant un bon ajustement du modèle. Tout motif récurrent pourrait suggérer des insuffisances du modèle ou des effets de cohorte non capturés.

- **LEE & CARTER** : Les résidus obtenus avec le modèle de LEE & CARTER présentent des structures diagonales, tant pour les femmes que pour les hommes. Ces structures suggèrent que le modèle a du mal à capturer certaines dynamiques intergénérationnelles ou des variations spécifiques aux cohortes, conduisant à des résidus non aléatoires.

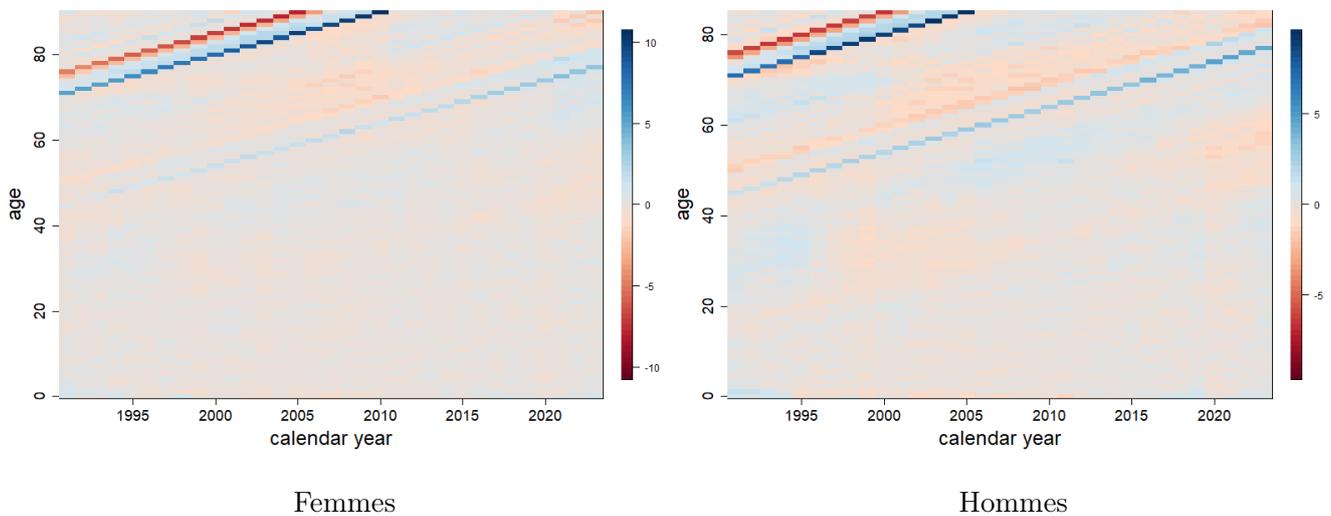


FIGURE 2.7 – Heatmap des résidus de Pearson du modèle de LEE & CARTER

- **LOG POISSON** : Les résultats obtenus avec le modèle LOG POISSON montrent des caractéristiques similaires à celles observées avec le modèle de LEE & CARTER. Les résidus présentent également des structures diagonales, tant pour les femmes que pour les hommes.

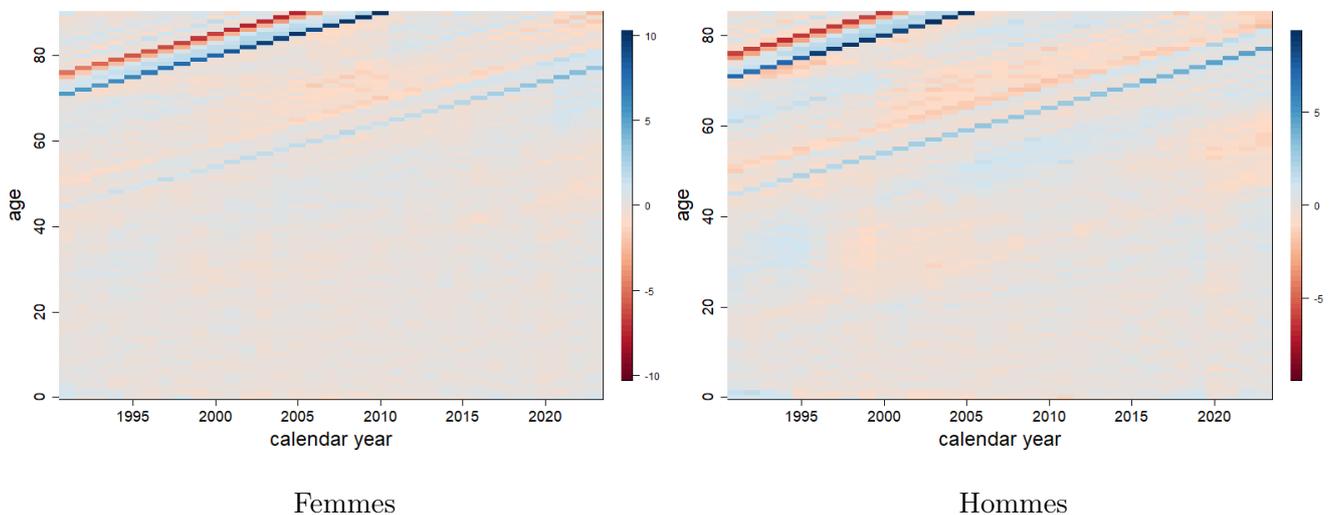


FIGURE 2.8 – Heatmap des résidus de Pearson du modèle LOG POISSON

- **RENSHAW & HABERMAN** : Les résidus obtenus avec le modèle de RENSRAW & HABERMAN présentent une structure différente de celle observée avec les modèles de LEE & CARTER et LOG POISSON. Contrairement aux deux premiers modèles, les heatmaps de résidus obtenus avec le modèle de RENSRAW & HABERMAN montrent une distribution plus aléatoire, suggérant un bon ajustement du modèle.

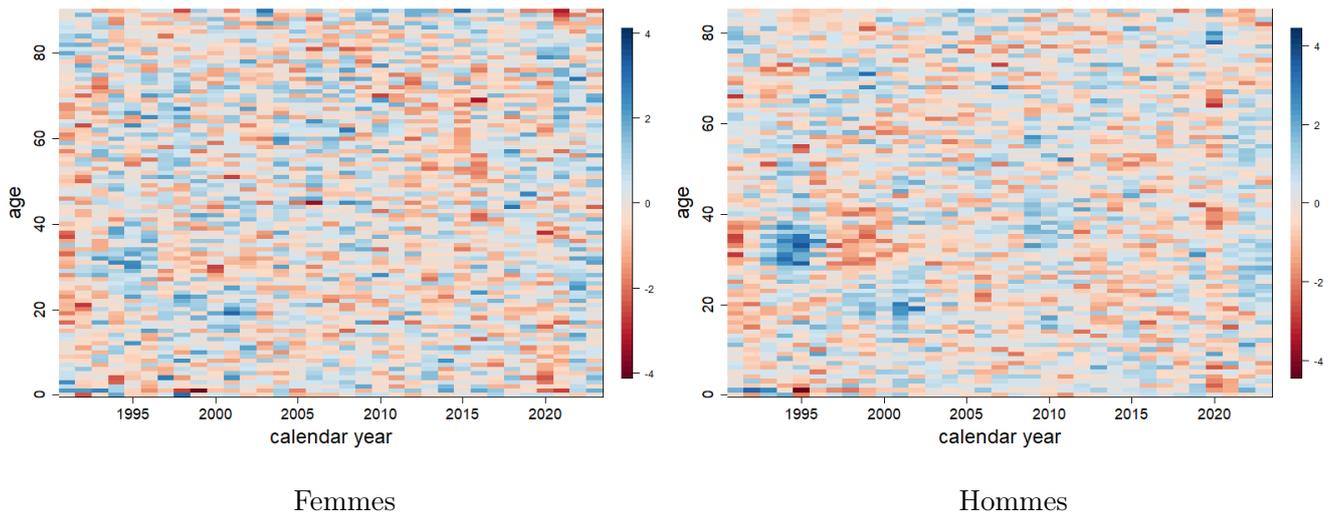


FIGURE 2.9 – Heatmap des résidus de Pearson du modèle RENSHAW & HABERMAN

2.4 L'étude du *Standardised Mortality Ratio*

L'analyse des ratios de mortalité standardisés, appelés SMR (*Standardised Mortality Ratio*), constitue un indicateur essentiel pour évaluer la qualité des prédictions de mortalité effectuées par les différents modèles. Le SMR permet de comparer de manière directe les décès observés à ceux prédits par le modèle.

Cet indicateur est défini par :

$$SMR_{x,t} = \frac{\text{Nombre de décès observés à l'âge } x \text{ et au temps } t}{\text{Nombre de décès prédits à l'âge } x \text{ et au temps } t}$$

$$SMR_{x,t} = \frac{D_{x,t}}{E_{x,t}\hat{q}_{x,t}} = \frac{D_{x,t}}{\hat{D}_{x,t}}$$

avec :

- $\hat{q}_{x,t}$, le taux de décès prédits à l'âge x et au temps t .
- $E_{x,t}$, l'exposition soumise au risque à l'âge x et au temps t .
- $D_{x,t}$, le nombre de décès observés à l'âge x et au temps t .
- $\hat{D}_{x,t}$, le nombre de décès prédits à l'âge x et au temps t .

L'interprétation du $SMR_{x,t}$ est particulièrement intuitive et permet d'évaluer directement la performance des modèles :

- Un $SMR_{x,t}$ **égal à 1** signifie qu'à l'âge x et au temps t , le nombre de décès prédits par le modèle correspond exactement au nombre de décès observés, indiquant une capacité de prédiction parfaite. Toutefois, cette situation représente un cas théorique idéal, rarement atteint en pratique.
- Un $SMR_{x,t}$ **supérieur à 1** révèle une sous-estimation de la mortalité par le modèle. Cela signifie qu'à l'âge x et au temps t , le nombre de décès observés dépasse les prédictions, suggérant que le modèle sous-évalue les risques de mortalité.

- Un $SMR_{x,t}$ inférieur à 1 signale une surestimation de la mortalité. Dans ce cas, à l'âge x et au temps t , le modèle prédit plus de décès qu'il en a été observé, ce qui peut indiquer une surévaluation des risques de mortalité par le modèle.

L'objectif principal de cette analyse est d'examiner si les SMR calculés par âge, par année calendaire et au global, se rapprochent de 1, sans toutefois les dépasser de manière significative. Un SMR proche de 1 mais inférieur signale une bonne prédiction avec une légère prudence.

- **L'étude des ratios par âge :** L'analyse des ratios par âge montre que dans l'ensemble les SMR des trois modèles sont très proches de 1, ce qui rend difficile de différencier les prédictions sur la base de ce critère. Les variations observées sont principalement attribuables à la variabilité des données plutôt qu'à des différences entre les modèles. Les fluctuations avant 30 ans, bien que perceptibles, ne sont pas particulièrement préoccupantes car elles ne concernent que 3,8% des décès pour les hommes et seulement 2% pour les femmes. Un décrochage se remarque après 60 ans pour les hommes et 80 ans pour les femmes, différenciant ainsi le modèle de LEE & CARTER des modèles de RENSHAW & HABERMAN et LOG POISSON.

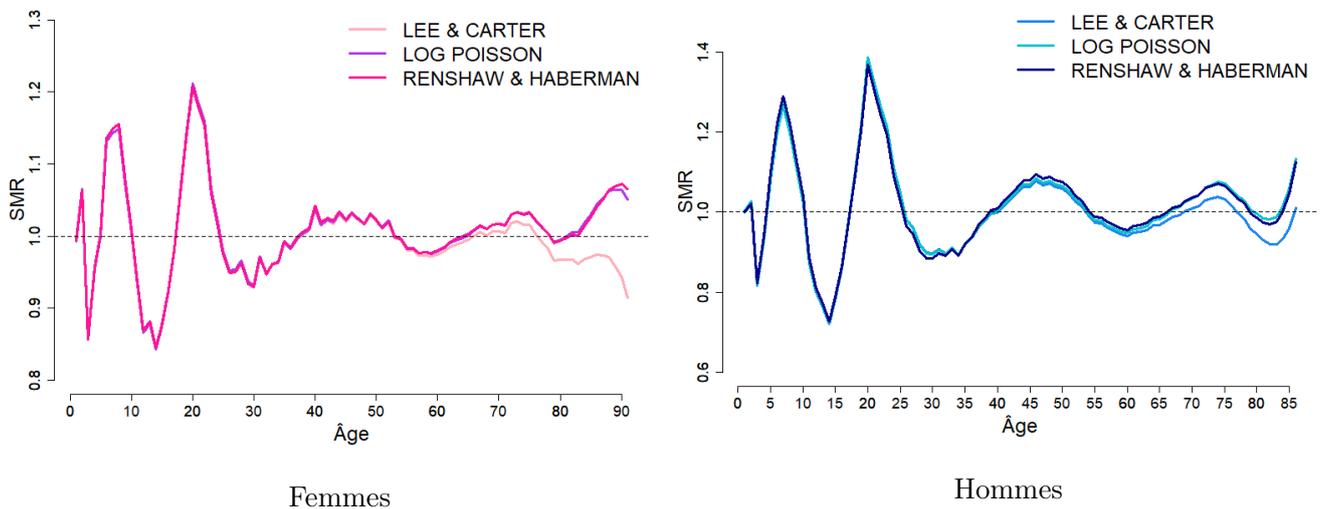


FIGURE 2.10 – Les SMR par âge, toutes années confondues

- **L'étude des ratios par année :** Pour les hommes et les femmes, les ratios de mortalité standardisés (SMR) restent généralement proches de 1 tout au long des années, quel que soit le modèle utilisé, indiquant une bonne correspondance globale entre décès observés et prédits. Toutefois, le modèle de LEE & CARTER se distingue légèrement des autres en présentant des SMR inférieurs à 1, suggérant une légère tendance à surestimer la mortalité. En revanche, le modèle LOG POISSON et celui de RENSHAW & HABERMAN montrent une légère sous-estimation de la mortalité, avec des SMR supérieurs à 1, bien qu'ils restent très proches de cette valeur.

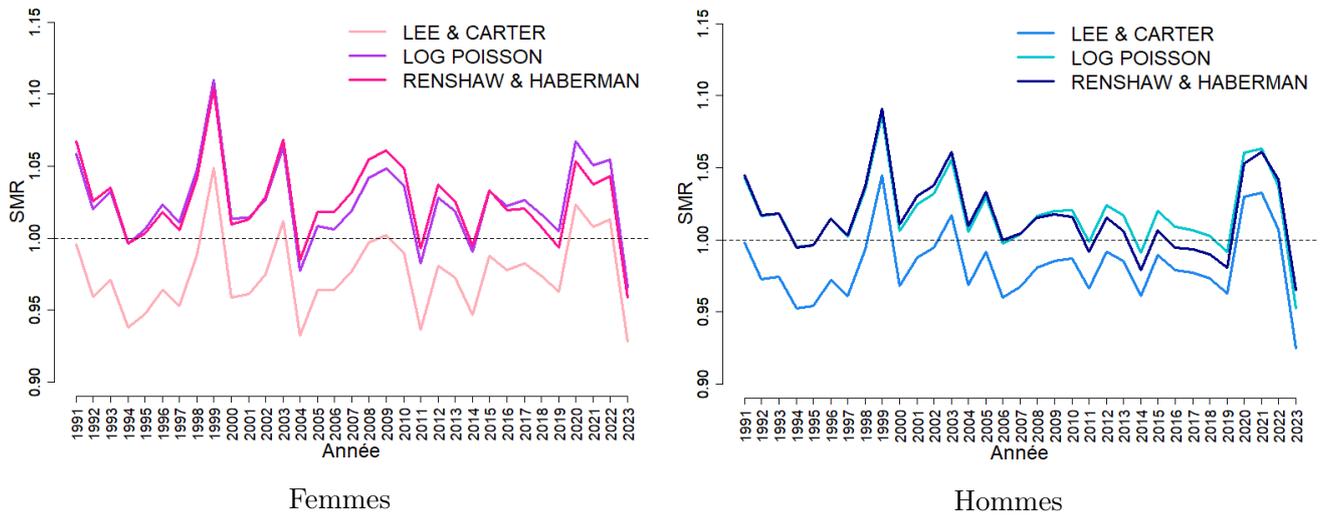


FIGURE 2.11 – Les SMR par année, tous âges confondus

— L'étude des ratios globaux :

	Femmes	Hommes
LEE & CARTER	0,9754	0,9821
LOG POISSON	1,0249	1,018
RENSHAW & HABERMAN	1,0255	1,0163

TABLE 2.1 – Les SMR globaux, tous âges et toutes années confondus

Pour les hommes comme pour les femmes, les valeurs de SMR obtenues avec les trois modèles sont très proches de 1, rendant difficile de distinguer leurs prédictions sur ce critère. Les écarts constatés sont très faibles : le modèle de LEE & CARTER affiche des SMR légèrement inférieurs à 1, tandis que les modèles LOG POISSON et RENSHAW & HABERMAN présentent des SMR légèrement supérieurs à 1, indiquant une sous estimation de la mortalité d'environ 2%. Ces faibles variations montrent que, dans l'ensemble, les résultats des trois modèles sont très similaires.

2.5 L'étude du χ^2

L'adéquation entre les données observées et les données prédites est évaluée en comparant les valeurs du χ^2 . Cet indicateur mesure la qualité de l'ajustement des modèles en comparant la distribution des décès observés avec celle des décès prédits. La valeur du χ^2 , se définit comme suit :

$$\chi^2 = \sum_{x,t} \frac{(D_{x,t} - \hat{D}_{x,t})^2}{\hat{D}_{x,t}}$$

avec :

- $D_{x,t}$, le nombre de décès observés à l'âge x et au temps t .
- $\hat{D}_{x,t}$, le nombre de décès estimés à l'âge x et au temps t .

L'objectif de cette étude est de déterminer si un modèle s'ajuste mieux, par âge, par année et au global, aux données en présentant des valeurs de χ^2 plus petites.

- **L'étude des χ^2 par âge :** L'analyse des valeurs de χ^2 par âge montre que, pour les femmes comme pour les hommes, les trois modèles présentent des valeurs similaires qui augmentent de manière exponentielle avec l'âge, suggérant des erreurs de prédiction plus importantes à partir de 70 ans. Pour les hommes, un léger pic est observé autour de 20 ans pour les trois modèles. En conséquence, il est difficile de conclure à la supériorité d'un modèle par rapport aux autres sur la base de cette analyse.

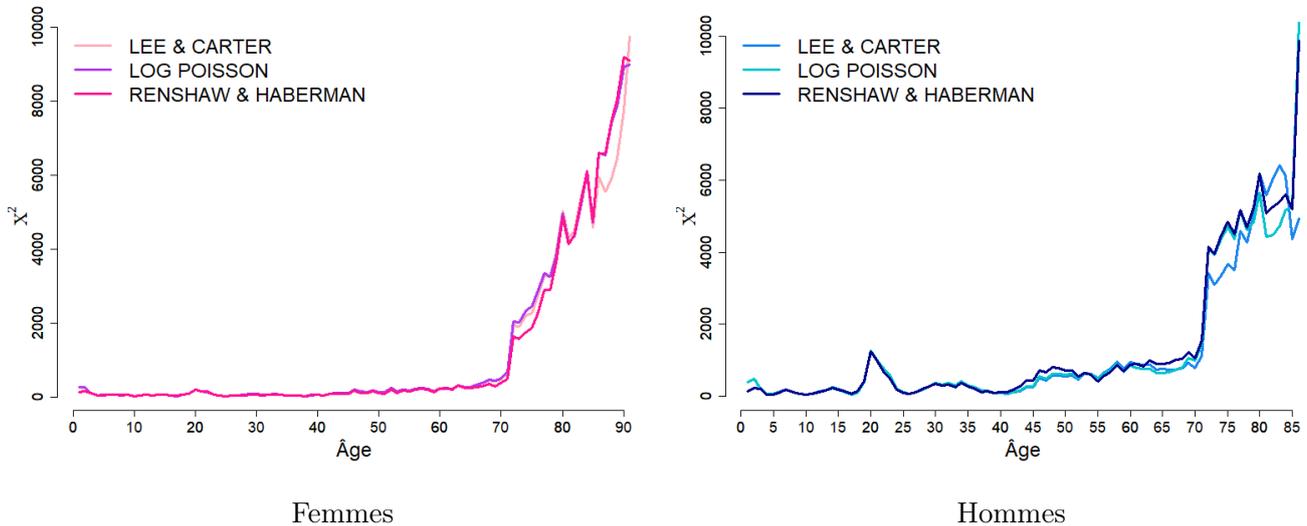


FIGURE 2.12 – Les valeurs du χ^2 par âge, toutes années confondues

- **L'étude des χ^2 par année :** Les trois modèles affichent des valeurs de χ^2 globalement similaires, avec des légères variations d'une année à l'autre. Bien que certaines différences se manifestent, notamment dans les années 1990 et 2000, il reste difficile de déterminer la supériorité d'un modèle par rapport aux autres sur la base de ces valeurs de χ^2 . Dans l'ensemble, les trois modèles semblent offrir des performances similaires en termes de prédiction de la mortalité au fil du temps.

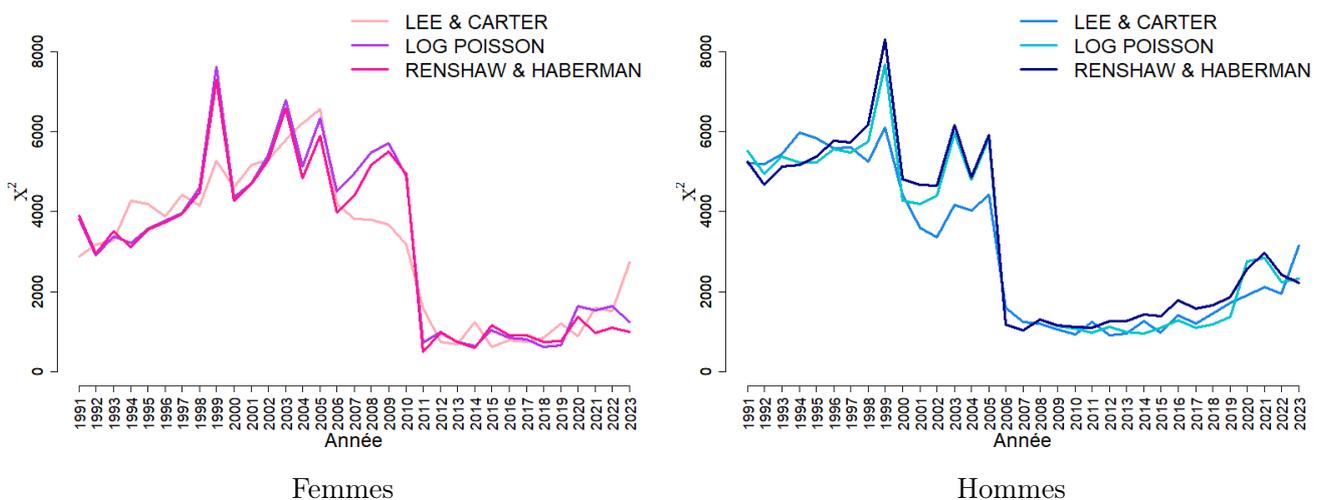


FIGURE 2.13 – Les valeurs du χ^2 par année, tous âges confondus

— L'étude des χ^2 globaux :

	Femmes	Hommes
LEE & CARTER	103 028	100 477
LOG POISSON	108 292	106 292
RENSHAW & HABERMAN	103 806	112 008

TABLE 2.2 – Les valeurs globales du χ^2

L'analyse des valeurs globales de χ^2 présentées dans le tableau ci-dessus montre que, pour les femmes, le modèle de LEE & CARTER affiche la valeur la plus faible de χ^2 , suivi de près par le modèle de RENSHAW & HABERMAN, tandis que le modèle LOG POISSON présente une valeur légèrement plus élevée. Pour les hommes, les écarts de χ^2 entre les modèles sont également réduits, ce qui rend difficile de distinguer les modèles sur la base de ce critère seul. Ces faibles différences suggèrent que les trois modèles ont des performances globalement comparables en termes d'ajustement.

2.6 L'étude de l'erreur moyenne des prédictions

Le *Mean Absolute Percentage Error* (MAPE) est couramment utilisée pour évaluer l'exactitude des prévisions. Cette métrique, exprimée en pourcentage, offre une interprétation claire et intuitive de l'écart entre les prévisions et les valeurs réellement observées. Elle correspond à la moyenne des écarts entre les décès prédits et les décès observés, en valeur absolue :

$$MAPE = \frac{1}{n} \sum_{x,t} \frac{|D_{x,t} - \hat{D}_{x,t}|}{D_{x,t}} \times 100$$

où :

- $D_{x,t}$ est le nombre de décès observés pour un âge x et une année t .
- $\hat{D}_{x,t}$ est le nombre de décès prédits pour un âge x et une année t .
- n est le nombre total d'observations.

L'objectif de cette analyse est de comparer les estimations obtenues à partir des différents modèles par âge, par année et au global.

- **L'étude du MAPE par âge :** Pour les femmes comme pour les hommes, les erreurs de prédiction relatives sont plus marquées aux âges les plus jeunes et dans une moindre mesure aux âges les plus avancés, tandis qu'elles sont plus faibles pour les âges intermédiaires, 30-70 ans. Dans l'ensemble, les trois modèles présentent des performances similaires, avec un léger avantage pour le modèle de RENSHAW & HABERMAN, en particulier pour les femmes à certains âges. Néanmoins, ces écarts sont faibles, ce qui rend difficile de conclure à la supériorité d'un modèle par rapport aux autres sur la base des valeurs du MAPE.

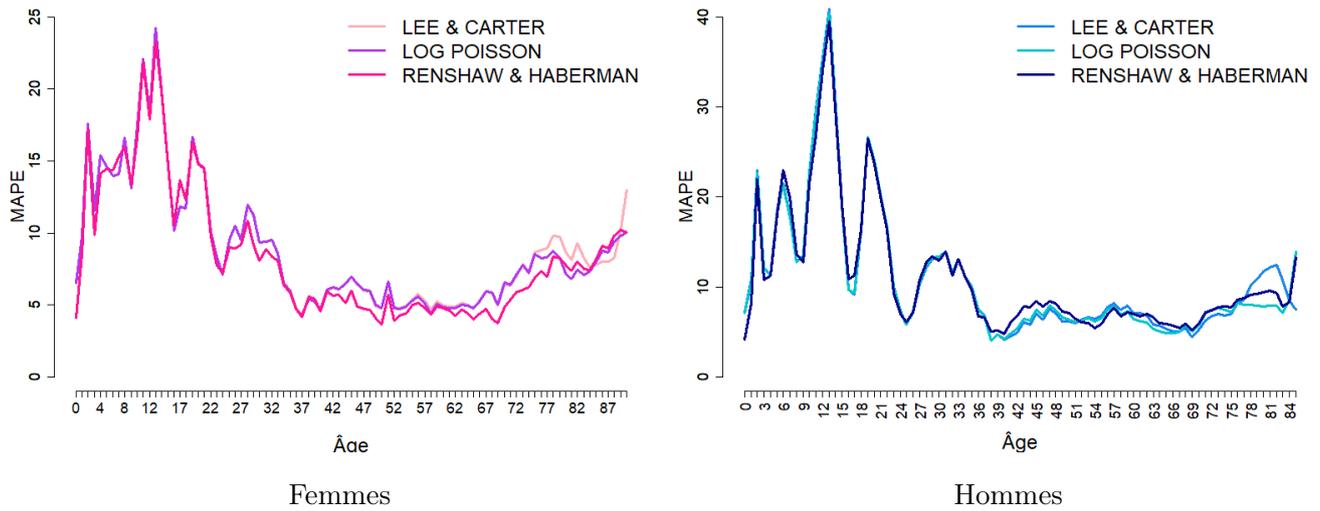


FIGURE 2.14 – Les valeurs du MAPE par âge, toutes années confondues

— **L'étude du MAPE par année** : Les estimations des trois modèles sont très proches, avec des différences peu marquées tout au long de la période. Le modèle de RENSRAW & HABERMAN se montre parfois légèrement meilleur, notamment entre 2000 et 2010, ainsi qu'à partir de 2021. Cependant, ces écarts sont minimes et ne permettent pas de conclure à une véritable supériorité d'un modèle sur les autres.

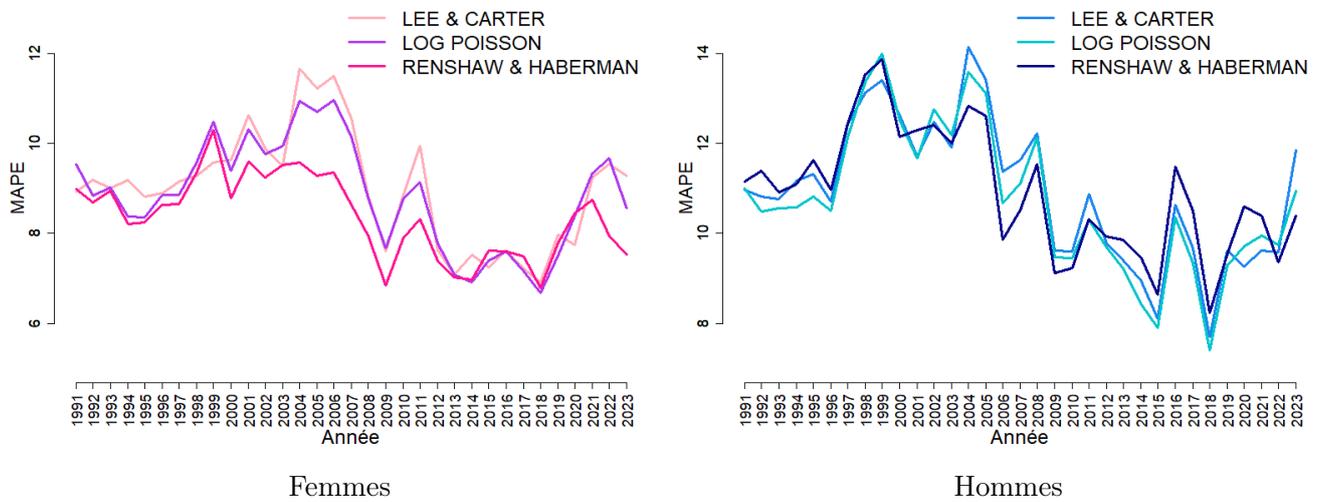


FIGURE 2.15 – Les valeurs du MAPE par année, tous âges confondus

— **L'étude des MAPE globaux** :

	Femmes	Hommes
LEE & CARTER	8,999%	10,926%
LOG POISSON	8,865%	10,742%
RENSRAW & HABERMAN	8,377%	10,919%

TABLE 2.3 – Les valeurs globales du MAPE

Les valeurs globales du MAPE montrent que les résultats des trois modèles sont très proches, avec des différences minimales. Pour les femmes, le modèle de RENSHAW & HABERMAN présente la plus faible valeur (8,377%), suivi de près par LOG POISSON (8,865%) et LEE & CARTER (8,999%). Pour les hommes, les erreurs sont légèrement plus élevées mais restent également proches entre les modèles (de 10,742% à 10,926%). Ces résultats suggèrent que les trois modèles fournissent globalement de bonnes prédictions, avec des erreurs relativement faibles, rendant difficile de conclure à la supériorité de l'un par rapport aux autres sur la base de cette étude.

2.7 L'étude du coefficient de détermination

L'adéquation entre les prédictions des modèles et les observations réelles est évaluée à l'aide du coefficient de détermination R^2 . Ce coefficient, compris entre 0 et 1, représente la proportion de la variance totale des observations qui est expliquée par le modèle. Plus la valeur du R^2 est proche de 1, plus le modèle s'ajuste bien aux données, indiquant une forte corrélation entre les valeurs observées et les valeurs prédites. À l'inverse, un R^2 faible suggère que le modèle laisse une grande part de l'incertitude inexpliquée. Le coefficient R^2 est défini par la formule suivante :

$$R^2 = 1 - \left(\frac{\sum_{x,t} (q_{x,t} - \hat{q}_{x,t})^2}{\sum_{x,t} (q_{x,t} - (\sum_{x,t} \frac{q_{x,t}}{n}))^2} \right)$$

Avec :

- $q_{x,t}$ les taux de mortalité bruts à l'âge x et à l'année t .
- $\hat{q}_{x,t}$ les taux de mortalité estimés à l'âge x et à l'année t . n le nombre total d'observations.

L'objectif de cette analyse est de comparer les différents modèles en termes de précision des prévisions de mortalité par âge, par année et au global.

- **L'étude du R^2 par âge :** Les valeurs de R^2 pour les deux sexes montrent que les trois modèles fournissent un bon ajustement des données de mortalité, bien que la précision soit légèrement réduite pour les âges très jeunes et, dans une moindre mesure, pour les âges très avancés. Les différences entre les modèles sont faibles, rendant difficile de déterminer la supériorité d'un modèle sur la base de cette étude.

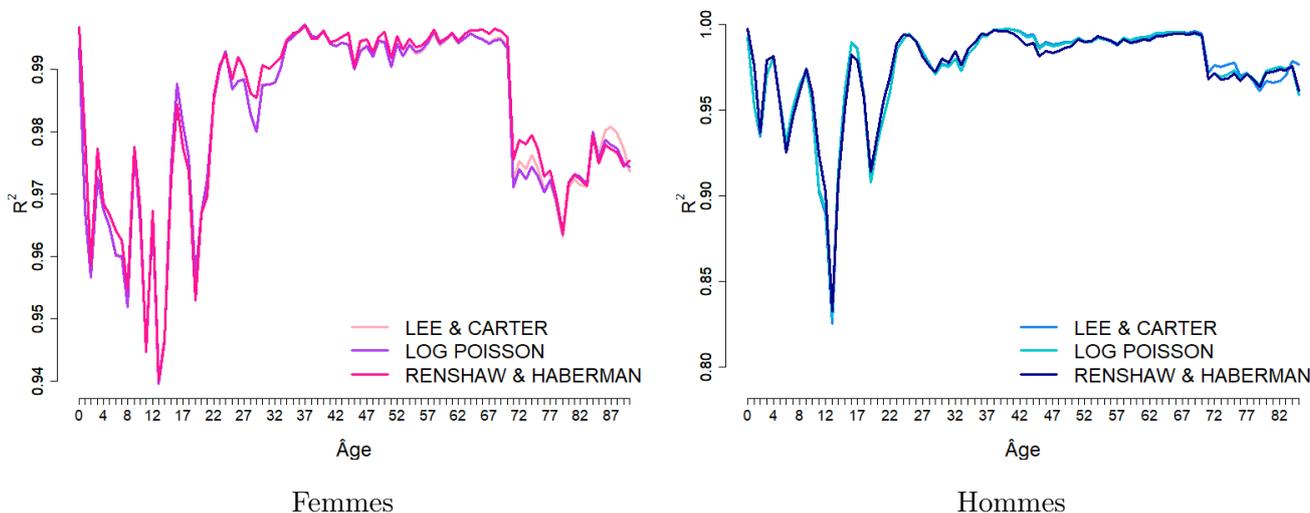


FIGURE 2.16 – Les valeurs du R^2 par âge, toutes années confondues

- **L'étude du R^2 par année :** Les trois modèles offrent globalement de très bonnes prédictions, avec des résultats proches de 1. Des variations mineures existent entre le modèle de LEE & CARTER et les autres : parfois, il présente des résultats légèrement meilleurs, et à d'autres moments, ce sont les modèles LOG POISSON et RENSHAW & HABERMAN qui montrent une légère supériorité. Depuis 2010 pour les femmes et 2007 pour les hommes, les résultats des trois modèles sont particulièrement similaires, rendant difficile de conclure sur la supériorité de l'un par rapport aux autres.

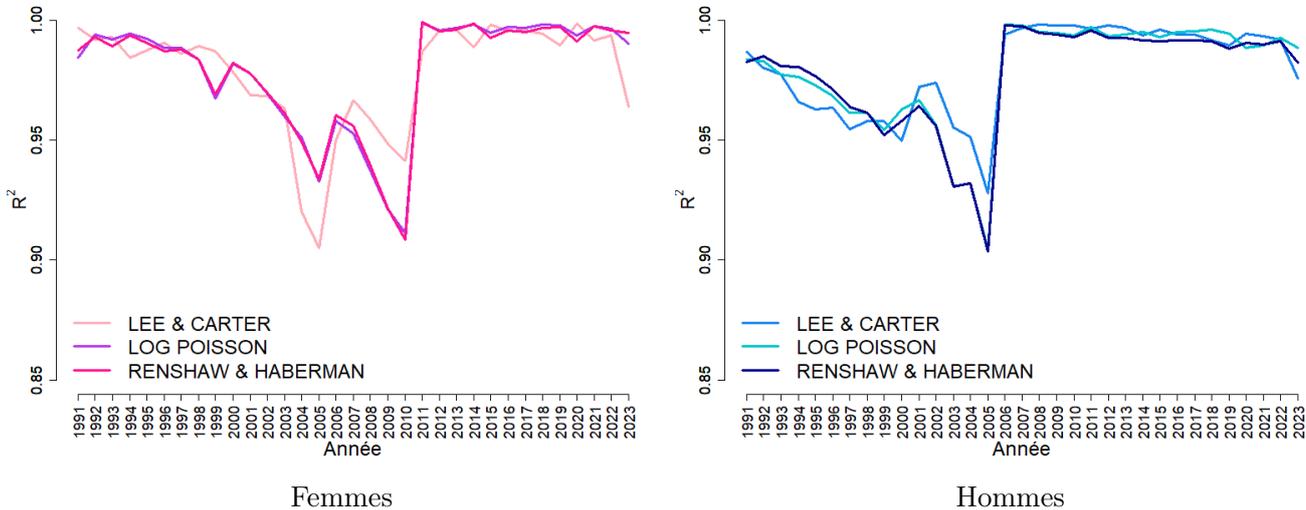


FIGURE 2.17 – Les valeurs du R^2 par année, tous âges confondus

- **L'étude des R^2 globaux :**

	Femmes	Hommes
LEE & CARTER	0,978	0,975
LOG POISSON	0,977	0,973
RENSHAW & HABERMAN	0,977	0,973

TABLE 2.4 – Les valeurs globales du R^2

L'étude des valeurs globales du coefficient de détermination R^2 , présentée dans le tableau ci-dessus, montre que pour les femmes comme pour les hommes, les résultats des trois modèles sont très proches de 1. Ainsi, ces modèles se révèlent globalement efficaces pour expliquer la variance observée dans les données, assurant une très bonne adéquation entre les prévisions et les observations.

2.8 L'étude du Critère d'Information Bayésien

Cette section se concentre sur l'évaluation des modèles à l'aide du Critère d'Information Bayésien (BIC). Cet indicateur est essentiel pour comparer la qualité des ajustements des différents modèles tout en tenant compte de leur complexité. Le BIC cherche à identifier le modèle qui offre le meilleur compromis entre l'ajustement des données et la simplicité du modèle, en pénalisant les modèles trop complexes.

Le BIC est calculé selon la formule suivante :

$$BIC = -2\log(L) + k\log(n)$$

où :

- $\log(L)$ est le logarithme de la fonction de vraisemblance maximale du modèle, représentant la qualité de l'ajustement aux données.
- k est le nombre de paramètres du modèle qui mesure la complexité du modèle.
- n est le nombre d'observations dans les données.

Plus la valeur du BIC est basse, meilleur est l'équilibre entre la qualité de l'ajustement et la complexité du modèle. Ainsi, le modèle ayant la plus petite valeur de BIC est considéré comme le plus performant.

	Femmes	Hommes
LEE & CARTER	115 603	102 418
LOG POISSON	109 983	98 040
RENSHAW & HABERMAN	30 910	30 926

TABLE 2.5 – Les valeurs du BIC

L'étude des valeurs du BIC, présentée dans le tableau ci-dessus, met en évidence l'efficacité du modèle de RENSRAW & HABERMAN dans la modélisation des données de mortalité, avec des valeurs de BIC nettement les plus faibles.



Ce qu'il faut retenir du chapitre :

Dans le processus de construction de tables de mortalité, valider et contrôler les estimations est une étape déterminante. Il est nécessaire de prédire le plus précisément possible les observations.

- * Les trois modèles étudiés fournissent de bonnes prédictions, avec des estimations très proches des observations. Ces résultats confirment l'intérêt de ces modèles pour l'analyse de la mortalité.
- * Les résultats des modèles sont très similaires pour plusieurs critères d'évaluation, ce qui rend leur distinction difficile sur ces bases. Cependant, le modèle de RENSRAW & HABERMAN se distingue par son Critère d'Information Bayésien (BIC) plus faible, indiquant un bon compromis entre complexité et précision des prédictions.
- * Les heatmaps des résidus montrent que seul le modèle de RENSRAW & HABERMAN présente des résidus aléatoires, ce qui met en évidence l'intérêt d'introduire un terme de cohorte dans les études de mortalité. Cette propriété renforce la pertinence de ce modèle pour capter les dynamiques intergénérationnelles et les effets de cohorte.

Le modèle de RENSRAW & HABERMAN est retenu pour la suite de ce mémoire, car il s'ajuste étroitement aux données observées et parvient à bien capturer les effets de cohorte, tout en offrant un bon compromis entre complexité et précision des prédictions.

Le prochain chapitre se concentrera sur la projection de sa composante temporelle, ce qui permettra d'obtenir des taux de mortalité projetés et d'explorer leurs applications futures.

Chapitre 3

La projection des taux de mortalité

Une compréhension approfondie des tendances passées et présentes, ainsi que la capacité à projeter les taux de mortalité et à anticiper les évolutions du portefeuille, est au cœur de l'actuariat. Adapter les stratégies de gestion des risques en fonction des tendances futures garantit ainsi une stabilité et une performance optimale des engagements financiers.

Classiquement, dans les modèles prospectifs, la composante temporelle est assimilée à un processus stochastique, en particulier à une marche aléatoire avec dérive. Cette approche est un cas particulier de modèle de type *AutoRegressive Integrated Moving Average* (*ARIMA*), précisément un modèle *ARIMA*(0, 1, 0) avec une dérive.

Ce chapitre a pour objectif d'appliquer ce modèle de projection, classiquement utilisé, à la composante κ_t du modèle de RENSHAW & HABERMAN.

3.1 Les processus *ARIMA*(p, d, q)

3.1.1 Définition

Les processus *AutoRegressive Integrated Moving Average* (*ARIMA*) sont largement utilisés pour l'analyse et la prévision des séries chronologiques. Ils permettent d'estimer les paramètres à partir des données historiques pour générer des trajectoires futures de la composante temporelle, et ainsi projeter les taux de mortalité. Ces processus sont caractérisés par trois composantes : l'autorégression (*AR*), la différenciation (*I*) et la moyenne mobile (*MA*). En combinant ces trois composantes, les processus *ARIMA* fournissent ainsi un cadre robuste pour l'analyse des séries temporelles, permettant l'obtention des prévisions précises et fiables.

- **La composante autorégressive** (*AR*) du modèle *ARIMA* se base sur l'idée que la valeur d'une série temporelle à un instant donné t , X_t , peut être exprimée comme une combinaison linéaire des p valeurs passées. Mathématiquement, un processus autorégressif d'ordre p , noté *AR*(p), s'exprime par :

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$$

où :

- ϕ_1, \dots, ϕ_p sont les coefficients autorégressifs qui déterminent l'influence des valeurs passées sur la valeur actuelle.
- X_{t-1}, \dots, X_{t-p} sont les p valeurs passées de la série temporelle.
- μ est la constante, aussi appelée dérive, qui capture la tendance générale de la série.

- ϵ_t est le terme d'erreur représentant la composante aléatoire non expliquée par les valeurs passées.

— **La composante moyenne mobile (MA)** modélise la série temporelle à un instant donné t , X_t , comme une fonction des q erreurs de prévision précédentes. Le modèle $MA(q)$ est défini comme :

$$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q}$$

où :

- $\theta_1, \dots, \theta_q$ sont les coefficients de moyenne mobile qui déterminent l'impact des erreurs passées sur la valeur actuelle.
- $\epsilon_{t-1}, \dots, \epsilon_{t-q}$ sont les q erreurs passées.
- μ est une constante.
- ϵ_t est le terme d'erreur.

— **La partie d'intégration (I)** implique la différenciation de la série temporelle pour la rendre stationnaire. Un processus, X_t est stationnaire si :

- $\mathbb{E}[X_t]$ est indépendante de t .
- $cov(X_{t+h}, X_t)$ est indépendante de t , pour tout h .

L'ordre de différenciation d correspond au nombre de fois que la série doit être différenciée pour obtenir une série stationnaire.

Ainsi, un processus stochastique $(X_t)_{t \geq -p-d}$ est un processus $ARIMA(p, d, q)$ s'il combine p termes autorégressifs, d différenciations pour rendre le processus stationnaire et q termes de moyennes mobiles. Autrement dit, il est défini comme suit :

$$\Delta^d X_t = \mu + \phi_1 \Delta^d X_{t-1} + \dots + \phi_p \Delta^d X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

Où $\Delta^d X_t$ représente la série X_t différenciée à l'ordre d .

Le cas particulier d'un $ARIMA(0, 1, 0)$ avec dérive :

La marche aléatoire avec dérive est un cas particulier de modèle ARIMA, spécifiquement un modèle $ARIMA(0, 1, 0)$ avec une constante μ qui représente la dérive. Ce modèle est particulièrement utile pour les séries temporelles qui présentent une tendance constante à long terme. Mathématiquement, la marche aléatoire avec dérive est exprimée comme suit :

$$X_t = \mu + X_{t-1} + \epsilon_t$$

où :

- μ est la dérive, représentant la tendance moyenne de la série.
- X_{t-1} est la valeur de la série à l'instant précédent.
- ϵ_t est le terme d'erreur aléatoire.

3.1.2 La méthodologie BOX & JENKINS

La modélisation des processus *ARIMA* nécessite une approche méthodique pour garantir une précision et une adéquation aux données. C'est dans ce contexte que l'ouvrage de BOX & JENKINS, *Time Series Analysis, Forecasting and Control*, publié en 1970, a proposé une démarche de prévision pour les séries univariées, fondée sur l'utilisation de processus *ARIMA*. Cette méthode guide l'identification, l'estimation et la validation des modèles *ARIMA*, garantissant ainsi que le modèle choisi capture au mieux les dynamiques de la série temporelle. Cette approche repose sur les étapes suivantes :

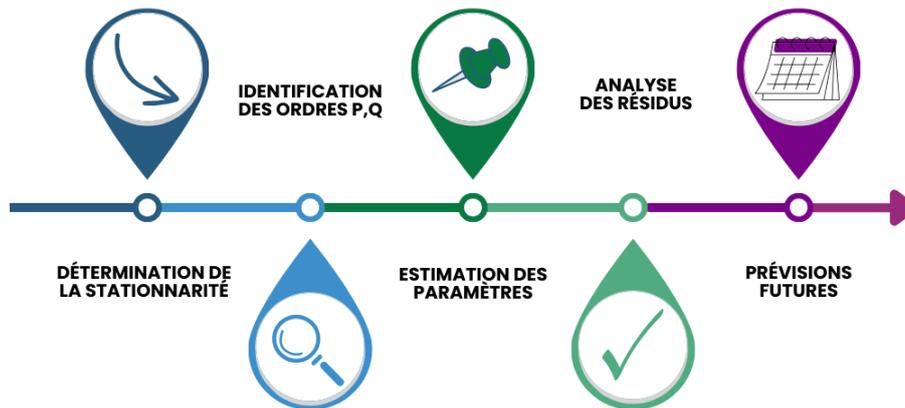


FIGURE 3.1 – La méthodologie BOX & JENKINS

1. La détermination de la stationnarité :

Pour rappel, une série temporelle X_t est dite stationnaire, si sa moyenne et ses covariances sont indépendantes du temps t . Pour rendre une série stationnaire, une technique appelée "différenciation" est utilisée. Cette technique permet de stabiliser la série en éliminant les tendances ou autres formes de non-stationnarité. Mathématiquement, cela consiste à calculer la différence entre les valeurs consécutives d'une série temporelle. La série différenciée de X_t , notée ΔX_t , est définie par :

$$\Delta X_t = X_t - X_{t-1}$$

Pour atteindre la stationnarité, des différenciations d'ordre supérieur peuvent être appliquées. Par exemple, une différenciation d'ordre deux s'écrit :

$$\begin{aligned}\Delta X_t^2 &= \Delta X_t - \Delta X_{t-1} \\ &= X_t - X_{t-1} - (X_{t-1} - X_{t-2}) \\ &= X_t - X_{t-2}\end{aligned}$$

L'analyse de la stationnarité peut être effectuée à l'aide d'une étude des autocorrélogrammes. Ces graphiques représentent les coefficients d'autocorrélation en fonction des décalages temporels (ou lags).

- L'autocorrélogramme d'une série non stationnaire montre souvent une décroissance lente et prolongée, indiquant que les valeurs passées continuent d'influencer fortement les valeurs futures.
- L'autocorrélogramme d'une série stationnaire montre une décroissance rapide vers zéro des coefficients d'autocorrélation, indiquant que les valeurs passées ont une influence limitée sur les valeurs futures.

2. L'identification des ordres p, q :

L'identification des ordres p et q dans un modèle *ARIMA* est importante. Ces ordres indiquent combien de lags influencent la valeur actuelle. Les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) sont utilisées à cette fin.

- **L'identification de l'ordre q** : L'ACF est utilisée pour déterminer l'ordre q d'un modèle *ARIMA*. Dans un processus $MA(q)$, l'ACF affiche une coupure nette après le décalage q , car la série temporelle est influencée par les erreurs aléatoires des q périodes antérieures. Par conséquent, l'ACF présente des valeurs significatives jusqu'au décalage q puis devient nulle ou presque nulle pour les décalages supérieurs, ce qui permet d'identifier directement l'ordre q .
- **L'identification de l'ordre p** : Le PACF est utilisé pour identifier l'ordre p d'un modèle *ARIMA*. Pour un processus $AR(p)$, le PACF montre une coupure nette après le lag p , car il mesure la corrélation directe entre une observation et un lag particulier, en éliminant l'effet des lags intermédiaires. Dans un modèle auto-régressif, la PACF montre des valeurs significatives jusqu'au lag p et devient nulle au-delà, ce qui permet d'identifier clairement l'ordre p .

3. L'estimation des paramètres :

Il est ensuite nécessaire d'estimer les coefficients $\theta_1, \dots, \theta_q$ et ϕ_1, \dots, ϕ_p .

En supposant que les résidus du modèle sont normalement distribués et indépendants, ces paramètres peuvent être ajustés par la méthode du maximum de vraisemblance. Cette méthode consiste à maximiser la vraisemblance suivante :

$$L(\kappa, \phi, \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \times \frac{1}{\sqrt{\det(\Gamma)}} \times \exp\left(-\frac{1}{2\sigma^2}\kappa'\Gamma^{-1}\kappa\right)$$

où

- κ : le vecteur des données différenciées.
 - ϕ et θ : les paramètres du modèle.
 - σ^2 : la variance des résidus.
 - Γ : la matrice de covariance de κ .
 - n : le nombre d'observations.
- **L'estimation de σ^2** : L'estimateur de la variance σ^2 est donné par :

$$\hat{\sigma}^2 = \frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} (\kappa_t - \hat{\kappa}_t)^2$$

- **L'estimation de ϕ et θ** : Les estimateurs de ϕ et θ sont obtenus en minimisant la fonction :

$$l(\phi, \theta) = \ln\left(\frac{1}{t_n - t_1} \sum_{t=t_1}^{t_n} (\kappa_t - \hat{\kappa}_t)^2\right) + \sum_{j=1}^{t_n - t_1} \ln(\epsilon_{j-1})$$

La convergence de ϕ et θ vers une distribution normale permet d'écrire l'équation suivante et d'en déduire les intervalles de confiance des paramètres.

$$\begin{pmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_q \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \\ \theta_1 \\ \vdots \\ \theta_q \end{pmatrix}, \frac{\sigma^2}{t_n - t_1} \Gamma_n \right)$$

4. L'analyse de résidus :

Pour valider l'adéquation du modèle *ARIMA* retenu, il est essentiel d'analyser les résidus, c'est-à-dire les écarts entre les valeurs observées et celles prédites par le modèle. Cette analyse permet de vérifier que les hypothèses sous-jacentes du modèle sont respectées : les résidus doivent être centrés, indépendants et normalement distribués.

— La centralité :

L'analyse de la centralité des résidus commence par l'examen graphique de leur répartition autour de la ligne de zéro. Idéalement, les résidus devraient se distribuer symétriquement autour de cette ligne.

Le calcul de la moyenne des résidus permet de quantifier leur centralité, qui doit idéalement être proche de zéro :

$$\bar{\epsilon} = \frac{1}{n} \sum_{t=1}^n \epsilon_t$$

où ϵ_t représente les résidus à chaque instant t , et n est le nombre total d'observations.

Pour évaluer si cette moyenne est statistiquement différente de zéro, un test de STUDENT (ou t-test) peut être réalisé. L'hypothèse nulle (\mathcal{H}_0) stipule que la moyenne des résidus est égale à zéro.

— L'indépendance :

Pour vérifier l'indépendance des résidus, une analyse graphique des autocorrélogrammes (ACF) peut être réalisée. Idéalement, pour que les résidus soient indépendants, les valeurs de l'ACF doivent se situer à l'intérieur des intervalles de confiance, indiquant une absence de corrélation significative.

Pour confirmer cette indépendance, le test de BOX-PIERCE peut être effectué. Ce test évalue l'hypothèse nulle (\mathcal{H}_0) selon laquelle les résidus sont indépendants.

— La distribution normale :

L'examen des histogrammes des résidus avec une courbe de distribution normale superposée permet d'avoir une première visualisation pour évaluer si les résidus suivent une distribution normale.

L'analyse QQ-plots permet de comparer les quantiles des résidus observés avec ceux d'une distribution normale théorique. Si les résidus sont normalement distribués, les points du QQ-plot doivent se situer proche de la ligne diagonale.

Pour confirmer la normalité des résidus, le test de SHAPIRO peut être réalisé. Ce test permet de vérifier l'hypothèse nulle, \mathcal{H}_0 , selon laquelle les résidus suivent une distribution normale.

5. Les prévisions futures

Les prévisions futures constituent la dernière étape de la méthodologie de BOX & JENKINS. En appliquant les modèles *ARIMA* précédemment identifiés, estimés, et validés, les valeurs futures de la série temporelle sont obtenues.

Application sur les données :

La récente perturbation des dynamiques de mortalité soulève plusieurs questions sur son impact sur les prévisions futures. Afin de quantifier cet impact et de tester la robustesse de ce modèle face aux changements récents, deux scénarios de projection sont envisagés. Le premier scénario inclut les années marquées par cette perturbation (étude de la série de 1991 à 2023), tandis que le second scénario les exclut (étude de la série de 1991 à 2019).

Seuls les résultats de modélisation pour le scénario incluant toutes les années jusqu'en 2023 sont présentés en détail. Le scénario excluant ces années suit le même processus. Il est présenté en annexe. Toutefois, l'étape des prévisions futures est présentée pour les deux scénarios, afin d'évaluer l'impact de l'intégration des années marquées par une perturbation sur les projections futures.

1. La détermination de la stationnarité :

Les autocorrélogrammes de la série κ_t montrent des signes de non-stationnarité, tant chez les hommes que chez les femmes, caractérisés par une décroissance lente des coefficients d'autocorrélation.

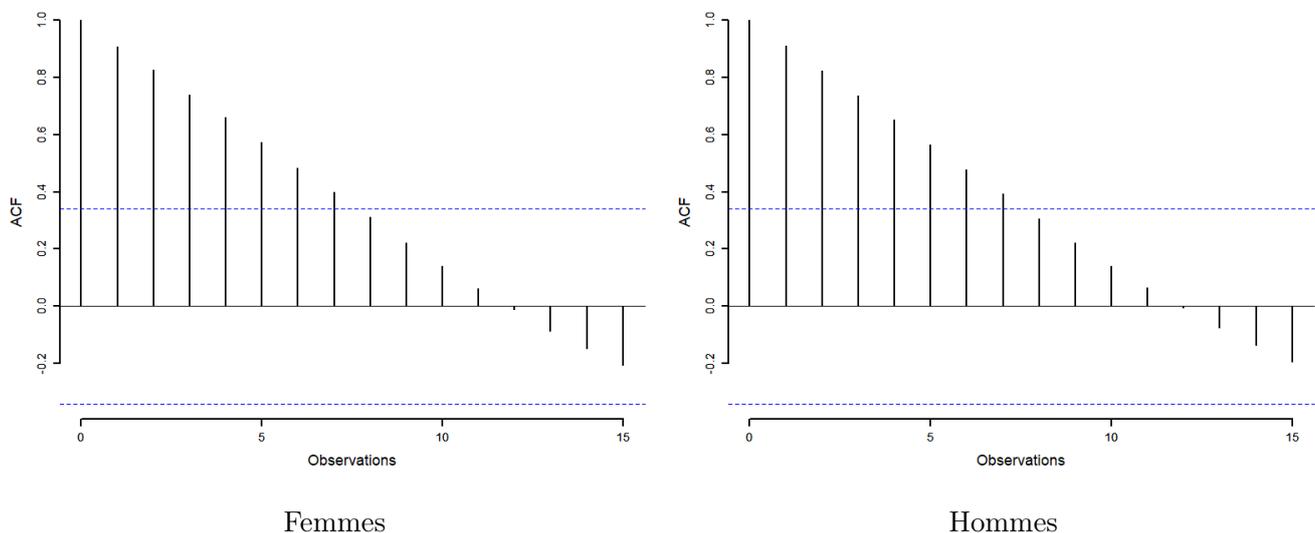


FIGURE 3.2 – Les autocorrélogrammes (ACF) des séries κ_t de 1991 à 2023

Après avoir différencié la série κ_t une fois, une décroissance rapide des coefficients d'autocorrélation est observée, indiquant que la série est devenue stationnaire.

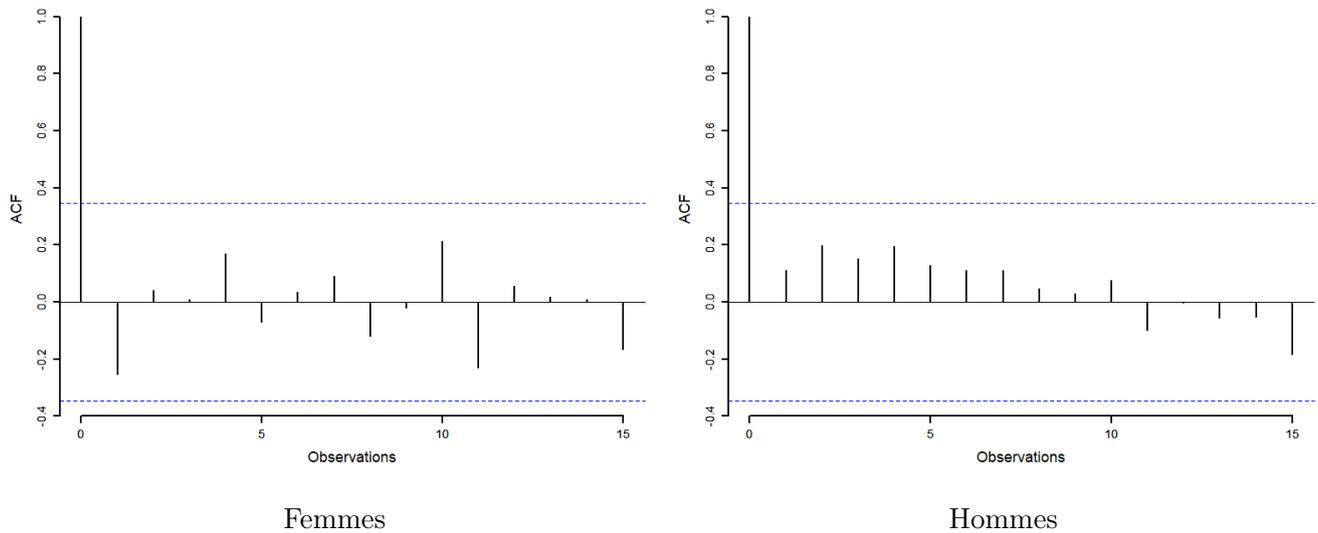


FIGURE 3.3 – Les autocorrélogrammes (ACF) des séries κ_t différenciées une fois de 1991 à 2023

→ Pour les hommes et pour les femmes, le nombre de différenciations retenu est $d = 1$.

2. L'identification des ordres p, q :

Les ACF des séries κ_t différenciées, présentés ci-dessus, pour les hommes et pour les femmes, montrent une chute notable après le premier lag, suggérant un ordre q probablement faible pour le modèle MA .

Les PACF des séries κ_t différenciées montrent des valeurs proches de zéro, suggérant un ordre p très faible ou nul pour les hommes et les femmes.

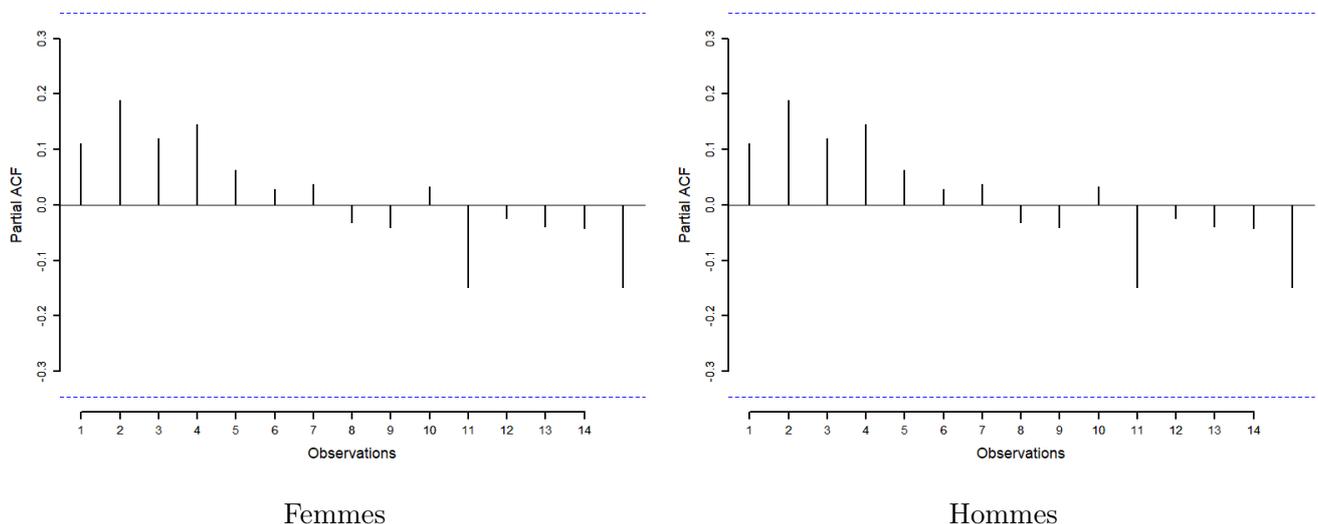


FIGURE 3.4 – Les autocorrélogrammes partiels (PACF) des séries κ_t différenciées une fois de 1991 à 2023

Pour affiner la sélection du modèle le plus approprié, la fonction `auto.arima()` de R est utilisée. Cette fonction automatise le processus de sélection en identifiant le modèle $ARIMA$ qui présente la meilleure adéquation aux données selon le Critère d'Information Bayésien (BIC). Le BIC permet de comparer différents modèles en pénalisant la complexité excessive tout en cherchant à maximiser l'ajustement aux données, ce qui en fait un critère particulièrement pertinent pour la modélisation des séries temporelles.

Pour les femmes, le modèle $ARIMA(0, 1, 0)$ avec une dérive égale à $-3,7728$ a été identifié comme étant le modèle optimal selon le critère BIC. Le BIC pour ce modèle est de 164,34. Mathématiquement, ce modèle peut être formulé comme suit :

$$X_t = -3,7728 + X_{t-1} + \epsilon_t$$

Avec X_{t-1} la valeur de la série à l'instant précédent et ϵ_t , le terme d'erreur.

Pour les hommes, le modèle $ARIMA(0, 1, 0)$ avec une dérive égale à $-6,2472$ a été identifié comme étant le modèle optimal selon le critère BIC. Le BIC pour ce modèle est de 152,05. Mathématiquement, ce modèle peut être formulé comme suit :

$$X_t = -6,2472 + X_{t-1} + \epsilon_t$$

Avec X_{t-1} la valeur de la série à l'instant précédent et ϵ_t , le terme d'erreur.

→ Pour les hommes et pour les femmes, les ordres p et q sont fixés à $p = 0, q = 0$.

3. L'estimation des paramètres :

Le modèle retenu ne comportant pas de composante autorégressive ni de moyenne mobile ($p = 0, q = 0$), aucune estimation de paramètre n'est nécessaire.

4. L'analyse de résidus :

— La centralité :

Les résidus du modèle ARIMA apparaissent centrés autour de zéro sur les graphiques ci-dessous.

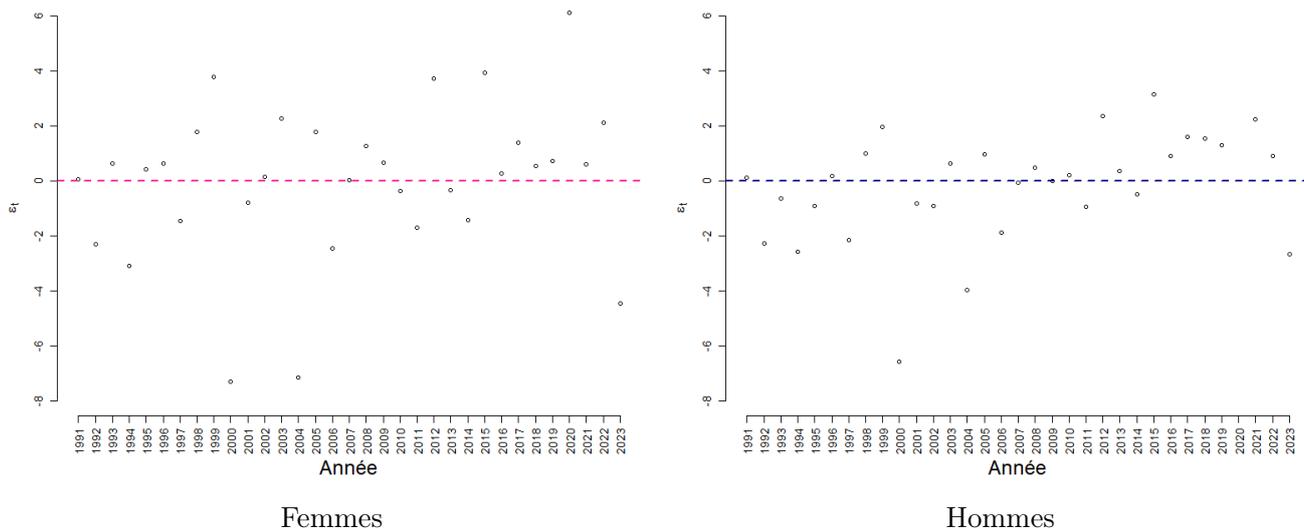


FIGURE 3.5 – Les nuages des points des résidus de la série 1991-2023

Pour les femmes, la moyenne des résidus est très proche de zéro, 0,002, ce qui suggère une bonne centralité. Pour les hommes, la moyenne des résidus est égale à 0,0035, indiquant également une centralité proche de zéro.

Les valeurs élevées des p-values, du test de STUDENT, indiquent que l'hypothèse nulle, selon laquelle la moyenne des résidus est égale à zéro, ne peut être rejetée. Ainsi, les résidus sont bien centrés autour de zéro pour les hommes et pour les femmes.

	Femmes	Hommes
p-value	0,9966	0,9931

TABLE 3.1 – Les p-values du test de STUDENT pour la série de 1991 à 2023

— **L'indépendance :**

Les autocorrélogrammes ci-dessous montrent que la majorité des coefficients d'autocorrélation sont proches de zéro et se trouvent à l'intérieur de l'intervalle de confiance, tant pour les femmes que pour les hommes. Cette distribution suggère qu'il n'y a pas d'autocorrélation résiduelle significative parmi les résidus.

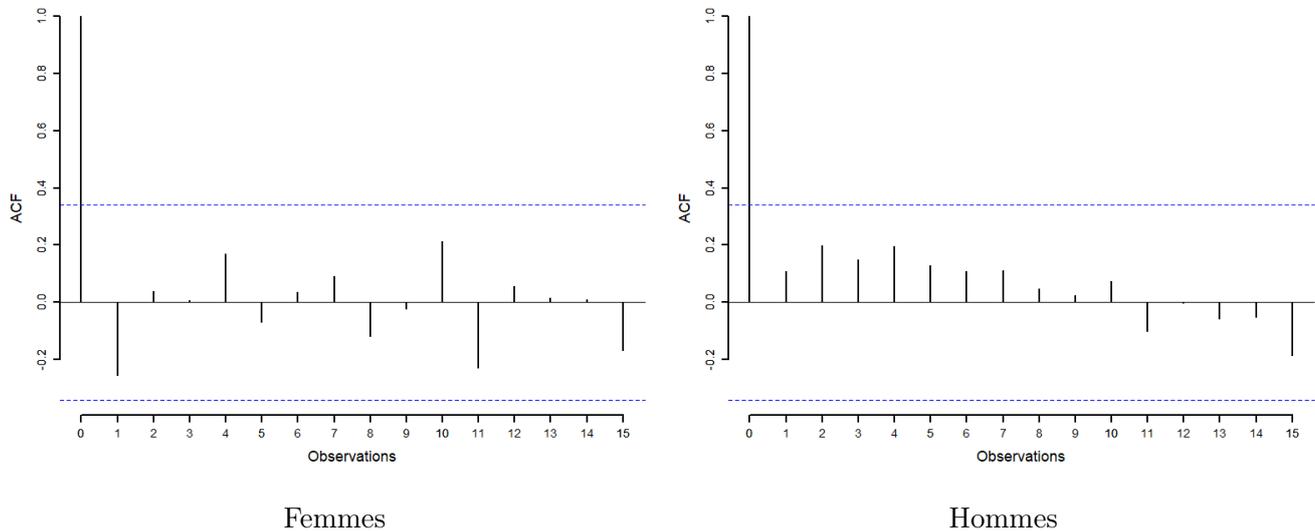


FIGURE 3.6 – Les autocorrélogrammes des résidus de la série 1991-2023

Pour les femmes, le test de BOX-PIERCE donne une valeur de p-value supérieure au seuil de 5%, indiquant qu'il n'y a pas d'autocorrélation significative dans les résidus. Pour les hommes, le test confirme également l'absence d'autocorrélation significative. Cela suggère que les résidus, du modèle *ARIMA* retenu, sont bien indépendants.

	Femmes	Hommes
p-value	0,1452	0,5296

TABLE 3.2 – Les p-values du test de BOX PIERCE pour la série de 1991 à 2023

— **La distribution normale :**

Les histogrammes ci-dessous montrent que les distributions semblent bien ajustées à la courbe normale, suggérant que les résidus sont proches d'une distribution normale.

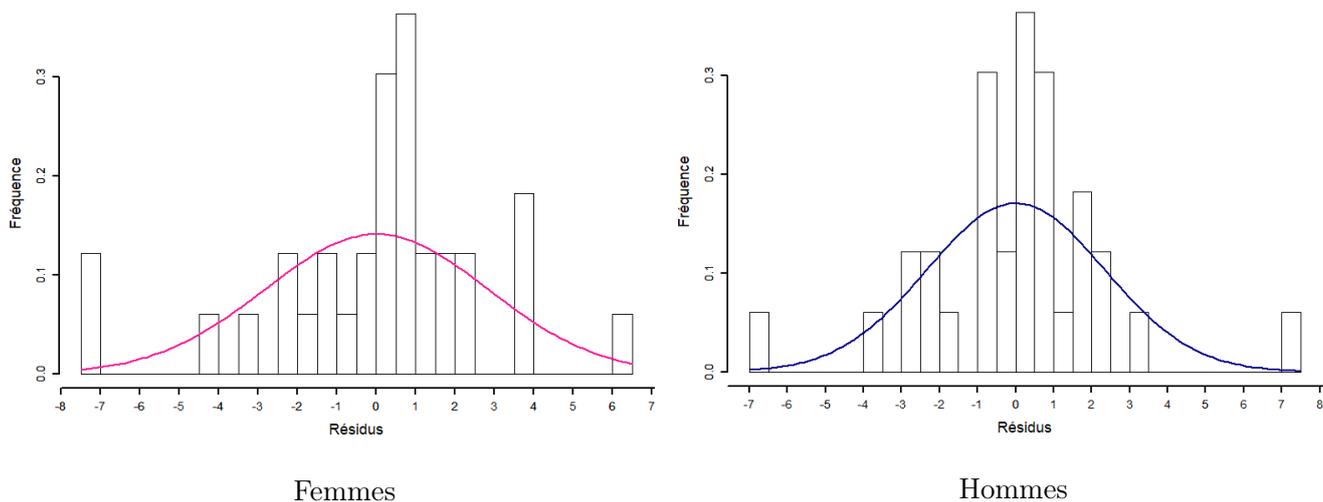


FIGURE 3.7 – Les histogrammes des résidus de la série 1991-2023

Les points des QQ-plot suivent globalement la diagonale, ce qui indique que les résidus sont proches d'une distribution normale, bien que quelques valeurs s'écartent légèrement de la droite théorique.

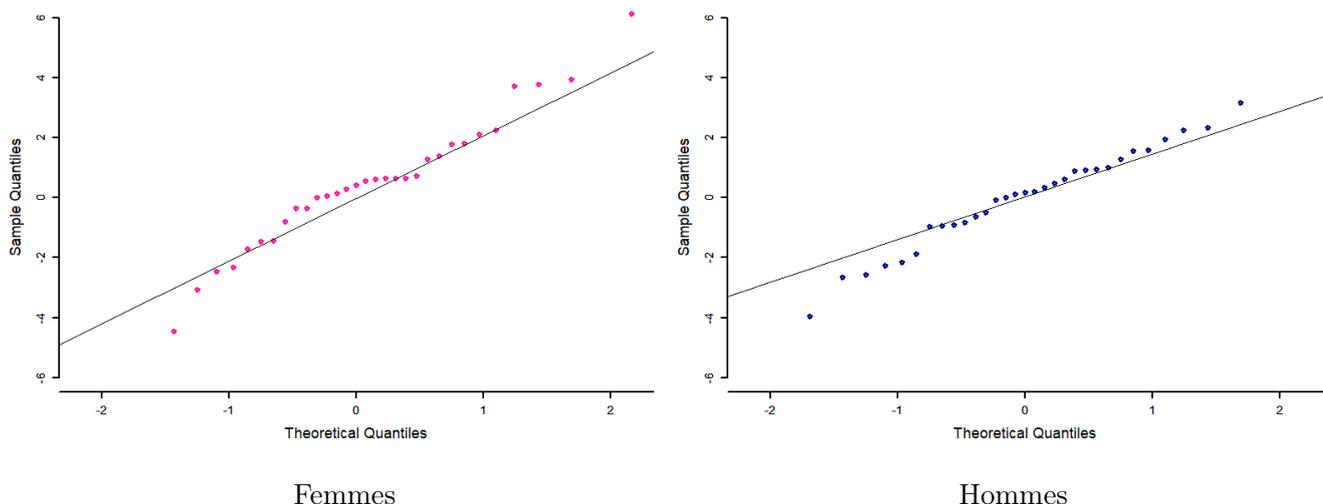


FIGURE 3.8 – Les QQ-plots des résidus de la série 1991-2023

Tant pour les femmes que pour les hommes, la p-value du test de SHAPIRO est supérieure au seuil de 5%. L'hypothèse nulle, selon laquelle les résidus suivent une distribution normale, ne peut pas être rejetée.

	Femmes	Hommes
p-value	0,06597	0,09097

TABLE 3.3 – Les p-values du test de SHAPIRO pour la série de 1991 à 2023

5. Les prévisions futures

En appliquant les modèles *ARIMA* précédemment identifiés, estimés et validés, pour les deux scénarios, les projections suivantes sont obtenues.

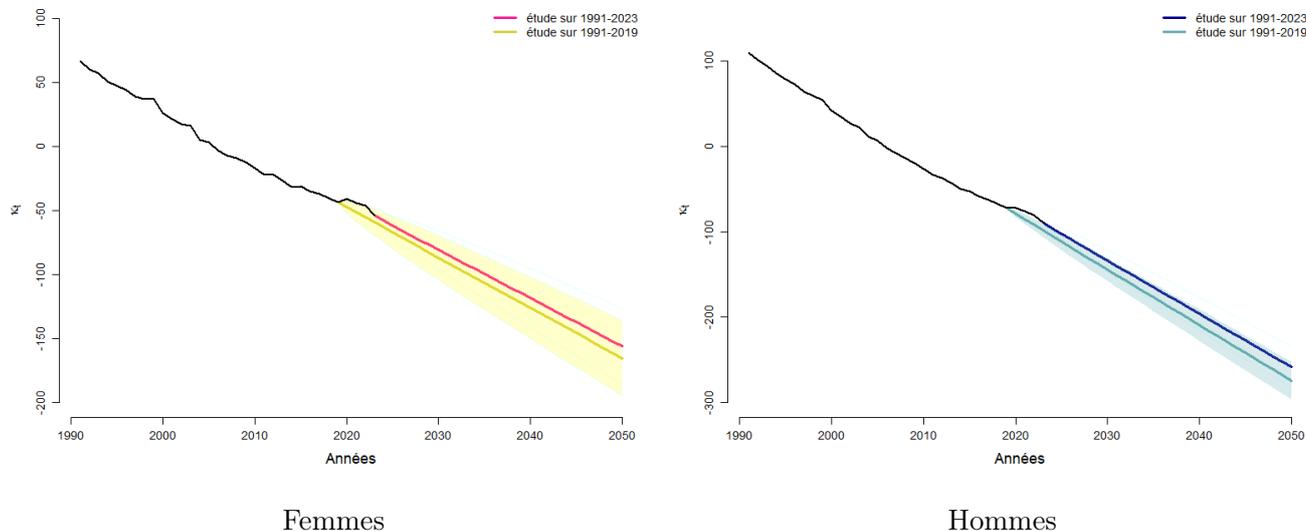


FIGURE 3.9 – Les projections de la série κ_t avec le modèle *ARIMA*

Les projections futures de la composante temporelle κ_t , obtenues à l'aide des modèles *ARIMA*, continuent de suivre la tendance historique observée, avec une diminution progressive de κ_t . Cette décroissance reflète une amélioration continue des taux de mortalité, en prolongeant les dynamiques de baisse observées au cours des dernières décennies.

Toutefois, l'intégration des données les plus récentes, marquées par des perturbations, modifie cette trajectoire. Lorsque ces dernières années sont incluses dans les projections, la pente de la décroissance de κ_t devient moins marquée, indiquant un ralentissement de l'amélioration de la mortalité. Cela souligne l'importance de prendre en compte ces nouvelles données pour ajuster les modèles et refléter au mieux les dynamiques actuelles.

Les projections basées sur les données de 1991 à 2023 se situent toujours dans l'intervalle de confiance à 95% des prévisions fondées sur les données de 1991 à 2019. Cela indique que, même si l'intégration des données récentes entraîne une légère déviation, ces nouvelles projections restent compatibles avec les tendances historiques.

Une analyse quantitative montre que l'écart absolu entre les deux scénarios de projection (1991-2019 vs. 1991-2023) est de 5,17 pour les femmes et de 8,9 pour les hommes en 2024. Ces écarts se creusent progressivement à mesure que l'on s'éloigne dans le temps, pour atteindre 9,25 d'écarts en 2050 pour les femmes et 15,9 pour les hommes.

Ces observations mettent en évidence l'importance d'intégrer les données les plus récentes dans les projections de mortalité, car elles peuvent influencer de manière significative sur les trajectoires futures. Si ces changements se révèlent être de nature structurelle plutôt que conjoncturelle, ils pourraient transformer durablement les dynamiques de mortalité. À la lumière de ces résultats, il serait pertinent d'examiner des modèles de projection plus flexibles ou alternatifs pour mieux s'adapter à ces évolutions. Cette approche sera explorée dans la troisième partie de ce mémoire.

3.2 Les taux de mortalité projetés

Dans le cadre de ce mémoire, les taux de mortalité sont établis à l'aide du modèle de RENSCHAW & HABERMAN, une généralisation du modèle de LEE & CARTER intégrant un terme de cohorte. Cette approche permet de mieux capturer les dynamiques complexes de la mortalité liées non seulement à l'âge et au temps, mais aussi aux effets générationnels.

Pour réaliser les projections des taux de mortalité futurs, les paramètres α_x et β_x estimés et lissés précédemment sont donc utilisés.

Pour obtenir des projections précises des taux de mortalité, il est nécessaire de disposer d'estimations fiables des composantes temporelles. Ainsi, tout comme le paramètre κ_t , la composante de cohorte γ_{t-x} , qui dépend également du temps t , doit être projetée.

Les résultats ci-dessous sont présentés pour le scénario incluant les années marquées par une perturbation, les conclusions restant similaires pour le scénario qui les exclut.

Pour analyser les tendances de mortalité, la composante de cohorte γ_{t-x} a été projetée en utilisant des modèles *ARIMA*, ajustés sur les données de 1991 à 2023. Les modèles retenus pour les femmes et les hommes sont ceux présentant le meilleur BIC :

- Femmes : Un modèle *ARIMA*(0, 1, 1) avec dérive a été sélectionné :

$$X_t = 0,0253 + X_{t-1} + \epsilon_t - 0,3679 \times \epsilon_{t-1}$$

- Hommes : Un modèle *ARIMA*(2, 1, 2) avec dérive a été retenu :

$$X_t = 0,061 + X_{t-1} + 1,5147 \times (X_{t-1} - X_{t-2}) - 0,5403 \times (X_{t-2} - X_{t-3}) + \epsilon_t - 1,8265 \times \epsilon_{t-1} + 0,9 \times \epsilon_{t-2}$$

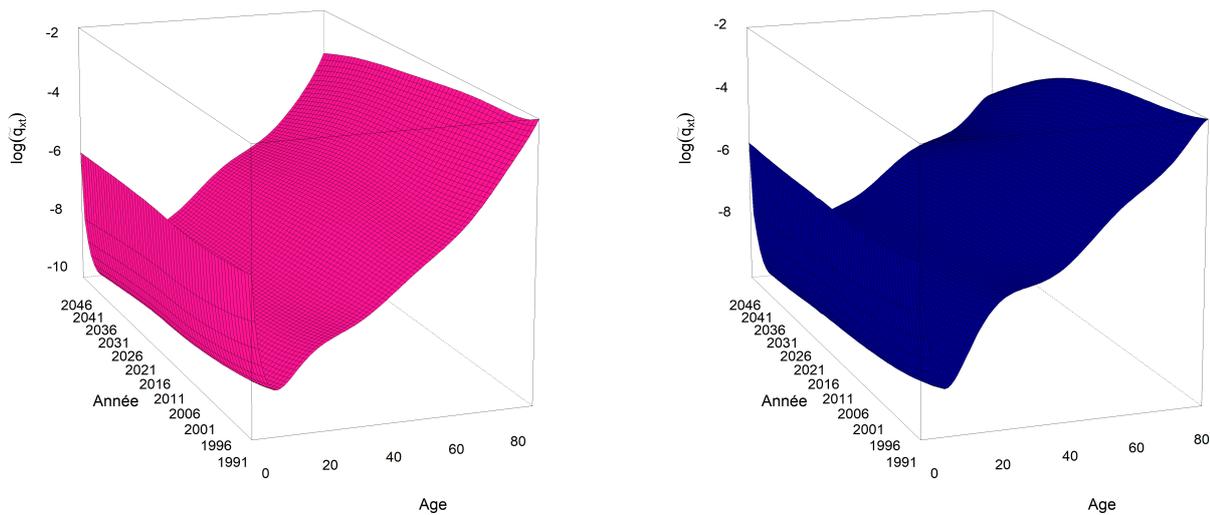
Les taux de mortalité peuvent ainsi se déduire. Pour rappel, la relation entre les taux de mortalité instantanés $\mu_{x,t}$ et les probabilités de décès $q_{x,t}$ peut être approximée par la formule suivante :

$$q_{x,t} = 1 - e^{-\mu_{x,t}}$$

où $\mu_{x,t} = \exp(\alpha_x + \beta_x \kappa_t + \gamma_{t-x})$

Les taux de mortalité projetés pour la période 2024-2050, ainsi calculés, montrent une évolution cohérente avec les dynamiques de mortalité observées de 1991 à 2023.

- Les taux projetés reflètent les schémas habituels de la mortalité par âge. La mortalité infantile élevée est suivie d'une baisse rapide des taux de mortalité, puis d'une "bosse des accidents". Les taux augmentent ensuite exponentiellement avec l'âge.
- En termes d'évolution temporelle, les projections des taux de mortalité de 2024 à 2050 montrent une amélioration au fil des années, avec une diminution progressive des taux. Cette tendance est en accord avec les dynamiques de mortalité observées de 1991 à 2023.



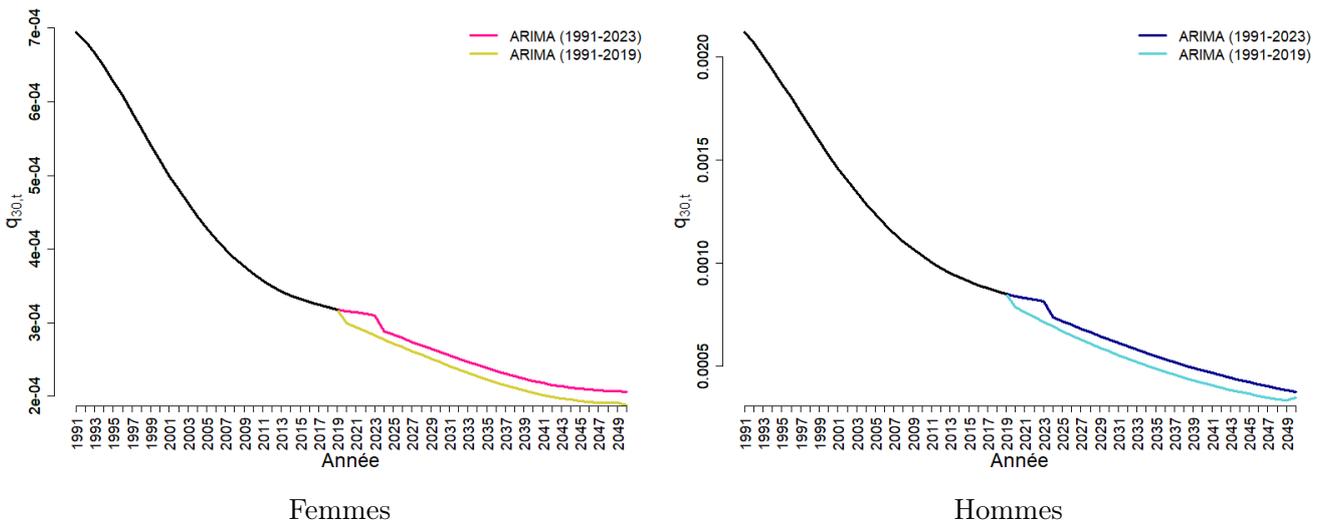
Femmes

Hommes

FIGURE 3.10 – Le logarithme des taux de mortalité projetés avec le modèle ARIMA (1991-2023)

Comparaison des taux de mortalité projetés selon les deux scénarios :

Les tendances observées précédemment sur la composante temporelle κ_t se retrouvent également sur les taux de mortalité : tout comme κ_t , les taux de mortalité continuent de diminuer, prolongeant ainsi les dynamiques historiques d'amélioration de la mortalité. Cette diminution, observée dans les projections, confirme que malgré les perturbations récentes, la tendance globale demeure orientée vers une diminution des taux de mortalité. Par exemple, les figures ci-dessous montrent clairement cette diminution continue pour les femmes et pour les hommes de 30 ans, même après l'intégration des années marquées par des perturbations.



Femmes

Hommes

FIGURE 3.11 – Les taux de mortalité projetés à 30 ans

Cependant, tout comme pour κ_t , l'inclusion des données des dernières années perturbe légèrement cette trajectoire. Lorsque les années récentes, marquées par une "bosse" de mortalité, sont intégrées dans les projections, la pente de la décroissance devient moins prononcée. Cela signifie que les taux de mortalité projetés sont plus élevés que ceux calculés en excluant ces années perturbées.

Globalement, cette tendance est confirmée par l'analyse quantitative : en moyenne, l'intégration des années de perturbation entraîne une augmentation des taux de mortalité projetés de 4,85% pour les femmes et de 7,8% pour les hommes en 2024. Ces écarts s'amplifient avec le temps, atteignant 11,38% pour les femmes et 17,93% pour les hommes en 2050, ce qui est attendu, car l'incertitude des projections augmente avec l'éloignement temporel.

L'intégration des données récentes affecte de manière significative les projections futures. Si ces changements s'avèrent structurels plutôt que temporaires, ils pourraient entraîner des conséquences majeures sur l'évolution des taux de mortalité à long terme. Il est donc crucial de prendre en compte les années les plus récentes dans les modèles de projection, afin de capter ces nouvelles dynamiques et d'assurer des prévisions fiables, reflétant avec précision les réalités démographiques actuelles et futures.

À la suite de cette analyse, deux tables de mortalité distinctes ont été élaborées pour chaque sexe :

- Pour les femmes, les taux de mortalité sont disponibles de 0 à 90 ans sur la période de 1991 à 2050. La première table est construite en se basant sur les données disponibles de 1991 à 2019, tandis que la seconde inclut les données jusqu'en 2023.
- Pour les hommes, les taux de mortalité sont disponibles de 0 à 85 ans sur la même période. De même, une table est établie à partir des données de 1991 à 2019, et une autre en intégrant les données jusqu'en 2023.



Ce qu'il faut retenir du chapitre :

- * Dans les modèles de mortalité prospectifs, comme le modèle de RENSHAW & HABERMAN, la composante temporelle est classiquement assimilée à un processus stochastique, généralement modélisé sous forme de marche aléatoire avec dérive ($ARIMA(0, 1, 0)$). Ce type de modèle combine des éléments d'autorégression (AR), de différenciation (I), et de moyenne mobile (MA). La méthodologie de BOX & JENKINS permet de déterminer la structure optimale des modèles, d'estimer leurs paramètres et de valider leur précision à travers l'analyse des résidus. Dans les projections, la composante temporelle κ_t tend à diminuer avec le temps, mais cette décroissance est moins prononcée lorsque les années récentes, marquées par une "bosse" de mortalité, sont incluses.
- * Les projections des taux de mortalité, basées sur l'évolution de la composante temporelle κ_t , confirment une tendance globale à la diminution des taux de mortalité avec le temps. Cependant, cette diminution est plus modérée lorsque les années récentes, marquées par des perturbations, sont incluses. En conséquence, les taux de mortalité projetés s'avèrent plus élevés que ceux calculés en excluant ces années perturbées. Cela illustre l'impact significatif des changements récents sur les prévisions futures de mortalité et souligne l'importance d'intégrer ces données dans les modèles, en particulier si ces perturbations reflètent des changements structurels durables.

Le chapitre suivant abordera la fermeture de la table de mortalité, en explorant les méthodes d'extrapolation des taux de mortalité aux âges les plus élevés pour assurer une estimation complète et cohérente des trajectoires de mortalité.

Chapitre 4

La fermeture de la table

Aux âges très avancés, le manque de données disponibles entraîne une grande volatilité des taux de mortalité. Pour éviter ces fluctuations et garantir des projections continues et cohérentes, la fermeture des tables de mortalité s'impose. Ainsi, les actuaires et démographes ont développé des modèles spécifiques afin de prolonger les taux de mortalité aux âges avancés.

Ce chapitre vise à élaborer avec précision des taux de mortalité pour les âges les plus élevés. Deux modèles sont principalement considérés pour cette extrapolation : la méthode de COALE & KISKER, développée en 1990, et celle de DENUIT & GODERNIAUX, introduite en 2005.

4.1 La méthode de COALE & KISKER

La méthode de COALE & KISKER, introduite en 1990, est une approche utilisée pour extrapoler les taux de mortalité aux âges avancés lorsque les données empiriques ne sont plus disponibles ou fiables. Cette méthode se base sur le constat que la progression des taux de mortalité tend à ralentir aux âges les plus avancés.

Selon la méthode de COALE & KISKER, le taux de mortalité instantanés, à l'âge x pour l'année t est défini par¹ :

$$\hat{\mu}_x = \hat{\mu}_{65} e^{k_x(x-65)}$$

où :

- $\hat{\mu}_x$ représente le taux de mortalité à l'âge x .
- k_x est le taux de croissance de la mortalité à l'âge x par rapport à 65 ans.
- $x \geq 65$.

Les auteurs observent un pic du taux de croissance de la mortalité, k_x , aux alentours de 80 ans, suivi d'une décroissance linéaire. Cette observation les conduit à modéliser l'hypothèse suivante :

$$k_x = k_{80} + s(x - 80)$$

où :

- k_{80} est le taux de croissance de la mortalité à l'âge de 80 ans par rapport à 65 ans, $k_{80} = \frac{\hat{\mu}_{80} - \hat{\mu}_{65}}{(80 - 65)}$.
- s est le coefficient de pente représentant le changement linéaire de k_{80} au-delà de 80 ans.

1. L'indice t est ici omis pour des questions de lisibilité.

Dans cette méthode, l'âge ultime est fixé à 110 ans, avec des taux de mortalité limites définis comme suit :

$$\hat{\mu}_{110} = \begin{cases} 0,8 & \text{pour les femmes} \\ 1 & \text{pour les hommes} \end{cases}$$

Pour extrapoler les taux de mortalité au-delà de 80 ans, la méthode utilise l'expression suivante :

$$\begin{aligned} \hat{\mu}_{110} &= \hat{\mu}_{79} \exp \left(\sum_{x=80}^{110} k_x \right) \\ &= \hat{\mu}_{79} \exp \left(\sum_{x=80}^{110} k_{80} + s(x - 80) \right) \\ &= \hat{\mu}_{79} \exp \left(31 k_{80} + s \sum_{x=80}^{110} (x - 80) \right) \\ &= \hat{\mu}_{79} \exp \left(31 k_{80} + s \frac{30 \times 31}{2} \right) \end{aligned}$$

Le coefficient de pente, s , est ainsi déterminé à partir des valeurs de mortalité à 79 ans et à 110 ans, en utilisant la formule suivante :

$$s = - \frac{\ln \left(\frac{\hat{\mu}_{79}}{\hat{\mu}_{110}} \right) + 31 k_{80}}{465}$$

Ainsi, les taux de mortalité pour les âges supérieurs à 80 ans se calculent comme suit :

$$\hat{\mu}_x = \hat{\mu}_{x-1} \exp (k_{80} + s(x - 80)), \quad x \in \{80, 81, \dots, 108, 109\}$$

Dans le cadre de ce mémoire, la méthode de COALE & KISKER a été appliquée aux différentes tables de mortalité construites selon deux scénarios : l'un incluant les années récentes marquées par une "bosse" de mortalité et l'autre les excluant. Dans les deux cas, les surfaces de mortalité ne présentent pas de cassure que ce soit pour les femmes ou pour les hommes. Cette notion est illustrée ci-dessous avec le scénario prenant en compte la "bosse" de mortalité observée ces dernières années.

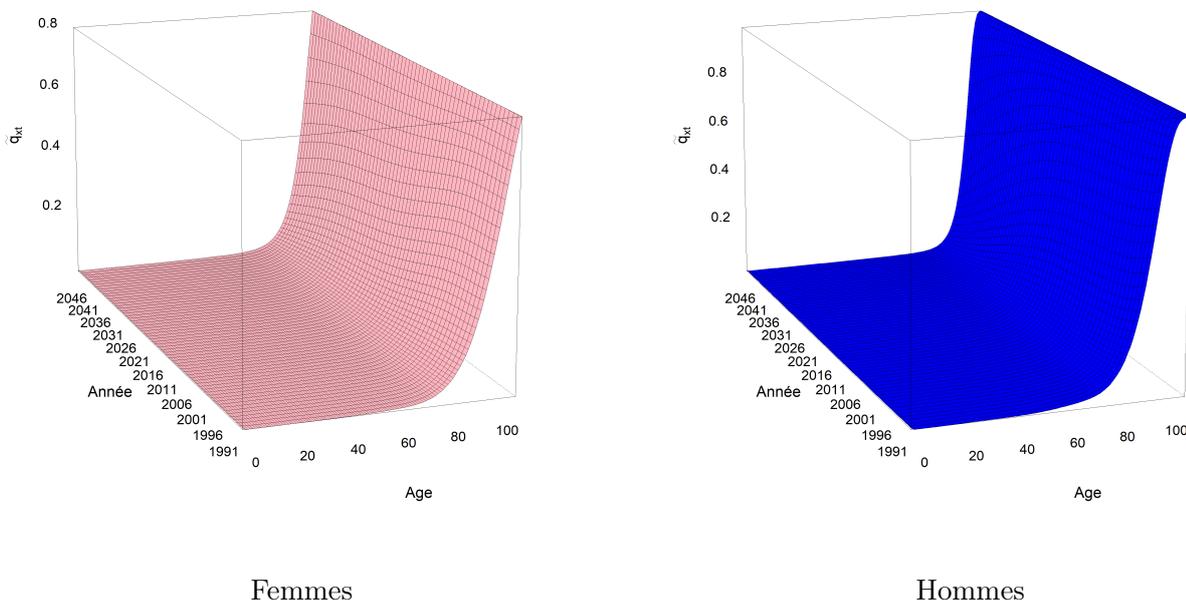


FIGURE 4.1 – Les taux de mortalité fermés avec COALE & KISKER (scénario 1991-2023)

4.2 La méthode de DENUIT & GODERNIAUX

La méthode de DENUIT & GODERNIAUX (2005) est une approche couramment utilisée pour la fermeture des tables de mortalité aux âges très élevés. Cette méthode repose sur l'ajustement par moindres carrés du modèle polynomial suivant :

$$\ln(\hat{q}_{x,t}) = a_t + b_t x + c_t x^2 + \epsilon_{x,t}$$

où :

- $\hat{q}_{x,t}$ est la probabilité de décès à l'âge x pour l'année t .
- a_t , b_t et c_t sont les coefficients à estimer pour chaque année t .
- $\epsilon_{x,t}$ sont des résidus indépendants et identiquement distribués selon une loi normale de moyenne 0 et de variance σ^2 : $\epsilon_{x,t} \sim \mathcal{N}(0, \sigma^2)$.

Le modèle pose que l'âge maximal est fixé à 130 ans, au-delà duquel la probabilité de décès est supposée égale à 1, $q_{130,t} = 1$.

L'hypothèse suivante, signifiant que la pente de la courbe des taux de mortalité est nulle à l'âge de 130 ans pour toute année t , est introduite :

$$\left. \frac{\partial q_{x,t}}{\partial x} \right|_{x=130} = 0$$

Ces hypothèses permettent de déduire la relation suivante pour chaque année t :

$$a_t + b_t x + c_t x^2 = c_t(130^2 - 260x + x^2)$$

Dans l'application de cette méthode, l'âge de raccord, à partir duquel l'extrapolation est effectuée, est déterminé de manière à maximiser le coefficient de détermination R^2 pour chaque année t . Cependant, un âge de raccord variant d'une année à l'autre peut entraîner des irrégularités dans la surface de mortalité aux âges élevés. Pour éviter cela, un âge de raccord unique est choisi sur une plage d'âge raisonnable.

Dans le cadre de ce mémoire, les âges raccords suivants ont été retenus pour les deux scénarios de mortalité :

- Pour les femmes : âge de raccord de 86 ans, sélectionné sur la plage 80-90 ans.
- Pour les hommes : âge de raccord de 70 ans, sélectionné sur la plage 70-85 ans.

Pour garantir une extrapolation plus lisse, une moyenne géométrique est ensuite appliquée sur les cinq années autour de cet âge de raccord.

Pour les hommes comme pour les femmes, des surfaces de mortalité sans cassures sont observées pour les deux scénarios de mortalité. Les graphiques ci-dessous illustrent ce résultat pour le scénario incluant les dernières années de l'étude.

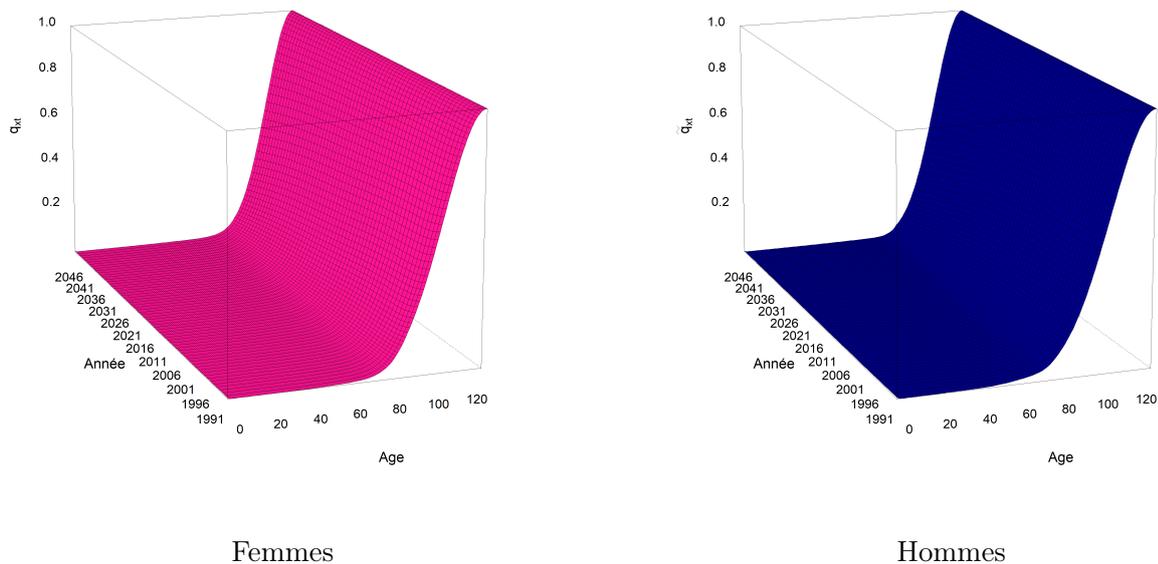


FIGURE 4.2 – Les taux de mortalité fermés avec DENUIT & GODERNIAUX (scénario 1991-2023)

4.3 La comparaison des méthodes de fermeture

Afin de déterminer la méthode d'extrapolation des taux de mortalité la plus appropriée, les deux approches sont comparées. Cette comparaison vise à évaluer leur capacité à reproduire fidèlement les tendances de mortalité aux âges avancés et à identifier celle qui offre la meilleure continuité des taux projetés.

Les deux méthodes d'extrapolation des taux de mortalité montrent une continuité aux âges avancés, sans rupture ou cassure évidente. Cela indique que les deux approches sont bien appropriées, comme illustré dans les graphiques ci-dessous pour les taux de mortalité en 2023, dans le scénario incluant les dernières années de l'étude.

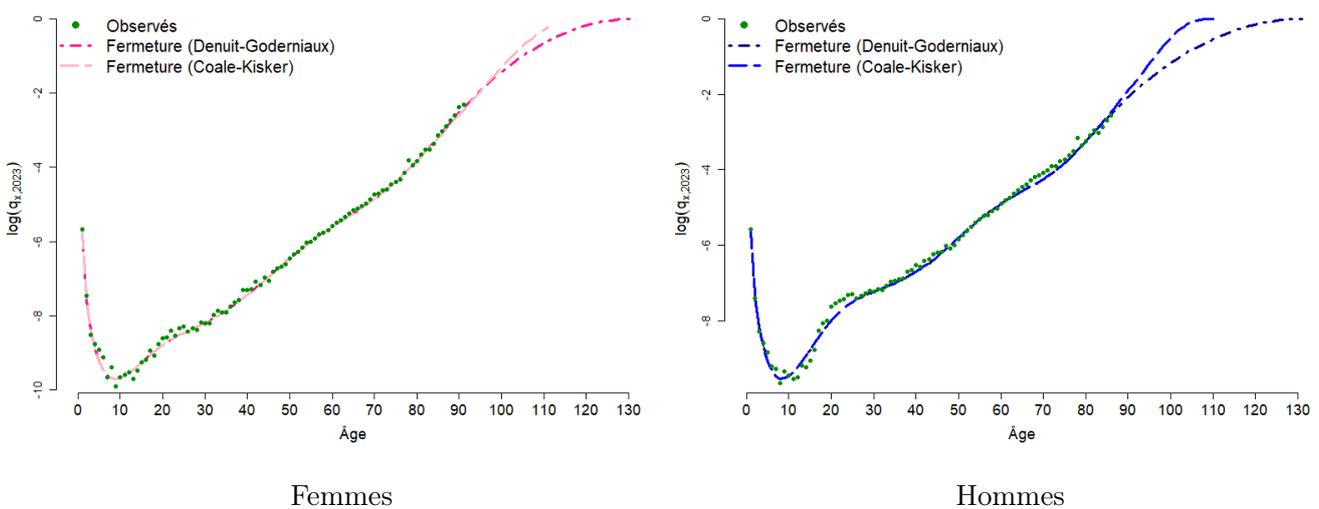


FIGURE 4.3 – Les logarithmes des taux de mortalité fermés pour l'année 2023

La méthode de DENUIT & GODERNIAUX a été retenue pour le prolongement des taux de mortalité aux âges élevés.

Les tables, désormais complètes, couvrent pour chaque sexe les âges de 0 à 130 ans sur la période de 1991 à 2050. Deux versions ont été élaborées : l'une est basée sur les données allant de 1991 à 2019, tandis que l'autre inclut les données jusqu'en 2023. Ces tables seront utilisées dans la dernière partie de ce mémoire pour calculer les provisions d'un régime de retraite supplémentaire, illustrant ainsi les impacts potentiels d'un changement dans les dynamiques de mortalité.



Ce qu'il faut retenir du chapitre :

Les projections précises aux âges les plus avancés sont essentielles pour garantir la stabilité financière des organismes d'assurance proposant des rentes viagères. Les méthodes de COALE & KISKER et de DENUIT & GODERNIAUX ont été utilisées pour extrapoler les taux de mortalité précédemment estimés et projetés. Ces deux méthodes ont montré une continuité des taux, sans cassure. La méthode de DENUIT & GODERNIAUX, qui extrapole les taux jusqu'à 130 ans, a été retenue.

Dans cette partie, l'effet de l'inclusion des dernières années, marquées par une "bosse" de mortalité, dans la construction des tables de mortalité a été étudié. Maintenant, il est pertinent d'explorer d'autres méthodes de projection pour la composante temporelle, κ_t , afin de mieux tenir compte des dynamiques récentes si elles s'avèrent être structurelles. La partie suivante se concentrera sur l'examen de ces méthodes alternatives.

Schématisation synthétique de la partie :



1. ESTIMATIONS

Trois méthodes de construction de tables générationnelles ont été utilisées : LEE & CARTER, LOG POISSON, RENSHAW & HABERMAN.



2. VALIDATION

La proximité des taux de mortalité prédits et observés a été examinée.



Le modèle de RENSHAW & HABERMAN est retenu.



3. PROJECTIONS

Les taux de mortalité ont été projetés avec une méthode classiquement utilisée, une marche aléatoire avec dérive. Deux scénarios ont été comparés : une table inclut les années affectées par un changement de dynamique et l'autre les exclut.



4. FERMETURE

Deux méthodes ont été comparées pour extrapoler les taux de mortalité aux très grands âges : COALE & KISKER et DENUIT & GODERNIAUX.



La méthode de DENUIT & GODERNIAUX a été retenue.



Des tables de mortalité par sexe, couvrant la période de 1991 à 2050 et les âges de 0 à 130 ans, ont été construites selon deux scénarios :

- **Les données de 1991 à 2019 ont été utilisées pour projeter les taux de mortalité de 2020 à 2050.**
- **Les données de 1991 à 2023 ont été utilisées pour projeter les taux de mortalité de 2024 à 2050.**

FIGURE 4.4 – Le schéma synthétique de la partie

Troisième partie

Modélisation alternative de la composante temporelle d'un modèle prospectif

Chapitre 1

Les modèles de projection alternatifs

Dans le contexte actuel, marqué par des perturbations récentes dans les dynamiques de mortalité, il devient pertinent de réévaluer l'efficacité de la méthode de projection classiquement utilisée dans les modèles de mortalité prospectifs, à savoir une marche aléatoire avec dérive. Si les changements observés dans les tendances de mortalité sont de nature structurelle plutôt que conjoncturelle, ce modèle pourrait ne plus être en mesure de capturer correctement ces nouvelles dynamiques.

Dans cette optique, il est pertinent d'envisager des modèles de projection plus sophistiqués, tels que les modèles ETS et PROPHET, qui offrent une meilleure flexibilité pour capturer les fluctuations récentes dans les dynamiques de mortalité.

1.1 Le modèle ETS

1.1.1 La présentation du modèle

Le modèle ETS (Error, Trend, Seasonal) est un modèle de séries temporelles basé sur le lissage exponentiel, qui décompose une série en trois composantes : l'erreur (Error), la tendance (Trend), et la saisonnalité (Seasonal). Chaque composante peut être modélisée de différentes manières, offrant une grande flexibilité pour capturer divers comportements de séries temporelles.

Ce modèle est particulièrement utile pour les prévisions, car il peut s'adapter à des schémas complexes de données avec des variations dans le niveau, la tendance, ou la saisonnalité. Le modèle ETS se concentre sur la décomposition des séries temporelles en différentes composantes, offrant ainsi une interprétation directe et intuitive des dynamiques sous-jacentes.

Les composantes du modèle :

Les modèles ETS se composent de trois éléments principaux qui peuvent être combinés de différentes manières :

1. **La composante d'erreur** représente les écarts entre les valeurs observées et les valeurs prévues.

L'erreur peut être :

- *Additive (A)* : Les écarts sont constants en amplitude, quelle que soit la valeur de la série.
- *Multiplicative (M)* : Les écarts sont proportionnels au niveau de la série, augmentant ou diminuant en fonction de celui-ci.

2. **La composante de tendance** reflète l'évolution générale de la série au cours du temps. Elle peut être :

- *Nulle (N)* : Absence de tendance dans la série.
- *Additive (A)* : La tendance est modélisée comme un changement linéaire constant dans le temps.
- *Additive amortie (A_d)* : La tendance linéaire est progressivement atténuée par un facteur d'amortissement ϕ , ce qui limite sa croissance ou décroissance à long terme.

3. **La composante de saisonnalité** capture les variations périodiques récurrentes dans les données. Elle peut être :

- *Nulle (N)* : Pas d'effet saisonnier.
- *Additive (A)* : L'effet saisonnier est constant au fil du temps.
- *Multiplicative (M)* : L'effet saisonnier varie proportionnellement au niveau de la série.

Les combinaisons de composantes :

Les différentes combinaisons de composantes permettent de définir plusieurs modèles ETS, chacun étant représenté par une notation tripartite (E, T, S) où chaque lettre représente respectivement l'erreur, la tendance, et la saisonnalité. Par exemple, un modèle ETS(A, A, N) présente une erreur additive, une tendance additive et aucune composante saisonnière.

Les tableaux ci-dessous présentent les différentes combinaisons possibles :

Erreur additive :

		<i>Saisonnalité</i>		
		N	A	M
<i>Tendance</i>	N	ETS (A, N, N)	ETS (A, N, A)	ETS (A, N, M)
	A	ETS (A, A, N)	ETS (A, A, A)	ETS (A, A, M)
	A_d	ETS (A, A_d , N)	ETS (A, A_d , A)	ETS (A, A_d , M)

Erreur multiplicative :

		<i>Saisonnalité</i>		
		N	A	M
<i>Tendance</i>	N	ETS (M, N, N)	ETS (M, N, A)	ETS (M, N, M)
	A	ETS (M, A, N)	ETS (M, A, A)	ETS (M, A, M)
	A_d	ETS (M, A_d , N)	ETS (M, A_d , A)	ETS (M, A_d , M)

TABLE 1.1 – Les combinaisons des modèles ETS

La formulation mathématique :

La formulation mathématique d'un modèle ETS dépend de la nature de ses composantes, comme l'illustre les exemples suivants :

- Le modèle ETS(A,Ad,N) est défini par :

$$\begin{cases} y_t = \ell_{t-1} + \phi b_{t-1} + \epsilon_t, & \text{(équation de prédiction)} \\ \ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \epsilon_t, & \text{(équation de lissage du niveau)} \\ b_t = \phi b_{t-1} + \beta \epsilon_t, & \text{(équation de lissage de la tendance)} \end{cases}$$

où :

- y_t : la valeur observée de la série temporelle à l'instant t .
- ℓ_t : le niveau estimé à l'instant t .
- α : le paramètre de lissage du niveau.
- b_t : la pente de la tendance.
- β : le paramètre de lissage de la tendance.
- ϕ : le paramètre de lissage de la tendance amortie.
- ϵ_t : le terme d'erreur.

- Le modèle ETS(A,Ad,A) est défini par :

$$\begin{cases} y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \epsilon_t, & \text{(équation de prédiction)} \\ \ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \epsilon_t, & \text{(équation de lissage du niveau)} \\ b_t = \phi b_{t-1} + \beta \epsilon_t, & \text{(équation de lissage de la tendance)} \\ s_t = s_{t-m} + \gamma \epsilon_t, & \text{(équation de lissage de la saisonnalité)} \end{cases}$$

où :

- s_t : le terme de saisonnalité.
- γ : le paramètre de lissage de la saisonnalité.

- Le modèle ETS(M,A,A) est défini par :

$$\begin{cases} y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \epsilon_t), & \text{(équation de prédiction)} \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t, & \text{(équation de lissage du niveau)} \\ b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t, & \text{(équation de lissage de la tendance)} \\ s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\epsilon_t, & \text{(équation de lissage de la saisonnalité)} \end{cases}$$

L'ensemble des formulations mathématiques possibles est présenté en annexe.

La sélection du modèle ETS adapté :

Pour sélectionner le modèle ETS le plus adapté parmi les différentes configurations possibles, des critères tels que le Critère d'Information d'Akaike (AIC), le Critère d'Information d'Akaike corrigé (AICc) et le Critère d'Information Bayésien (BIC) sont utilisés. Ces critères permettent de comparer les modèles en prenant en compte leur ajustement aux données et leur complexité.

Pour rappel, le Critère d'Information Bayésien (BIC) est défini par :

$$BIC = 2 \operatorname{Log}(L) + k \log(n)$$

où :

- $\log(L)$ est le logarithme de la fonction de vraisemblance maximale du modèle, représentant la qualité de l'ajustement aux données.
- k est le nombre de paramètres du modèle qui mesure la complexité du modèle.
- n est le nombre d'observations dans les données.

Le Critère d'Information d'Akaike (AIC) est défini par :

$$AIC = -2\log(L) + 2k$$

où :

- $\log(L)$ est le logarithme de la fonction de vraisemblance maximale du modèle, représentant la qualité de l'ajustement aux données.
- k est le nombre de paramètres du modèle qui mesure la complexité du modèle.

Le Critère d'Information d'Akaike corrigé (AICc) est défini par :

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

où $\frac{2k(k+1)}{n-k-1}$ représente la correction du biais, k étant le nombre de paramètres du modèle et n la taille de l'échantillon.

Le modèle présentant la plus faible valeur de ces critères est généralement considéré comme le meilleur.

Pour évaluer la performance d'un modèle prédictif, il est essentiel de vérifier qu'il ne présente pas de sur-apprentissage (*overfitting*), c'est-à-dire qu'il ne s'ajuste pas trop précisément aux spécificités des données d'entraînement, au point de perdre sa capacité à généraliser à de nouvelles données. Un modèle en sur-apprentissage peut obtenir d'excellents résultats sur les données d'entraînement, mais échouer lors de la prédiction de nouvelles observations.

Pour prévenir ce problème, une méthode couramment utilisée est la validation croisée (*cross-validation*). Elle consiste à diviser l'ensemble des données disponibles en plusieurs sous-ensembles et à répéter plusieurs itérations d'entraînement et de test. Le modèle est ajusté sur une partie de ces sous-ensembles (données d'entraînement), qui représentent généralement 70% à 80% des données totales, et testé sur les sous-ensembles restants (données de validation).

En répétant ce processus avec différentes partitions des données, cette méthode permet de vérifier la stabilité des performances du modèle et de s'assurer qu'elles ne dépendent pas d'un seul ensemble de test, réduisant ainsi la variabilité des résultats et minimisant le risque de surapprentissage.

Lorsque les données ne sont pas ordonnées, la séparation des bases d'entraînement et de validation est réalisée de manière aléatoire. Cette approche classique est illustrée ci-dessous.

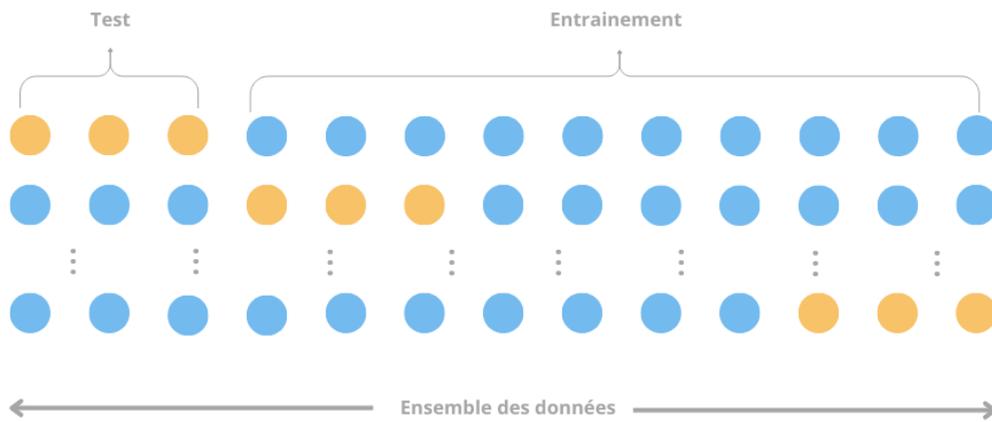


FIGURE 1.1 – La validation croisée dans le cas de données non ordonnées

En revanche, lorsque les données sont ordonnées, comme c'est le cas pour les séries temporelles, la validation croisée classique n'est pas directement applicable en raison de la nature séquentielle des données. Les observations étant corrélées dans le temps, il est essentiel de respecter l'ordre chronologique. Dans ce contexte, une méthode spécifique appelée validation croisée avec découpage temporel (*Time Series Cross-Validation*) est utilisée. Les données sont divisées en plusieurs "blocs" où les premières parties sont utilisées pour l'entraînement et les parties suivantes pour la validation. À chaque itération, un nouveau bloc de données est ajouté à l'ensemble d'entraînement, et les prédictions sont testées sur le bloc suivant. Cette approche est illustrée dans le graphique suivant.

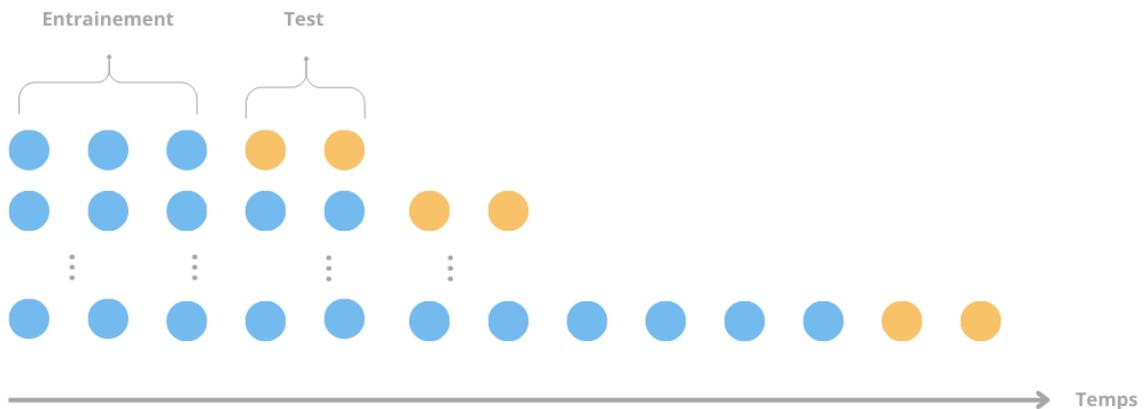


FIGURE 1.2 – La validation croisée dans le cas de données ordonnées

1.1.2 L'application aux données

Le modèle ETS se distingue par sa flexibilité à combiner différentes composantes : erreurs additives ou multiplicatives, tendances linéaires ou amorties, et effets saisonniers. Cette flexibilité est particulièrement pertinente dans l'analyse de la mortalité en France, où les fluctuations récentes des taux de mortalité pourraient refléter des changements structurels sous-jacents.

Dans ce contexte, le modèle ETS devrait intuitivement présenter une tendance linéaire amortie, correspondant à une décroissance progressive de la composante temporelle κ_t au fil du temps. Cette tendance reflète un ralentissement des améliorations de la mortalité, suggérant que la diminution des taux ne suit plus un rythme constant, mais tend à s'atténuer. Une erreur additive est définie. Enfin, étant donné que les données sont annuelles et qu'aucune variation saisonnière significative n'est observée, il est raisonnable de ne pas inclure de composante saisonnière.

Pour vérifier cette intuition, plusieurs configurations du modèle ETS seront modélisées et comparées :

- **ANN** : une erreur additive, aucune tendance et aucune saisonnalité.
- **AAN** : une erreur additive, une tendance additive et aucune saisonnalité.
- **AAdN** : une erreur additive, une tendance amortie et aucune saisonnalité.
- **ANA** : une erreur additive, aucune tendance et une saisonnalité additive.
- **AAA** : une erreur additive, une tendance additive et une saisonnalité additive.
- **AAdA** : une erreur additive, une tendance amortie et une saisonnalité additive.

Les graphiques suivants illustrent les projections de mortalité pour les différentes configurations appliquées aux données des femmes et des hommes. Ces graphiques permettent de comparer visuellement comment chaque modèle reproduit les tendances historiques et projette les évolutions futures.

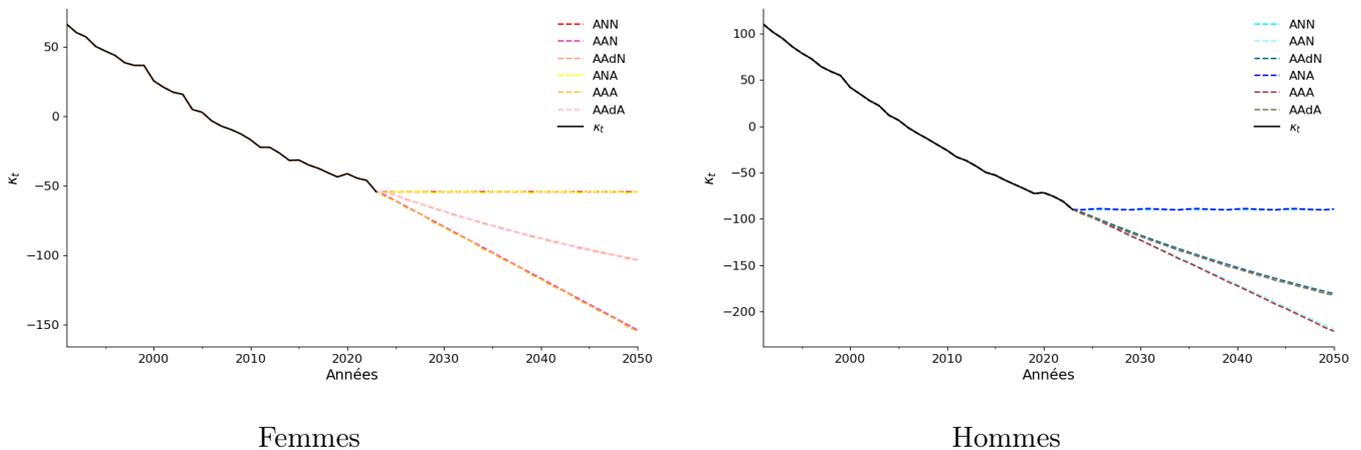


FIGURE 1.3 – Les projections avec différentes combinaisons ETS

Les modèles ETS sans composante de tendance n'indiquent aucune poursuite de la diminution de la mortalité et se contentent de projeter une stabilisation. Cela limite leur pertinence pour explorer les dynamiques de mortalité.

Avec l'application d'une tendance amortie, les projections montrent que la diminution de κ_t devient moins accentuée. Cela reflète un ralentissement progressif des améliorations de la mortalité. Ce mémoire se concentre sur l'exploration de ce ralentissement des améliorations de la mortalité, justifiant ainsi l'utilisation d'une tendance amortie pour mieux capturer ces dynamiques actuelles de mortalité et fournir des projections robustes.

Compte tenu de ces observations, la comparaison des modèles est affinée en se concentrant sur les configurations AAdA (erreur additive, tendance amortie, saisonnalité additive) et AAdN (erreur additive, tendance amortie, aucune saisonnalité). Pour déterminer le modèle le plus adapté, les critères AIC, AICc, et BIC sont calculés et comparés entre ces deux configurations.

	Femmes		Hommes	
	AAdN	AAdA	AAdN	AAdA
BIC	173,82	193,9	159,5	181,11
AIC	164,84	175,94	150,52	163,15
AICc	168,07	191,54	153,77	178,75

TABLE 1.2 – Les critères d’information du modèle ETS

Le modèle AAdN qui combine une erreur additive, une tendance amortie et aucune composante saisonnière, présente de meilleurs résultats pour la modélisation de la dynamique temporelle de mortalité. Le choix de ce modèle s’aligne avec l’intuition initiale selon laquelle une tendance amortie reflète correctement le ralentissement progressif des améliorations de la mortalité en France, sans nécessiter d’ajouter une composante saisonnière, étant donné la nature annuelle des données et l’absence de variations saisonnières significatives. Dans le cadre de ce mémoire, qui vise à élaborer une table de mortalité prospective pouvant servir de référence si les changements récents de dynamique de mortalité se révèlent être structurels, le modèle AAdN constitue une approche robuste pour projeter les tendances futures.

1.2 Le modèle PROPHET de Facebook

1.2.1 La présentation du modèle

Le modèle PROPHET, développé en 2017 par une équipe de data scientists de Facebook dirigée par S.J. TAYLOR et B. LETHAM, est décrit, dans leur article *Forecasting at Scale*, comme une solution innovante pour la prévision des séries temporelles. PROPHET est une alternative aux modèles plus classiques, tels qu’ARIMA. Celui-ci est accessible tout en conservant une grande flexibilité et une robustesse face aux données manquantes, aux valeurs aberrantes, et aux changements de tendance.

Conçu pour être utilisé aussi bien par des data scientists que par des experts métiers, PROPHET répond aux besoins de prévision à grande échelle. Son principal atout réside dans sa capacité à intégrer directement les ajustements et l’expertise des analystes dans le modèle, un concept que les auteurs nomment "*analyst-in-the-loop*". Cela permet aux experts de domaine d’affiner les prévisions en fonction de leurs connaissances, rendant ainsi le modèle plus intuitif et mieux adapté aux réalités opérationnelles.

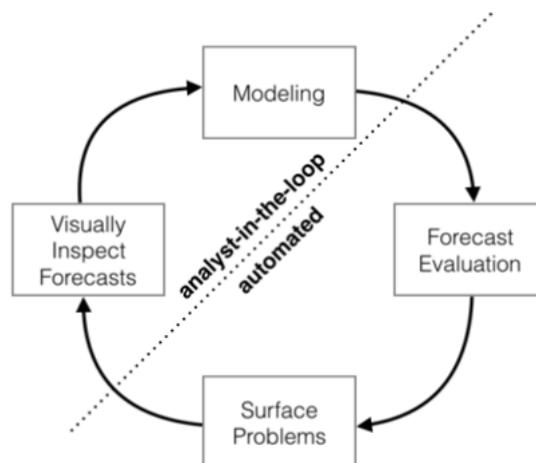


FIGURE 1.4 – L’*analyst-in-the-loop*

Source : *Forecasting at Scale*

Ce modèle additif, décomposé en plusieurs composantes, notamment la tendance, la saisonnalité et les effets d'évènements particuliers, est également capable d'incorporer des régresseurs externes pour enrichir les prévisions et les adapter à des contextes spécifiques. C'est un outil puissant pour la prévision des séries temporelles, offrant à la fois précision et interprétabilité.

La définition du modèle :

Mathématiquement, le modèle (sans ajout de régresseur) s'écrit comme suit :

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Avec :

- $g(t)$: modélisant les changements non périodiques de la série temporelle.
- $s(t)$: représentant les changements périodiques de la série temporelle.
- $h(t)$: représentant les effets d'une série temporelle selon des calendriers potentiellement irréguliers.
- ϵ_t : étant le terme d'erreur, supposé normalement distribué.

Les composantes du modèle :

1. **Le trend** : Le modèle de tendance dans PROPHET comprend deux approches pour décrire l'évolution d'une série temporelle dans le temps : un modèle de croissance saturée non linéaire et un modèle de tendance linéaire avec des changements de tendance.

- Dans le cas d'une **croissance saturée non linéaire**, PROPHET utilise une fonction logistique pour modéliser la croissance d'une série temporelle. Ce modèle s'inspire des dynamiques naturelles, telles que la croissance d'une population dans un écosystème, où la croissance commence rapidement mais finit par ralentir en atteignant un point de saturation appelé capacité de charge.

La fonction logistique utilisée par PROPHET est définie par l'équation suivante :

$$g(t) = \frac{C}{1 + \exp(-k(t - m))}$$

où

- $g(t)$ représente la tendance générale de la série au fil du temps.
- C est la capacité de charge, ou le niveau maximum que la série peut atteindre.
- k est le taux de croissance. Ce paramètre contrôle la vitesse à laquelle la série se rapproche de sa capacité de charge.
- m est le point médian, indiquant le moment où la moitié de la capacité maximale est atteinte.

Le modèle de croissance logistique standard présente deux limitations majeures : une capacité limite fixe et un taux de croissance constant. PROPHET surmonte ces limitations en introduisant des capacités variables dans le temps, $C(t)$, et des taux de croissance ajustables en fonction de points de changement, s_j avec $j = 1, \dots, s$. À chaque point de changement, une modification du taux de croissance est définie par un vecteur d'ajustement des taux $\delta \in \mathbb{R}^s$, où chaque δ_j représente la variation du taux de croissance à l'instant s_j .

Ces points de changements peuvent être définis manuellement, lorsque l'analyse à connaissance de nouvelles dynamiques. Cependant, PROPHET offre également la possibilité d'une sélection automatique des points de changement pour ajuster le modèle de manière optimale aux données historiques. Placés à des intervalles de temps réguliers sur les données d'entraînement, ces points de changement δ_j suivent une distribution de Laplace centrée autour de zéro avec un paramètre d'échelle τ :

$$\delta_j \sim \text{Laplace}(0, \tau)$$

Le paramètre τ contrôle directement la flexibilité du modèle dans l'altération de son taux de croissance. Une valeur élevée de τ permet des ajustements plus fréquents.

Le taux de croissance à l'instant t devient alors le taux de base k , auquel s'ajoute les ajustements accumulés jusqu'à ce point :

$$k(t) = k + \sum_{j|t > s_j} \delta_j$$

PROPHET définit un vecteur indicateur $a(t) \in 0, 1^S$ tel que :

$$a_j(t) = \begin{cases} 1, & \text{si } t \geq s_j \\ 0, & \text{sinon} \end{cases}$$

Ainsi, le taux de croissance à l'instant t peut se réécrire comme suit :

$$k(t) = k + a(t)^T \delta$$

Pour assurer la continuité de la courbe aux points de changement, le paramètre de décalage m doit également être ajusté. Ainsi l'ajustement approprié au point de changement j est calculé par :

$$\gamma_j = \left(s_j - m - \sum_{l < j} \gamma_l \right) \left(1 - \frac{k + \sum_{l \leq j} \delta_l}{k + \sum_{l < j} \delta_l} \right)$$

En incorporant ces ajustements, le modèle de croissance logistique devient :

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))}$$

Dans les projections, PROPHET simule des changements de taux de croissance similaires à ceux observés dans le passé. Concrètement, PROPHET suppose qu'il y a S points de changement sur un historique de T points, chacun avec un changement de taux $\delta_j \sim \text{Laplace}(0, \tau)$. Les points de changement futurs sont alors échantillonnés aléatoirement de manière à ce que la fréquence moyenne des points de changement corresponde à celle de l'historique :

$$\forall j > T, \begin{cases} \delta_j = 0 & \text{avec la probabilité } \frac{T-S}{T}, \\ \delta_j \sim \text{Laplace}(0, \lambda) & \text{avec la probabilité } \frac{S}{T}. \end{cases}$$

- Dans certains cas, la croissance d'une série temporelle ne présente pas de saturation et suit plutôt une progression linéaire. Ainsi, PROPHET peut incorporer une **tendance linéaire** qui peut également intégrer des points de changement pour capturer les variations de croissance à des moments spécifiques.

Le modèle de croissance linéaire est défini par :

$$g(t) = \underbrace{(k + a(t)^T \delta)}_{\text{La pente}} t + \underbrace{(m + a(t)^T \gamma)}_{\text{Le décalage}}$$

où les paramètres se définissent comme ceux présents dans la tendance saturée.

2. **La saisonnalité** : Les effets saisonniers dans PROPHET sont modélisés avec des séries de Fourier, ce qui permet de capturer des schémas périodiques dans les données. La fonction de saisonnalité est exprimée comme suit :

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi n t}{P} \right) + b_n \sin \left(\frac{2\pi n t}{P} \right) \right)$$

où :

- P est la période de la série temporelle. Par exemple, pour des données annuelles, $P = 365, 25$ jours.
- N représente le nombre de termes utilisés pour approcher la saisonnalité. En limitant la série à N termes, un filtre est appliqué permettant de gérer la complexité du modèle. Un N plus élevé rend le modèle capable de capturer des variations saisonnières rapides, mais accroît aussi le risque de surapprentissage.
- a_n et b_n sont les coefficients de Fourier à estimer.

Pour imposer un lissage de la saisonnalité, $\beta = [a_1, b_1, a_2, b_2, \dots, a_N, b_N]$ est supposé tel que $\beta \sim \mathcal{N}(0, \sigma^2)$.

Ainsi, la matrice de saisonnalité suivante est construite :

$$X(t) = \left[\cos \left(\frac{2\pi(1)t}{P} \right), \sin \left(\frac{2\pi(1)t}{P} \right), \dots, \dots, \cos \left(\frac{2\pi(N)t}{P} \right), \sin \left(\frac{2\pi(N)t}{P} \right) \right]$$

La composante de saisonnalité peut donc se réécrire comme suit :

$$s(t) = \beta X(t)$$

PROPHET permet de modéliser simultanément plusieurs types de saisonnalités. Par défaut, il intègre une composante de saisonnalité annuelle et hebdomadaire, mais d'autres périodes saisonnières peuvent être ajoutées selon les exigences de l'analyse.

Cette approche flexible permet à PROPHET de capturer efficacement les effets saisonniers dans une large variété de séries temporelles, tout en offrant la possibilité d'affiner les paramètres pour mieux correspondre aux particularités des données étudiées.

3. **Les évènements particuliers** : Le modèle PROPHET permet d'ajouter des évènements spécifiques, notés ($i = 1, \dots, L$), qui sont indépendants les uns des autres et dont les dates D_i peuvent varier dans le temps. Ces évènements représentent des occurrences exceptionnelles ou irrégulières (comme des jours fériés ou des évènements uniques) qui peuvent influencer de manière significative la série temporelle.

Une fonction indicatrice est utilisée pour déterminer si une date t correspond à l'une des dates D_i de l'évènement i . Elle est définie comme suit :

$$Z(t) = [\mathbb{1}_{t \in D_1}, \dots, \mathbb{1}_{t \in D_L}]$$

où $\mathbb{1}_{t \in D_i}$ prend la valeur 1 si t est une date de l'évènement D_i et 0 sinon.

À chaque évènement i , est associé un paramètre κ_i représentant l'impact de cet évènement sur la prévision. Ce paramètre suit une distribution normale $\kappa \sim \mathcal{N}(0, \nu^2)$, où ν^2 est la variance qui régule l'amplitude de l'effet de chaque évènement.

Ainsi, la composante des évènements peut être modélisée par l'expression :

$$h(t) = kZ(t)$$

où $h(t)$ représente l'impact total de tous les évènements sur la série temporelle à l'instant t .

Cette approche flexible permet de capturer l'effet ponctuel ou irrégulier des évènements sur les prévisions, en tenant compte de leur nature et de leur intensité spécifiques.

L'ajout de régresseurs externes :

PROPHET permet d'enrichir le modèle en intégrant des régresseurs externes, notés $x_i(t)$, c'est-à-dire des variables supplémentaires qui influencent la série temporelle d'intérêt, y_t . Ces régresseurs sont généralement additifs. Ainsi, mathématiquement, l'ajout de régresseurs se traduit par une extension de l'équation du modèle PROPHET :

$$y(t) = g(t) + s(t) + h(t) + \sum_{i=1}^k \beta_i x_i(t) + \epsilon_t$$

où :

- $g(t)$ représente le terme de tendance de la série temporelle.
- $s(t)$ est la composante saisonnière.
- $h(t)$ est la composante des évènements particuliers.
- $x_i(t)$ sont les régresseurs externes ajoutés au modèle. Ceux-ci doivent être connus et disponibles sur toute la période historique utilisée pour l'entraînement du modèle et sur la période de projection.
- β_i sont les coefficients estimés associés aux régresseurs externes, qui quantifient l'influence de chaque régresseur $x_i(t)$ sur la série temporelle cible.
- ϵ_t est le terme d'erreur, supposé normalement distribué.

1.2.2 L'application aux données

Dans cette section, plusieurs configurations du modèle PROPHET sont testées pour modéliser la dynamique de la composante temporelle de la mortalité, κ_t , en France. Ces différentes approches permettent d'explorer les tendances linéaires et saturées, ainsi que les effets potentiels des points de changement et de la saisonnalité, afin de déterminer le modèle le plus adapté aux données observées.

- **Une tendance linéaire avec des points de changement déterminés automatiquement par le modèle sur l'ensemble des données :** Cette première approche utilise une tendance linéaire avec des points de changement automatiquement sélectionnés sur l'ensemble des données disponibles, afin de capturer les changements de tendance significatifs tout au long de la série temporelle. Cette méthode est particulièrement pertinente pour intégrer les dernières années, marquées par des dynamiques de mortalité différentes.

Le critère de prise en compte des points de changement est déterminé en fonction de l'erreur quadratique moyenne (MSE). Différentes valeurs de flexibilité ont été testées pour mesurer leur impact sur la précision des prévisions. Le tableau ci-dessous montre les résultats obtenus :

Flexibilité du modèle	MSE Femmes	MSE Hommes
0,05	24,28	14,04
0,1	16,74	10,65
0,2	12,40	21,73
0,3	13,56	39,24
0,4	28,75	50,75
0,5	32,70	34,25

TABLE 1.3 – Les valeurs du MSE selon la flexibilité du modèle

L'analyse indique que la flexibilité de 0,3 permet au modèle d'être suffisamment sensible aux changements significatifs tout en minimisant le MSE pour les femmes, et d'éviter une réactivité excessive aux variations mineures. Cette valeur a également été retenue pour les hommes, car elle offre un compromis satisfaisant entre sensibilité et stabilité du modèle par rapport aux dynamiques des données historiques.

Les graphiques ci-dessous présentent les prédictions obtenues à partir de cette approche.

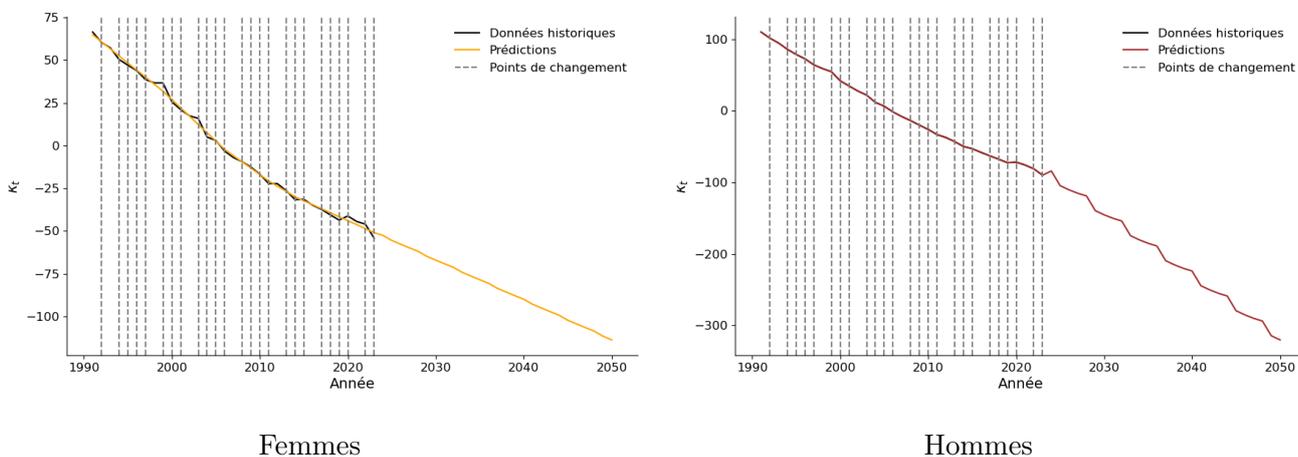


FIGURE 1.5 – Les projections avec le modèle de PROPHET (Approche 1)

Il est notable que les points de changement sont positionnés presque tous les ans, ce qui entraîne un risque d'*overfitting*, en particulier pour les hommes, où les courbes d'estimation et les valeurs réelles se superposent étroitement. Cela est corroboré par le faible MSE observé pour les hommes, égal à 0,0257, qui suggère un sur-ajustement aux données historiques.

- **Une tendance linéaire avec une limitation du nombre de points de changement :** Afin de réduire le risque de surajustement, une nouvelle approche est testée avec un nombre limité de points de changement, fixé à 6 entre 1991 et 2023. Cette méthode permet de capturer les changements de tendance les plus significatifs tout en évitant une réactivité excessive aux variations mineures des données historiques.

Comme pour la première approche, le critère de prise en compte des points de changement est déterminé en fonction de l'erreur quadratique moyenne (MSE). Une flexibilité de 0,3, offrant un compromis satisfaisant entre sensibilité et stabilité du modèle, est retenue pour les femmes et les hommes.

Les graphiques ci-dessous présentent les prédictions obtenues à partir de cette approche.

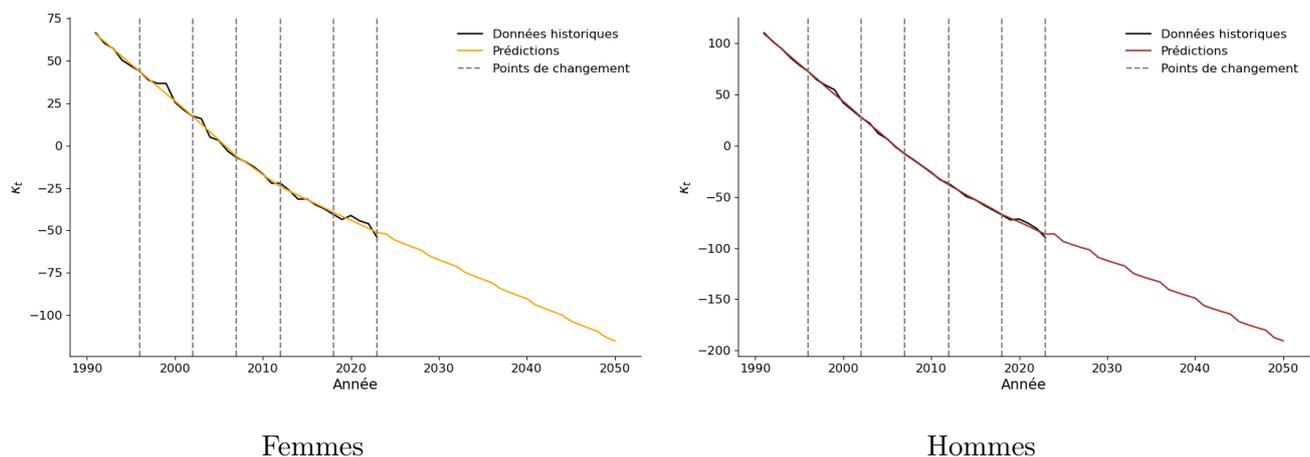


FIGURE 1.6 – Les projections avec le modèle de PROPHET (Approche 2)

Cette approche offre un bon ajustement du modèle aux données historiques. Les points de changement sélectionnés capturent efficacement les variations significatives de la mortalité, notamment la "bosse" observée ces dernières années, avec des points placés judicieusement en 2018 et 2023.

Pour les hommes, le risque de surajustement semble être mieux maîtrisé comparé à l'approche précédente. La courbe de prédiction suit correctement la tendance observée sans surréagir aux fluctuations mineures des données. Le MSE a augmenté par rapport à l'approche 1, mais il reste relativement faible, 2,25, confirmant que cette approche offre un compromis efficace entre précision et stabilité.

Ainsi, cette approche se révèle être plus robuste, que la précédente, en particulier pour éviter le surajustement, tout en intégrant de manière pertinente les dynamiques récentes de la mortalité.

- **Une tendance linéaire avec l'inclusion d'une composante de saisonnalité :** Cette configuration teste l'inclusion d'une composante saisonnière dans le modèle. L'objectif de cette approche est d'examiner si des schémas saisonniers significatifs peuvent être capturés dans la dynamique de la mortalité, ce qui justifierait l'ajout de cette composante au modèle. Des tests sont effectués pour

différents ordres de la série de Fourier afin de déterminer le niveau optimal de complexité pour la composante saisonnière. Le choix du modèle est basé sur la minimisation de l'erreur quadratique moyenne (MSE).

Ordres de Fourier	MSE Femmes	MSE Hommes
1	3,45	2,17
2	6,32	4,35
3	6,59	4,47

TABLE 1.4 – Les valeurs du MSE selon la saisonnalité

L'ordre de Fourier optimal pour capturer les schémas saisonniers dans la dynamique de la mortalité, sans provoquer de surajustement, semble être l'ordre 1.

Les graphiques ci-dessous illustrent les prédictions obtenues avec l'inclusion d'une composante de saisonnalité par rapport aux prédictions sans composante saisonnière.

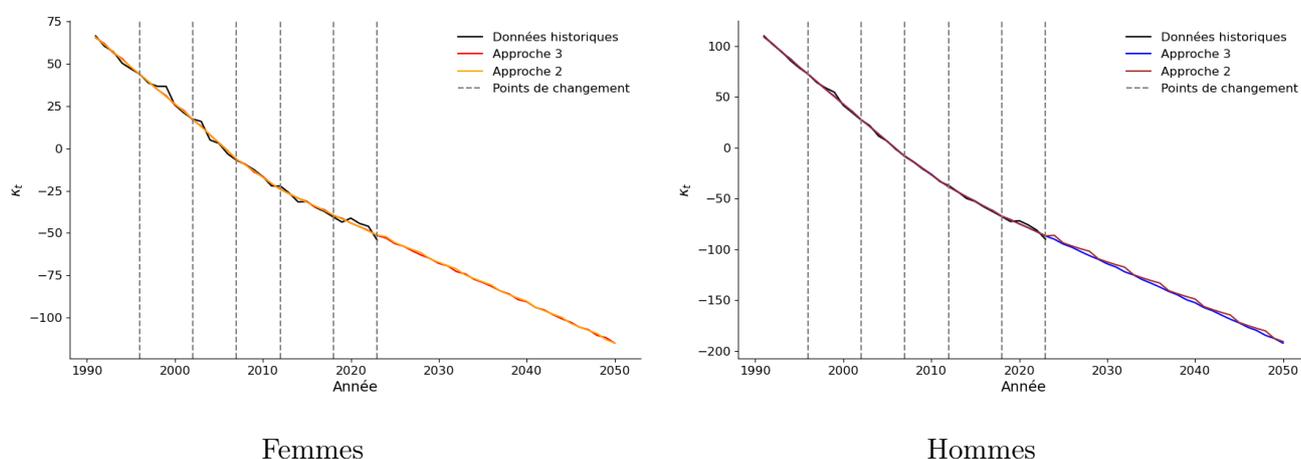


FIGURE 1.7 – Les projections avec le modèle de PROPHET (Comparaison approches 2 et 3)

Les prévisions des deux modèles sont très similaires. Toutefois, le modèle avec une composante de saisonnalité présente des projections légèrement plus stables au fil du temps. Les erreurs quadratiques moyennes sont légèrement plus faibles pour le modèle incluant la saisonnalité (3,45 pour les femmes contre 3,62 sans saisonnalité, et 2,17 pour les hommes contre 2,25 sans saisonnalité).

Cette approche qui combine une tendance linéaire avec 6 points de changement et une composante de saisonnalité d'ordre 1 est retenue pour la suite de ce mémoire.

- **Une tendance saturée :** Cette approche modélise une tendance saturée avec 6 points de changement fixés automatiquement sur l'ensemble de la période, incluant une composante de saisonnalité d'ordre de Fourier égal à 1. Le seuil minimum de saturation est ajusté pour refléter une limite réaliste pour les femmes et les hommes. Il est déterminé en fonction de la borne inférieure en 2050 du modèle ARIMA, présenté dans la partie précédente de ce mémoire, qui prend en compte la bosse de mortalité (entraîné sur la période 1991-2023). Ce seuil est fixé à -200 pour les femmes et -300 pour les hommes.

Comme le montrent les graphiques ci-dessous, les prédictions obtenues avec une tendance linéaire sont très proches de celles obtenues avec une tendance saturée.

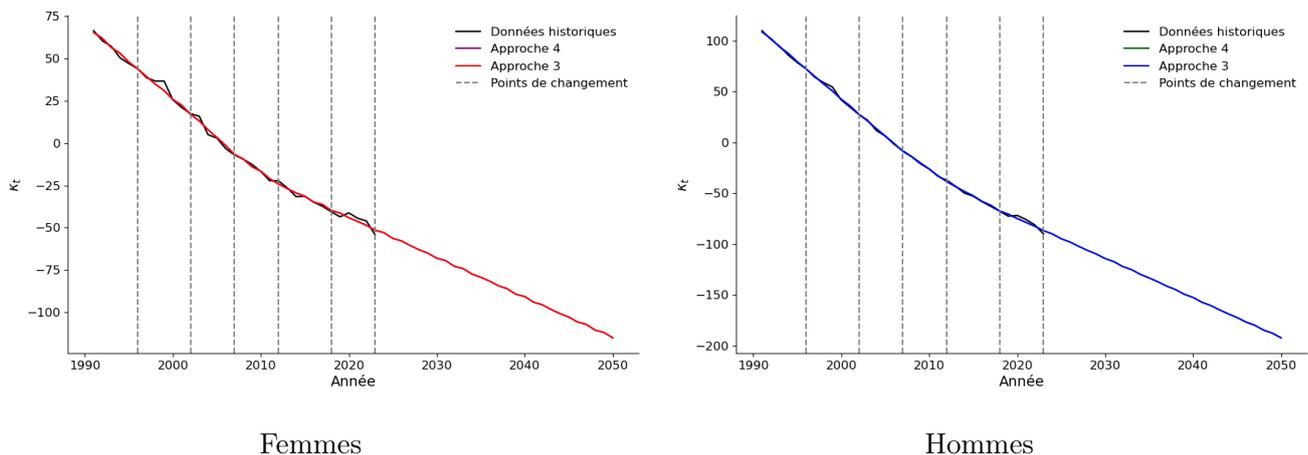


FIGURE 1.8 – Les projections avec le modèle de PROPHET (Comparaison approches 3 et 4)

Les ajustements obtenus avec les deux approches sont très proches, les courbes se superposant quasiment. En termes de MSE, les résultats sont également similaires : pour les femmes, l’approche 3 présente un MSE de 3,45 contre 3,50 pour l’approche 4, et pour les hommes, le MSE est de 2,17 pour l’approche 3 et de 2,47 pour l’approche 4.

Au vu de ces résultats, seule l’approche 3, qui modélise une tendance linéaire avec 6 points de changement choisis automatiquement sur l’ensemble des données et inclut une composante de saisonnalité d’ordre de Fourier égal à 1, est retenue pour la suite de ce mémoire.



Une idée d’exploration : Dans le cadre de ce mémoire, il aurait été intéressant d’explorer l’intégration de régresseurs dans le modèle PROPHET pour affiner l’analyse des dynamiques de mortalité. Dans la première partie de ce mémoire, plusieurs facteurs influençant la mortalité ont été identifiés, tels que les conditions climatiques extrêmes, la qualité de l’air et les comportements individuels. L’inclusion de régresseurs, comme les chroniques de température, les épisodes d’épidémies de grippe, ou les niveaux de pollution atmosphérique aurait pu apporter une précision supplémentaire aux projections en capturant l’impact de ces événements spécifiques sur la mortalité. De même, des indicateurs comme la consommation de tabac ou la prévalence de maladies chroniques auraient enrichi l’analyse en offrant une perspective plus complète sur les influences de comportements de santé et des conditions médicales. Bien que cette approche ne soit pas développée ici, elle offrirait une perspective d’étude intéressante pour enrichir les projections futures.

1.3 La comparaison des modèles alternatifs : ETS et PROPHET

L’objectif de cette section est de comparer les différentes approches retenues précédemment. Les deux modèles suivants sont ainsi examinés :

- **Le modèle ETS (Erreur-Tendance-Saisonnalité [AAAdN])** : Ce modèle est caractérisé par une erreur additive, une tendance additive amortie, et l’absence de composante saisonnière.
- **Le modèle PROPHET** : basé sur une tendance linéaire avec 6 points de changement, fixés automatiquement sur l’ensemble de la période d’observation, et une composante de saisonnalité d’ordre de Fourier égal à 1.

Les graphiques ci-dessous présentent les projections des deux modèles pour chaque sexe. Bien que ces projections soient globalement similaires, de légères divergences apparaissent au fur et à mesure que l’horizon de projection s’étend.

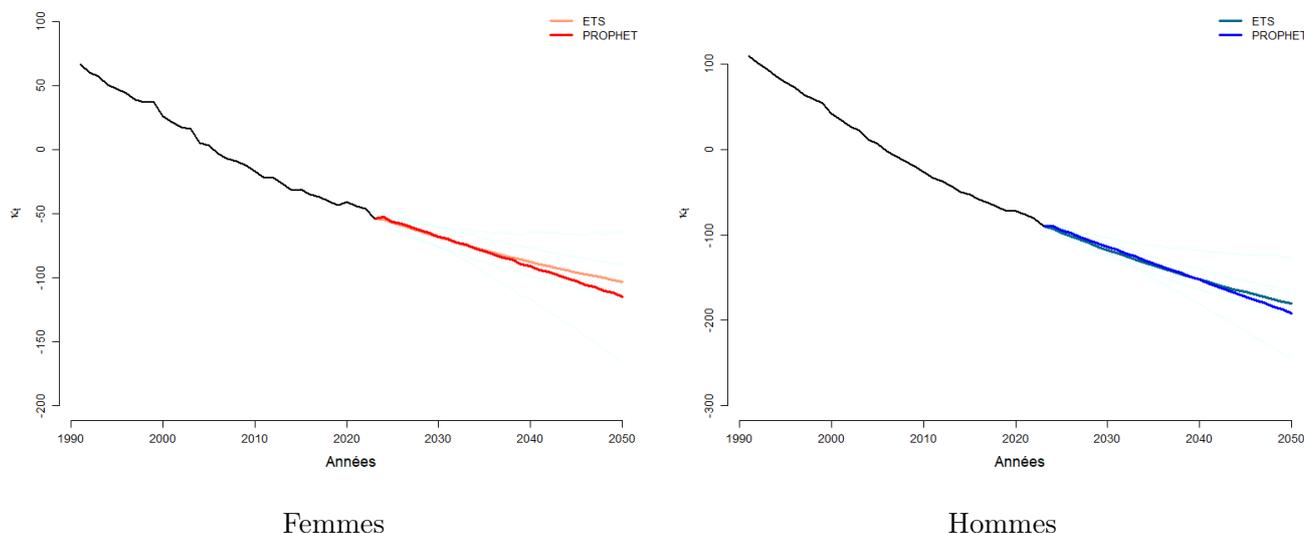


FIGURE 1.9 – Les projections avec les modèles ETS et PROPHET

Les erreurs quadratiques moyennes sont calculées pour chaque modèle et pour chaque sexe :

	Femmes	Hommes
ETS	6	3,89
PROPHET	3,45	2,17

TABLE 1.5 – Les valeurs du MSE des modèles ETS et PROPHET

Le modèle de PROPHET minimise les erreurs de prédiction par rapport au modèle ETS. Cette méthode alternative de projection est ainsi retenue pour la suite de ce mémoire.



Une idée d’exploration : Dans ce mémoire, les modèles ETS et PROPHET ont été utilisés comme alternatives au modèle ARIMA pour modéliser les dynamiques de mortalité. D’autres approches auraient pu être explorées, comme les réseaux de neurones récurrents (RNN) et les modèles *Long Short-Term Memory* (LSTM), qui capturent des relations complexes dans les séries temporelles. Des modèles récents comme DeepAR, développé par Amazon, combinant des approches probabilistes avec des techniques de *Machine Learning* avancées, ou Neural Prophet, une extension de PROPHET, intégrant des réseaux de neurones, auraient également offert des alternatives intéressantes pour enrichir les projections, notamment dans un contexte de changements potentiellement structurels.

1.4 La comparaison du modèle classique et du modèle alternatif retenu

Il est essentiel de comparer les projections obtenues avec le modèle classique (une marche aléatoire avec dérive) et le modèle PROPHET pour évaluer leur pertinence. Si les changements observés récemment dans les tendances de mortalité s'avèrent être de nature structurelle, il pourrait être nécessaire de réévaluer l'efficacité des méthodes de projection classiques et de considérer des approches plus complexes telles que PROPHET, qui intègrent plus facilement des variations et des changements de tendance. Les deux modèles ont été appliqués à la projection de la composante temporelle, κ_t , du modèle de RENSHAW & HABERMAN, comme illustré ci-dessous pour les hommes et les femmes.

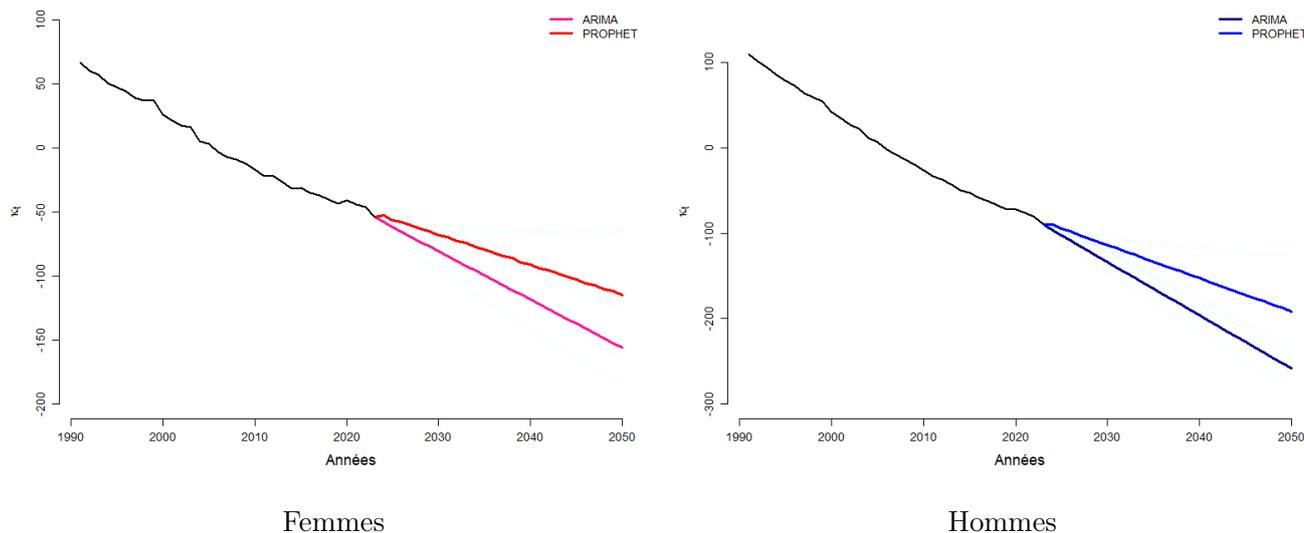


FIGURE 1.10 – Les projections avec les modèles ARIMA et PROPHET

Bien que les deux modèles, ARIMA et PROPHET, continuent d'indiquer une baisse des taux de mortalité au fil du temps, des différences notables émergent dans les résultats de ces projections. Le modèle PROPHET, en particulier, reflète plus fortement un potentiel ralentissement de l'amélioration de la mortalité, contrairement au modèle ARIMA qui tend à prolonger les tendances historiques de manière plus linéaire. En effet, le modèle PROPHET, en intégrant mieux les récents changements de tendance, suggère une décélération plus marquée des améliorations de la mortalité. À l'inverse, le modèle ARIMA maintient une perspective plus optimiste, alignée avec les tendances de baisse observées au cours des dernières décennies.

Cette divergence est confirmée par une analyse quantitative : l'écart moyen entre les projections des deux modèles s'élève en 2024 à 8,79% pour les hommes et 6,54% pour les femmes.

Ces différences soulignent l'importance de sélectionner le modèle de projection le plus approprié en fonction des dynamiques sous-jacentes de mortalité et des objectifs d'analyse. Dans un contexte où les récents changements de tendance pourraient être de nature structurelle, il est important de considérer des modèles plus flexibles, tels que PROPHET, qui sont capables de s'adapter à des variations potentielles dans les schémas de mortalité à long terme.



Une idée d'exploration : Ce mémoire se concentre sur la projection de l'amélioration de la mortalité dans le temps, en construisant différents scénarios à l'aide de diverses méthodes de projection. Il serait pertinent de les comparer à ceux élaborés par l'INSEE dans le cadre des projections de population 2021-2070, construites selon une méthode déterministe, fondée sur des tendances observées et enrichie par la consultation d'experts nationaux et internationaux.



Ce qu'il faut retenir du chapitre :

- * Le modèle ETS décompose les séries temporelles en trois composantes : erreur, tendance et saisonnalité, offrant une flexibilité dans le choix des paramètres. L'analyse montre que le modèle ETS avec erreur additive, tendance amortie, et sans saisonnalité (ETS AAdN) est le plus pertinent pour modéliser la composante temporelle, capturant le ralentissement progressif des améliorations de mortalité observé récemment.
- * Le modèle PROPHET, développé par Facebook, est un outil de prévision flexible et robuste pour les séries temporelles. L'analyse de différentes configurations montre que le modèle avec une tendance linéaire ajustée par 6 points de changement et une saisonnalité d'ordre 1 est le plus équilibré, offrant un bon ajustement aux données tout en minimisant le surajustement. Ce modèle, minimisant l'erreur quadratique moyenne, est retenu.
- * Bien que les modèles classique (ARIMA) et alternatif (PROPHET) projettent tous deux une baisse continue de la mortalité, le modèle PROPHET capture un ralentissement potentiel plus marqué des améliorations, contrairement au modèle ARIMA qui prolonge les tendances historiques de manière plus linéaire.

Le chapitre suivant se concentrera sur l'utilisation de la composante κ_t projetée par le modèle PROPHET pour construire une table de mortalité projetée, puis la méthode de DENUIT & GODERNIAUX sera appliquée pour extrapoler les taux aux grands âges.

Chapitre 2

Les nouvelles tables de mortalité

Ce chapitre présente l'élaboration de nouvelles tables de mortalité prospectives, développées à l'aide du modèle PROPHET, qui s'adapte aux récents changements de tendance. Contrairement aux méthodes classiquement utilisées, PROPHET offre une plus grande flexibilité grâce à l'intégration de points de changement, permettant une meilleure réactivité face aux dynamiques observées.

Les sections suivantes présentent la projection des taux de mortalité. Ces projections seront ensuite comparées à celles obtenues dans la partie précédente pour évaluer l'impact d'un changement de modèle. Les tables de mortalité seront également complétées aux grands âges à l'aide d'une méthode de fermeture.

2.1 La projection des taux de mortalité

Pour rappel, les taux de mortalité sont déterminés à l'aide du modèle de RENSHAW & HABERMAN, qui est une extension du modèle de LEE & CARTER incluant un terme de cohorte.

Comme dans la partie précédente, les paramètres α_x et β_x estimés et lissés sont utilisés.

Dans cette nouvelle approche, la composante temporelle κ_t a été projetée à l'aide du modèle PROPHET choisit pour sa capacité à s'adapter aux récents changements de tendance dans les dynamiques de mortalité, comme expliqué dans le chapitre précédent. Le modèle adopté est une tendance linéaire ajustée par 6 points de changement et une saisonnalité d'ordre 1.

Pour l'obtention des taux de mortalité projetés dans le temps, il a également été nécessaire de projeter la composante γ_{t-x} . Comme précédemment, un modèle ARIMA, ajusté sur les données de 1991 à 2023, a été modélisé.

Les projections des taux de mortalité pour la période 2024-2050, obtenues avec le modèle PROPHET, suivent les dynamiques de mortalité observées de 1991 à 2023.

- Les projections conservent les profils caractéristiques de la mortalité par âge : une mortalité infantile initialement élevée suivie d'une diminution rapide, une "bosse des accidents", puis une augmentation exponentielle des taux avec l'âge.
- En termes d'évolution temporelle, les projections pour 2024-2050 révèlent une amélioration continue, marquée par une diminution progressive des taux de mortalité, cohérente avec les tendances observées entre 1991 et 2023.

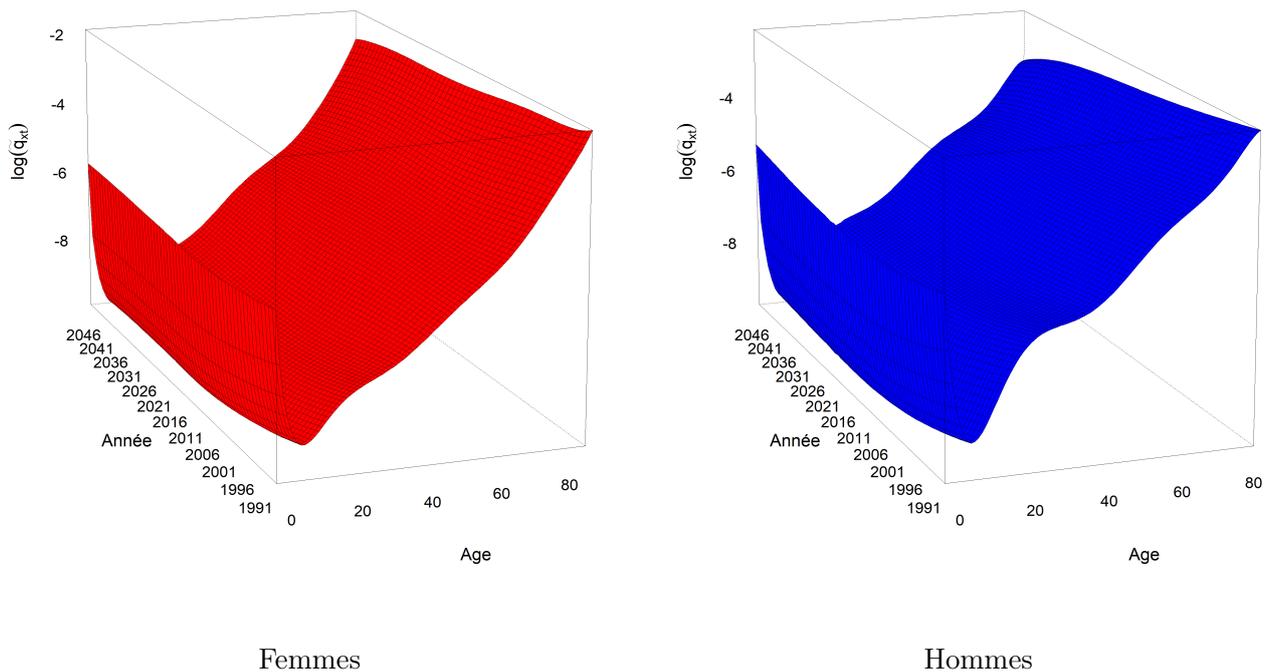


FIGURE 2.1 – Le logarithme des taux de mortalité projetés avec le modèle PROPHET



Une idée d'exploration : Dans cette étude, la composante de cohorte γ_{t-x} a été modélisée à l'aide d'un modèle ARIMA adapté pour capturer les tendances linéaires dans le temps. Toutefois, il serait intéressant d'envisager d'autres méthodes de modélisation. Par exemple, des modèles prenant en compte la saisonnalité, tels que SARIMA, ou des modèles intégrant des régresseurs, comme PROPHET ou SARIMAX, pourraient apporter des perspectives nouvelles. Ces modèles sont particulièrement appropriés pour traiter des cohortes ayant des dynamiques de mortalité spécifiques influencées par des événements exceptionnels tels que les guerres.

Comparaison des tendances d'évolutions temporelles :

Une comparaison des tables de mortalité obtenues avec le modèle classique de projection, basé sur un ARIMA, et un modèle alternatif, PROPHET, a été menée pour évaluer ces différences.

Les tendances observées dans les projections des taux de mortalité issues des deux modèles montrent à la fois des similarités et des divergences importantes :

- **Le modèle ARIMA :** Les tables obtenues avec le modèle ARIMA, fondées sur une prolongation des tendances historiques linéaires, montrent une baisse continue et régulière des taux de mortalité.
- **Le modèle PROPHET :** Le modèle PROPHET, en intégrant des points de changement, offre une plus grande réactivité aux dynamiques récentes. Les projections avec ce modèle montrent également une diminution des taux de mortalité, mais elles sont plus sensibles aux variations récentes, suggérant un possible ralentissement dans l'amélioration de la mortalité. Cette caractéristique est particulièrement pertinente si les perturbations des dernières années reflètent un changement structurel dans les dynamiques de mortalité.

Les figures ci-dessous illustrent ces différences pour les taux de mortalité à 30 ans chez les femmes et les hommes.

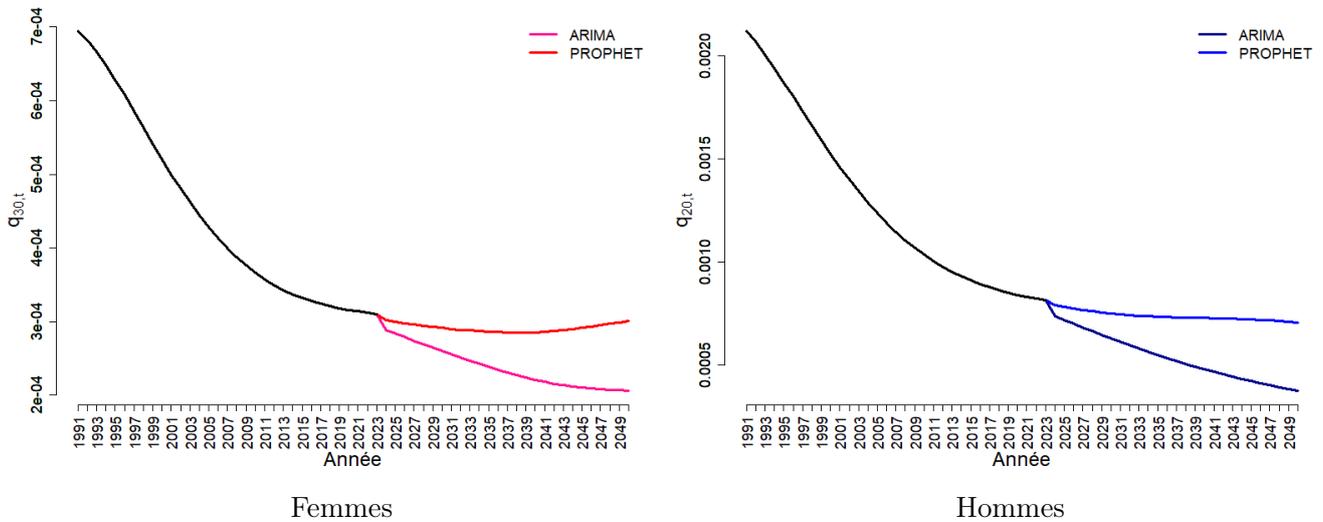


FIGURE 2.2 – Les taux de mortalité projetés à 30 ans avec le modèle classique et le modèle PROPHET

Une analyse quantitative révèle que les taux de mortalité projetés par le modèle PROPHET sont, en moyenne, supérieurs de 5,22% pour les femmes et de 8% pour les hommes sur 2024 par rapport aux projections du modèle ARIMA.

Cela suggère que le modèle PROPHET, en intégrant mieux les récents changements de tendance, anticipe un ralentissement plus marqué des améliorations de la mortalité.

2.2 La fermeture de la table de mortalité

La fermeture des tables de mortalité est essentielle pour obtenir des estimations fiables aux âges élevés, où les données empiriques sont souvent limitées. Dans cette étude, la méthode de DENUIT & GODERNIAUX est utilisée pour prolonger les taux de mortalité au-delà des âges pour lesquels les données sont disponibles.

Pour rappel, cette méthode permet de lisser les taux de mortalité au-delà d'un certain âge raccord, en assurant une transition cohérente entre les observations empiriques et les extrapolations. Ces âges de raccord sont fixés à 70 ans pour les hommes et 86 ans pour les femmes afin de garantir une continuité dans la courbe de mortalité.

Les surfaces de mortalité obtenues après application de cette méthode sont présentées ci-dessous. Elles montrent une continuité sans cassure visible aux âges de raccord, ce qui confirme la pertinence de l'utilisation de cette méthode pour la fermeture des tables de mortalité.

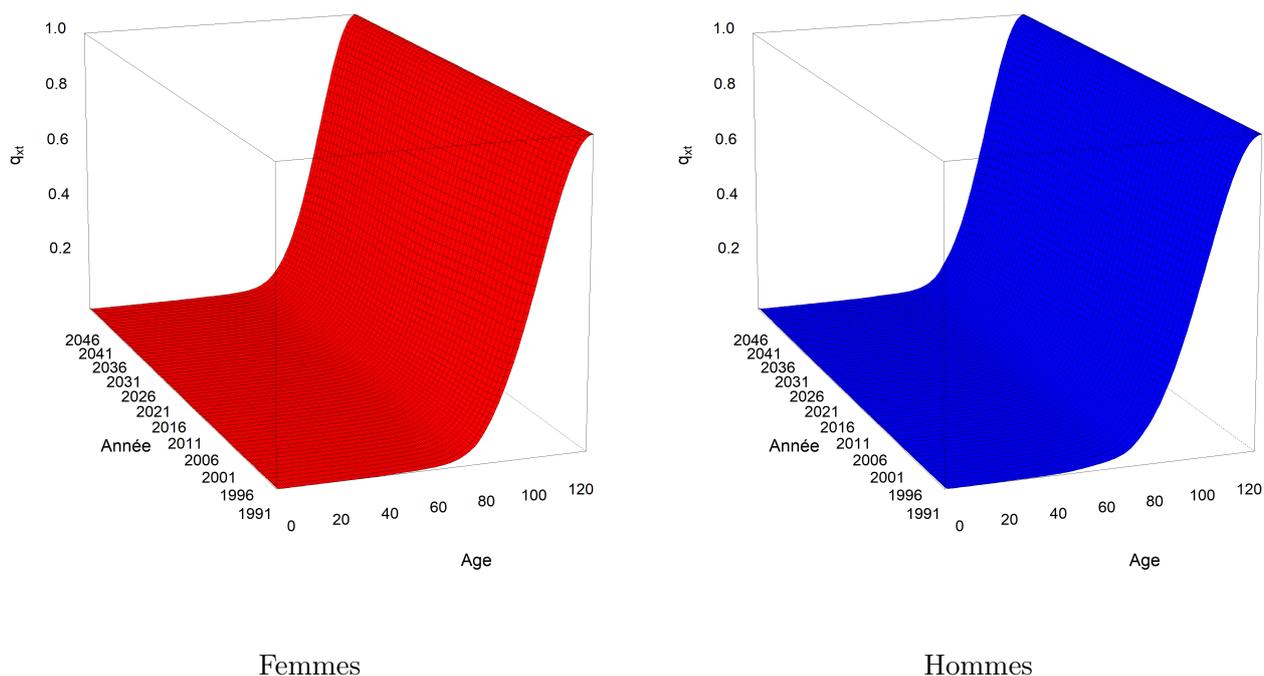


FIGURE 2.3 – Les taux de mortalité fermés par DENUIT & GODERNIAUX (PROPHET)

Ces travaux aboutissent à l'élaboration de nouvelles tables de mortalité prospectives pour les femmes et pour les hommes, couvrant la période de 1991 à 2050 et s'étendant sur les âges de 0 à 130 ans. Ces tables se distinguent par leur caractère innovant, résultant de l'utilisation d'un modèle de projection alternatif, PROPHET, plutôt que du modèle classique ARIMA. Cette approche permet de mieux capter les dynamiques récentes de mortalité, si elle s'évèrent être structurelles, tout en offrant des estimations robustes.



Ce qu'il faut retenir du chapitre :

- * Les nouvelles tables de mortalité ont été élaborées à l'aide du modèle PROPHET, capable d'intégrer des points de changement. Ce choix permet de mieux capter les récentes dynamiques de mortalité marquées par des perturbations, notamment si elles sont structurelles. Les projections réalisées avec ce modèle présentent des taux de mortalité supérieurs à ceux obtenus avec le modèle classique. Les différences observées suggèrent que PROPHET est plus sensible aux changements récents dans les tendances de mortalité, anticipant un ralentissement plus marqué des améliorations futures.
- * La méthode de DENUIT & GODERNIAUX a été utilisée pour prolonger les tables aux âges élevés. Cette approche assure une continuité sans cassure visible aux âges de raccord, garantissant des estimations fiables jusqu'à 130 ans.

Dans la partie suivante, les différentes tables de mortalité élaborées dans ce mémoire seront utilisées pour le calcul des provisions d'un régime de retraite supplémentaire. Cette application permettra d'illustrer l'impact d'un changement de table de mortalité sur les engagements financiers.

Schématisation synthétique de la partie :



1. ESTIMATIONS

Trois méthodes de construction de tables générationnelles ont été utilisées : LEE & CARTER, LOG POISSON, RENSHAW & HABERMAN.



2. VALIDATION

La proximité des taux de mortalité prédits et observés a été examinée.



Le modèle de RENSHAW & HABERMAN est retenu.



3. PROJECTIONS

Deux méthodes alternatives ont été explorées pour la projection de la composante temporelle : ETS et PROPHET.



Le modèle PROPHET est retenu.



4. FERMETURE

La méthode de DENUIT & GODERNIAUX a été utilisée pour extrapoler les taux de mortalité aux grands âges.



Des tables de mortalité par sexe, couvrant la période de 1991 à 2050 et les âges de 0 à 130 ans, ont été construites :

- **Les données de 1991 à 2023 ont été utilisées pour projeter les taux de mortalité de 2024 à 2050 avec le modèle de PROPHET.**

FIGURE 2.4 – Le schéma synthétique de la partie

Quatrième partie

Impact quantitatif sur le provisionnement d'un portefeuille de retraite supplémentaire

Chapitre 1

Contextualisation

Ce chapitre vise à fournir le contexte essentiel dans lequel s'inscrit le régime de retraite supplémentaire, pour lequel les Provisions Mathématiques Théoriques (PMT) seront calculées dans le chapitre suivant.

1.1 La retraite supplémentaire

La retraite supplémentaire, également appelée retraite surcomplémentaire, est un régime facultatif qui vient en complément de la retraite de base et de la retraite complémentaire. Elle repose sur le principe de la capitalisation : l'épargne constituée est investie sur des supports financiers, et les rendements générés sont capitalisés jusqu'au départ à la retraite. Ce régime peut être souscrit auprès d'institutions de prévoyance, de sociétés d'assurance ou d'établissements bancaires, qui collectent et investissent les cotisations pour constituer les réserves financières nécessaires au paiement des prestations futures.

La retraite supplémentaire peut prendre deux formes principales : la retraite supplémentaire collective proposée par les entreprises et souvent alimentée par des versements de l'employeur en complément des cotisations du salarié et la retraite supplémentaire individuelle qui permet aux individus de constituer leur propre épargne retraite via des versements volontaires. À l'âge de la retraite, les organismes assureurs versent les prestations conformément aux termes des contrats souscrits, garantissant ainsi un complément de revenus pour les bénéficiaires.

Ces régimes peuvent s'articuler autour de deux grandes typologies :

- **Les régimes à prestations définies (Branches 20 & 22) :** le montant de la pension est défini contractuellement et les cotisations du régime sont ajustées régulièrement. Dans ce type de régime, le risque lié à l'investissement et à la gestion du fonds de pension repose principalement sur l'organisme responsable du régime. L'objectif est de garantir aux participants une certaine sécurité financière à la retraite, en leur assurant un niveau de prestations prédéterminé.
- **Les régimes à cotisations définies (Branches 20 & 22 et Branche 26) :** le niveau des cotisations est défini contractuellement. Le niveau des prestations quant à lui fluctue en fonction du contexte économique, démographique et financier. Dans les régimes à cotisations définies, le montant de la pension de retraite n'est pas prédéterminé. Les participants assument donc le risque lié aux performances des investissements et à la gestion de leur compte individuel.

Bien que connaissant une augmentation des cotisations versées et des prestations reçues, la retraite supplémentaire reste encore marginale en comparaison avec l'ensemble des régimes de retraite. Elle ne représente que seulement 5,8% de l'ensemble des contributions de la retraite en 2023, ce qui souligne sa part relativement faible dans le paysage global de la retraite. De même, les prestations versées au titre de la retraite supplémentaire ne représentent que 2,3% des prestations totales de retraite en 2023.

1.2 Les régimes en points

Les régimes de retraite en points constituent une approche spécifique dans le cadre des dispositifs de retraite supplémentaire. Ils reposent sur un mécanisme d'accumulation et de conversion des droits en une rente viagère, permettant une gestion claire et lisible des cotisations et des prestations :

- **Acquisition des points** : Le principe d'acquisition des points repose sur la transformation des montants cotisés par l'employeur et/ou le salarié en points de retraite. Cette conversion s'effectue en fonction d'une valeur d'achat définie par l'organisme gestionnaire. Par exemple, si un salarié cotise 200 € dans un régime où la valeur d'achat d'un point est de 10 €, il acquiert 20 points pour cette période.
- **La conversion en rente** : À l'âge de la retraite, le total des points accumulés est utilisé pour calculer la rente viagère. La conversion s'effectue en fonction du nombre de points accumulés et de la valeur de service du point à cette date, définie par l'organisme gestionnaire au moment de la liquidation des droits.

$$\text{Rente} = \text{Nombre de points} \times \text{Valeur de service du point}$$

Par exemple, un retraité ayant accumulé 20 000 points et une valeur de service de 1,5 € par point percevra une rente annuelle de 30 000 €.

1.3 Le régime étudié

Le régime de retraite supplémentaire de l'étude est un dispositif collectif à cotisations définies, fonctionnant en points. Il est ainsi géré par capitalisation. Les assurés acquièrent des points en cotisant et cette épargne est investie sur des supports financiers. Les droits acquis ouvrent droit au versement d'une rente viagère annuelle réversible.

Le portefeuille d'assurés est constitué de 102 913 individus, répartis en trois catégories :

- **Les assurés actifs ou radiés**, comprenant 87 130 individus, soit près de 85% du total.
- **Les bénéficiaires de droits directs**, représentant 12 501 individus, soit environ 12% du total.
- **Les bénéficiaires de droits indirects** correspondant à 3 282 individus, soit environ 3% du total.

1.3.1 Les assurés actifs ou radiés

Les assurés actifs sont les participants au régime qui continuent d'acquérir des points de retraite en fonction de leurs cotisations. Ils n'ont ainsi pas encore liquidé leurs droits à la retraite supplémentaire.

Les assurés radiés sont ceux qui ont cessé de cotiser, mais qui n'ont pas encore liquidé leurs droits à la retraite supplémentaire.

L'analyse du portefeuille des assurés actifs ou radiés révèle une nette majorité d'hommes, représentant 75% des assurés, contre 25% de femmes. Tant chez les hommes que chez les femmes, ces assurés sont principalement âgés de 50 à 70 ans, avec un âge moyen de 54 ans pour les hommes et de 52 ans pour les femmes, comme l'illustre le graphique ci-dessous.

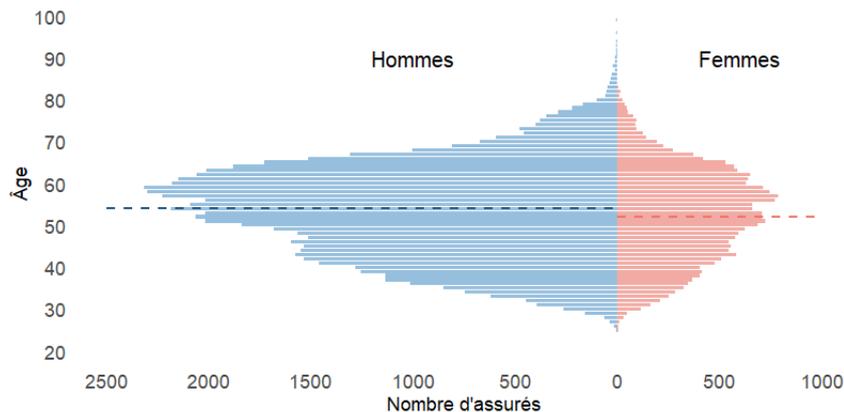


FIGURE 1.1 – La pyramide des âges des assurés actifs ou radiés

L'analyse des montants de rentes théoriques, obtenus en convertissant le nombre de points cumulés par assurés jusqu'à l'exercice N-1, met en évidence une répartition majoritairement concentrée sur des montants très faibles. En effet, plus de 80% des rentes sont inférieures à 500 euros par an et près de 46,5% sont inférieures à 100 euros. Le montant moyen des rentes est égal à 357 euros. Cela s'explique notamment par le fait que les participants n'ont pas encore liquidé leurs droits et que la majorité d'entre eux continue de cotiser (environ 80%), ce qui entraîne des montants théoriques faibles à ce stade.

1.3.2 Les bénéficiaires de droits directs

Les bénéficiaires de droits directs sont les participants qui ont liquidé leurs droits à la retraite supplémentaire et qui reçoivent une rente viagère, établie en fonction de leurs cotisations passées et du nombre de points acquis.

L'analyse du portefeuille des bénéficiaires de droits directs révèle une nette majorité d'hommes, représentant 84% des bénéficiaires, contre 16% de femmes. Tant chez les hommes que chez les femmes, ces bénéficiaires sont principalement âgés de 70 à 80 ans, avec un âge moyen de 80 ans pour les hommes et 79 ans pour les femmes, comme l'illustre le graphique ci-dessous.

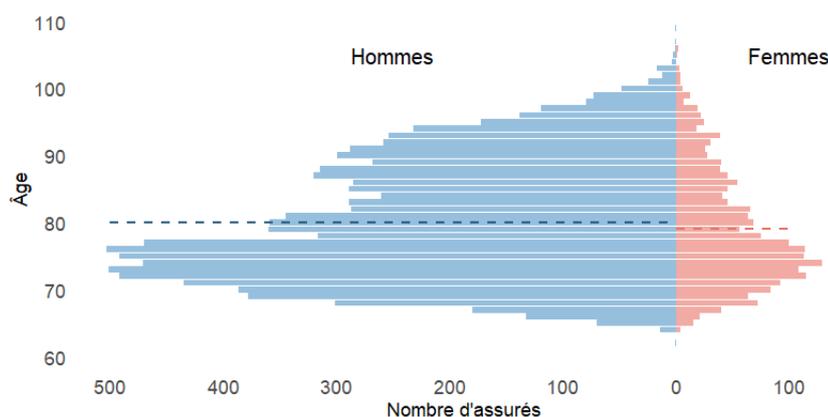


FIGURE 1.2 – La pyramide des âges des bénéficiaires de droits directs

L'analyse des rentes versées révèle une prédominance de rentes annuelles inférieures à 1 000 euros, représentant environ 70% du total avec un montant moyen de 1 085 euros.

1.3.3 Les bénéficiaires de droits indirects

Les **bénéficiaires de droits indirects** sont les ayants droit, principalement les conjoints survivants, qui perçoivent une rente réversible après le décès du participant principal. Cette rente n'est pas liée à leurs propres cotisations, mais provient des droits acquis par le participant décédé.

L'analyse du portefeuille des bénéficiaires de droits indirects révèle une nette majorité de femmes, représentant 98% des bénéficiaires, contre 2% d'hommes. Ces femmes sont principalement âgées de 80 à 100 ans, avec un âge moyen de 87 ans, comme l'illustre le graphique ci-dessous.

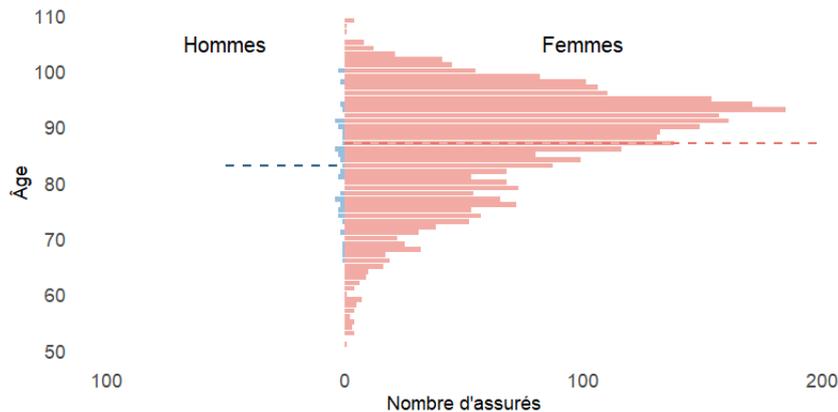


FIGURE 1.3 – La pyramide des âges des bénéficiaires de droits indirects

L'analyse des rentes versées montre une concentration notable de rentes annuelles inférieures à 500 euros, représentant environ 72% du total. Le montant moyen des rentes s'établit également autour de 500 euros



Ce qu'il faut retenir du chapitre :

Le régime étudié est un régime de retraite supplémentaire qui repose sur le principe des régimes en points. Il s'agit d'un dispositif collectif à cotisations définies où les droits des assurés sont matérialisés par des points acquis au fil des cotisations. Ces points, calculés en fonction des montants versés, seront convertis en une rente viagère réversible à l'âge de la retraite. Le chapitre suivant présente les calculs détaillés des Provisions Mathématiques Théoriques (PMT) associées à ce régime.

Chapitre 2

Le calcul des provisions

Ce chapitre offre l'opportunité d'examiner l'impact de l'utilisation des tables de mortalité élaborées dans ce mémoire sur le calcul des Provisions Mathématiques Théoriques (PMT). À travers un exemple chiffré appliqué à un portefeuille de retraite supplémentaire, l'objectif est non seulement de comparer les résultats issus de ces tables à ceux des tables réglementaires, mais également de confronter les résultats obtenus avec les différentes tables construites. Cela permettra de quantifier l'impact de l'intégration de données récentes, marquées par des perturbations, ainsi que d'évaluer l'influence des méthodes de projection utilisées.

2.1 Les Provisions Mathématiques Théoriques (PMT)

Les Provisions Mathématiques Théoriques (PMT) représentent les engagements financiers d'un régime de retraite, calculés pour garantir le paiement des rentes viagères. Elles correspondent à la somme des valeurs actuelles des prestations futures dues aux assurés, actualisées en fonction d'hypothèses démographiques et financières. En d'autres termes, elles traduisent la réserve nécessaire pour couvrir les droits acquis par les participants au régime.

La PMT est calculée à l'aide d'un coefficient actuariel, qui prend en compte les caractéristiques individuelles de chaque assuré, ainsi que des hypothèses prédéfinies, telles que les tables de mortalité. Ainsi, pour un assuré, la PMT s'exprime par la formule suivante :

$$\text{PMT} = \text{Rente} \times \text{Coefficient actuariel}$$

2.2 Les hypothèses de calcul

Les hypothèses retenues pour les calculs de PMT sont les suivantes :

- **Date de calcul** : Les calculs sont effectués au **31/12/2023**.
- **Âge de départ à la retraite** : L'âge de départ à la retraite est fixé à **64 ans**, comme simplification de la réforme des retraites de 2023.
- **Âge du conjoint** : Les hommes sont considérés comme ayant **2 ans de plus** que leurs épouses.
- **Taux de mariage** :

$$\text{Taux de mariage} = \begin{cases} 84\% & \text{si l'âge du participant est inférieur à 64 ans,} \\ 84\% - 1,7\% \times (\text{âge} - 64) & \text{si l'âge du participant dépasse 64 ans.} \end{cases}$$

- **Âge minimum de l’ayant droit pour percevoir la rente de réversion** : L’ayant droit doit être âgé au minimum de **55 ans** pour être éligible à la rente de réversion.
- **Taux de réversion** : La rente de réversion est fixée à **60%** des droits du participant principal.
- **Tables de mortalité** : Les PMT ont été calculées selon des **scénarios distincts**, chacun basé sur des tables de mortalité différentes :
 - **Scénario 1 (ARIMA 2019)** : Utilisation des tables construites à partir des données historiques, sans inclure les perturbations récentes, et avec un modèle de projection classique.
 - **Scénario 2 (ARIMA 2023)** : Utilisation des tables intégrant à la fois les données historiques et les perturbations récentes, tout en appliquant un modèle de projection classique.
 - **Scénario 3 (PROPHET 2023)** : Utilisation des tables prenant en compte les données historiques et les perturbations récentes, mais basées sur un modèle de projection alternatif.

Ces tables de mortalité construites auraient pu servir de références pour positionner le portefeuille en fonction de ses données spécifiques. En ajustant les tables aux caractéristiques démographiques précises du groupe étudié, des résultats plus représentatifs de la réalité de ce portefeuille auraient été obtenus. Deux méthodes de positionnement couramment utilisées en actuariat auraient pu être envisagées : l’application d’un coefficient de réduction ou de majoration de manière multiplicative et l’utilisation d’un modèle relationnel. Ces méthodes sont présentées en annexe. Toutefois, cette approche n’a pas été retenue en raison du manque de données nécessaires à sa mise en œuvre.

- **Taux technique** : Bien qu’il soit habituel d’utiliser le **taux EIOPA** au 31/12/2023 comme base d’actualisation, ce taux est fixé à 0 pour cette étude. Ce choix vise à neutraliser l’effet des taux et à isoler l’impact des tables de mortalité sur les résultats.

Ces hypothèses garantissent une prise en compte rigoureuse et cohérente des facteurs influençant les engagements du régime.

2.3 Les méthodologies de calculs

2.3.1 Les assurés actifs ou radiés

Dans le cadre des assurés actifs ou radiés, le coefficient actuariel tient compte, d’une part, de la probabilité de survie de l’assuré après l’âge de liquidation de la retraite. D’autre part, après le décès de l’assuré, le coefficient prend en compte la probabilité de réversion, c’est-à-dire la probabilité qu’un ayant droit, généralement un conjoint, continue de percevoir une rente réversible. Ce calcul inclut la durée pendant laquelle cette rente sera versée à l’ayant droit et à quel montant, en fonction du taux de réversion et des probabilités de survie du conjoint. En résumé, le coefficient actuariel permet d’actualiser l’ensemble des engagements futurs du régime de retraite, qu’il s’agisse des rentes directes ou réversibles.

$$\text{Coefficient actuariel} = \sum_{k=63-x} k p_x - \frac{11}{24} \times {}_{63-x} p_x + \text{taux}_{rev} \times \text{taux}_{mariage} \times \sum_{k=0} ({}_k q_x \times {}_k p_y \mathbb{1}_{y+k>55})$$

où :

- ${}_k p_x$: La probabilité de survie du participant entre x et $x+k$ (${}_k q_x$, la probabilité de décès entre x et $x+k$).
- taux_{rev} : Le taux de réversion.
- $\text{taux}_{mariage}$: Le taux de mariage.
- ${}_k p_y$: La probabilité de survie de l’ayant-droit entre y et $y+k$.

Exemple :

Un homme né le 24/09/1968 est âgé de 55 ans au 31/12/2023. Il est ainsi supposé que sa femme est âgée de 53 ans au 31/12/2023. Le taux de mariage utilisé est égal à 84%.

Jusqu'à l'exercice N-1, ce participant a acquis 425 points et la valeur de service du point au 31/12/2023 est de 0,3439 euros. Ainsi, sa rente annuelle est égale à 146,16 euros.

$$\text{Rente annuelle} = 425 \times 0,3439 \text{ €} = 146,16 \text{ €}$$

Les étapes de calcul de la PMT sont les suivantes :

— **Calcul de la probabilité de survie du participant à la liquidation :**

La probabilité que le participant soit vivant à l'âge de la liquidation, c'est-à-dire dans 9 ans, est calculée.

$${}_9p_{55} = \frac{l_{55+9}}{l_{55}}$$

— **Calcul des probabilités de survie et de décès à chaque période :**

- Pour chaque période, k , les probabilités de décès et de survie du participant entre 55 et $55 + k$ sont calculées.

$${}_kp_{55} = \frac{l_{55+k}}{l_{55}}$$

Ainsi, que les probabilités de décès associées :

$${}_kq_{55} = 1 - {}_kp_{55}$$

- Les mêmes calculs sont effectués pour l'ayant droit afin de déterminer la durée potentielle pendant laquelle une rente de réversion pourrait être versée.

Pour illustrer ces calculs, le graphique ci-dessous présente l'évolution des probabilités de survie du participant en fonction de son âge, selon les différents scénarios de mortalité.

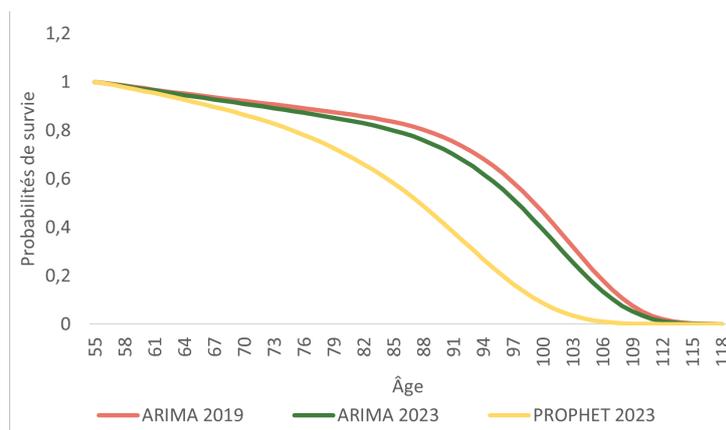


FIGURE 2.1 – Les probabilités de survie du participant (actifs-radiés)

Le scénario ARIMA 2023, qui intègre les données des dernières années marquées par des perturbations, présente des probabilités de survie légèrement plus faibles que celles du scénario ARIMA 2019 pour les âges avancés. Quant au scénario PROPHET 2023, il affiche des probabilités de survie globalement plus faibles que tous les autres scénarios.

— **Calcul des probabilités de versement d’une rente directe :**

Pour chaque période, la probabilité de versement d’une rente directe est ensuite calculée.

- Avant que le participant n’atteigne l’âge de liquidation, la probabilité de versement de la rente directe est nulle.
- La première année de versement intervient l’année où le participant atteint l’âge de liquidation ($k = 9$). Cette probabilité est ainsi égale à :

$${}_9p_{55} - \frac{11}{24} \times {}_9p_{55}$$

- À partir des périodes suivantes, seule la probabilité de survie du participant entre 55 et $55 + k$, ${}_kp_{55}$ est prise en compte.

— **Calcul des probabilités de versement d’une rente de réversion :**

Pour chaque période, la probabilité de versement d’une rente de réversion est ensuite calculée.

- Avant que l’ayant droit n’atteigne l’âge ouvrant les droits à une rente de réversion (55 ans), la probabilité de versement de la rente de réversion est nulle.
- À partir des périodes suivantes, la probabilité de verser une rente de réversion tient compte de la probabilité de décès de l’assuré, de la probabilité de survie de l’ayant droit et des taux de réversion et de mariage.

$$taux_{rev} \times taux_{mariage} \times {}_kq_{55} \times {}_kp_{53}$$

- **Actualisation des probabilités de versement :** Une actualisation de ces probabilités est effectuée pour chaque période k , à l’aide du coefficient d’actualisation suivant :

$$\frac{1}{(1+r)^k}$$

où r est le taux d’actualisation. Pour rappel, dans le cadre de cet exemple, ce taux d’actualisation est fixé à 0 pour neutraliser son effet sur le calcul des PMT.

— **Obtention du coefficient actuariel :**

En sommant ces probabilités actualisées, le coefficient actuariel est obtenu. Il représente le nombre d’années pendant lesquelles l’organisme assureur prévoit de verser une rente annuelle à ce participant. Les résultats ci-dessous sont obtenus selon les différents scénarios de mortalité. Entre le scénario ARIMA 2019, traduisant une amélioration constante de la mortalité, et le scénario PROPHET 2023, traduisant un ralentissement marqué de l’amélioration de la mortalité, un écart proche de 9 années de versement probable est observé.

	ARIMA 2019	ARIMA 2023	PROPHET 2023
Coefficient	35,4	34	26,5

TABLE 2.1 – Exemple actif-radié : Les coefficients actuariels

— **Obtention de la Provision Mathématique Théorique :**

La PMT, associée à ce participant, est ensuite calculée en multipliant la rente annuelle par le coefficient actuariel. L'écart observé sur ces coefficients entraîne un écart de 1 301 euros sur le montant de provision pour ce participant entre les scénarios ARIMA 2019 et PROPHET 2023.

	ARIMA 2019	ARIMA 2023	PROPHET 2023
PMT	5 174 €	4 969 €	3 873 €

TABLE 2.2 – Exemple actif-radié : Les montants de PMT

2.3.2 Les bénéficiaires de droits directs

Dans le cas des bénéficiaires de droits directs, le coefficient actuariel tient compte, d'une part, de la probabilité que l'assuré principal reste en vie, déterminant ainsi combien de temps la rente directe sera versée. D'autre part, après le décès de l'assuré, le coefficient prend en compte la probabilité de réversion, c'est-à-dire la probabilité qu'un ayant droit, généralement un conjoint, continue de percevoir une rente réversible. Ce calcul inclut la durée pendant laquelle cette rente sera versée à l'ayant droit et à quel montant, en fonction du taux de réversion et des probabilités de survie du conjoint. En résumé, le coefficient actuariel permet d'actualiser l'ensemble des engagements futurs du régime de retraite, qu'il s'agisse des rentes directes ou réversibles.

$$\text{Coefficient actuariel} = \sum_{k=0} ({}_k p_x + {}_k q_x \times \text{taux}_{rev} \times \text{taux}_{mariage} \times {}_k p_y) - \frac{11}{24}$$

où :

- ${}_k p_x$: La probabilité de survie du participant (${}_k q_x$, la probabilité de décès).
- taux_{rev} : Le taux de réversion.
- $\text{taux}_{mariage}$: Le taux de mariage.
- ${}_k p_y$: La probabilité de survie de l'ayant droit.

2.3.3 Les bénéficiaires de droits indirects

Dans le cas de bénéficiaires de droits indirects, le coefficient actuariel tient compte de la probabilité de survie du réversataire. Ce calcul reflète la durée pendant laquelle la rente de réversion sera versée à l'ayant droit réversataire.

$$\text{Coefficient actuariel} = \sum_{k=0} {}_k p_y - \frac{11}{24}$$

où ${}_k p_y$ est la probabilité de survie du réversataire.

2.4 Les montants de PMT

L'analyse met en évidence une sensibilité aux données utilisées, soulignant leurs impacts directs sur les résultats. L'intégration de données affectées par des perturbations modifie la dynamique de mortalité, comme l'illustre le scénario ARIMA 2023. Ce dernier traduit une mortalité plus élevée que celle observée dans le scénario ARIMA 2019, impliquant que les individus décèdent plus tôt. Cette évolution entraîne une diminution des engagements futurs de l'assureur, entraînant ainsi des montants de PMT inférieurs avec le scénario ARIMA 2023 par rapport à ceux obtenus avec le scénario ARIMA 2019.

- Pour les **participants actifs ou radiés**, les PMT sont inférieures de 7,7% avec le scénario ARIMA 2023 (1 0046M€ versus 965M€).
- Pour les **bénéficiaires de droits directs**, les PMT sont inférieures de 5,1% avec le scénario ARIMA 2023 (234M€ versus 222M€).
- Pour les **bénéficiaires de droits indirects**, les PMT sont inférieures de 3,4% avec le scénario ARIMA 2023 (17,5M€ versus 16,9M€).

L'utilisation d'une méthode de projection alternative, illustrée ici par le scénario PROPHET 2023, met en évidence une baisse encore plus marquée de la PMT. Ce scénario repose sur une approche qui prend en compte des ruptures de tendance, traduisant un possible ralentissement de l'amélioration de la mortalité à long terme. Une telle hypothèse révèle une forte sensibilité des résultats au modèle de projection utilisé. En comparaison, le scénario ARIMA 2023, construit à partir du même historique de données mais avec une méthode de projection différente, conduit à des montants de PMT plus élevés. Cet écart reflète les divergences entre les approches méthodologiques.

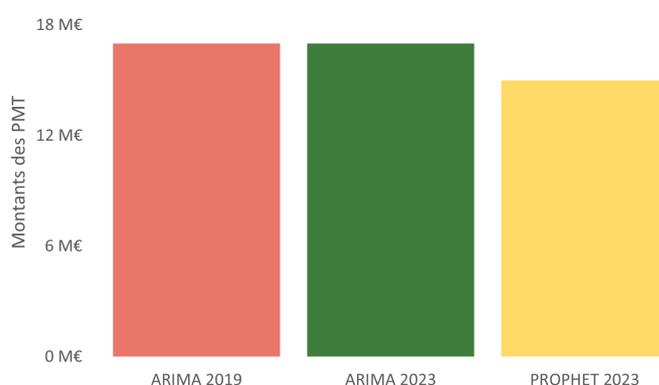
- Pour les **participants actifs ou radiés**, les PMT sont inférieures de 20% avec le scénario PROPHET 2023 (965M€ versus 769M€)
- Pour les **bénéficiaires de droits directs**, les PMT sont inférieures de 16% avec le scénario PROPHET 2023 (222M€ versus 186M€).
- Pour les **bénéficiaires de droits indirects**, les PMT sont inférieures de 10,7% avec le scénario PROPHET 2023 (16,9M€ versus 15,1M€).

Cette étude met en lumière l'incertitude liée à la projection de la mortalité. À ce stade, il est difficile de déterminer si le scénario PROPHET 2023 reflète une tendance pertinente ou s'il correspond à une hypothèse extrême. Un suivi rigoureux, incluant des *backtesting* réguliers, sera nécessaire dans les années à venir pour confronter ces hypothèses à la mortalité réelle observée au sein du portefeuille. Si la hausse de la mortalité observée ces dernières années devait s'avérer structurelle plutôt que conjoncturelle, ce scénario pourrait mériter une analyse plus approfondie. Toutefois, en l'état actuel, cette question reste ouverte et les résultats doivent être interprétés avec prudence.



Les assurés actifs ou radiés

Les bénéficiaires de droits directs



Les bénéficiaires de droits indirects

FIGURE 2.2 – Les sommes des PMT

Pour rappel, les PMT présentées jusqu'ici sont calculées sans actualisation. Cette absence amplifie le besoin de provisions en accordant un poids identique aux engagements de court et de long terme. L'application d'un taux d'actualisation permet de ramener les flux futurs à leur valeur actuelle, réduisant ainsi le besoin de provisions. Afin d'évaluer l'impact de l'actualisation sur les montants des provisions, les PMT du scénario ARIMA 2023, considéré comme un scénario central puisqu'il intègre les données récentes tout en conservant une méthode de projection standard, ont été recalculées pour les actifs radiés, qui représentent 85% du portefeuille total, en actualisant à différents niveaux : 1,75%, 1% et avec la courbe des taux EIOPA au 31/12/2023.

Les résultats obtenus sont présentés ci-dessous. Pour rappel, sans actualisation, les PMT s'élevaient à 965M€ :

	1,75%	1%	EIOPA 31/12/2023
Montant des PMT	636M€	754M€	536M€
Écart avec et sans actualisation	-34%	-21,9%	-44,5%

TABLE 2.3 – Les montants de PMT avec et sans actualisation

Ces résultats montrent que l'actualisation réduit significativement les montants des PMT, avec un effet plus marqué lorsque le taux d'actualisation est plus élevé. L'utilisation de la courbe des taux EIOPA au 31/12/2023 entraîne la réduction la plus importante, de -44,40%, illustrant l'impact majeur d'un taux du marché sur l'évaluation des engagements futurs.

Afin de mettre ces résultats en perspective avec les pratiques de marché, une comparaison a été effectuée entre les PMT obtenues avec le scénario ARIMA 2023 actualisé avec la courbe EIOPA et celles issues des tables réglementaires TGH-05/TGF-05, également actualisées avec la courbe EIOPA.

L'écart observé de -1,3% souligne une relative proximité avec les pratiques de marché (536M€ versus 529M€). Cette comparaison permet de valider la cohérence des projections du scénario ARIMA 2023 vis-à-vis des standards en vigueur, bien que des différences subsistent en raison des méthodologies utilisées.



Ce qu'il faut retenir du chapitre :

- * Une Provision Mathématique Théorique (PMT) représente la réserve financière nécessaire pour garantir le paiement des engagements futurs d'un régime de retraite.
- * Intégrer des données récentes marquées par des perturbations et/ou utiliser une méthode de projection alternative traduit une diminution de l'amélioration de la mortalité dans le temps. Cette réduction modifie les prévisions de durée pendant laquelle les rentes seront versées impactant directement le montant des PMT. Une durée de versement plus courte entraîne des PMT réduites, tandis qu'une durée prolongée les augmente.
- * L'application d'un taux d'actualisation modifie significativement les montants des PMT en ramenant les flux futurs à leur valeur actuelle.
- * Une comparaison avec les tables réglementaires a été réalisée à titre indicatif. Le faible écart observé est rassurant vis-à-vis du scénario ARIMA 2023.

Conclusion

Au cours des dernières décennies, les tendances de mortalité ont été marquées par une amélioration continue. Cependant, des perturbations récentes soulèvent la question de la nature des changements observés : s'agit-il d'évolutions structurelles ou de phénomènes conjoncturels temporaires ? Ce mémoire s'inscrit dans ce contexte d'incertitude, en proposant une démarche prospective pour construire des tables de mortalité de référence adaptées.

À partir des données de populations et de décès fournies par l'Institut National de la Statistique et des Études Économiques (INSEE), les taux de mortalité bruts ont été calculés par sexe. À l'aide de modèles prospectifs tels que RENSCHAW & HABERMAN, ces taux ont ensuite été décomposés pour isoler une composante temporelle.

En exploitant cette composante temporelle, des projections de mortalité ont été réalisées pour explorer différents scénarios et évaluer l'impact des perturbations récentes sur les dynamiques observées.

En utilisant une méthode de projection classique, basée sur un modèle ARIMA, deux approches ont été explorées pour évaluer l'influence des données récentes marquées par des perturbations. La première approche exclut ces années, prolongeant les tendances linéaires historiques observées. La seconde les intègre, entraînant un changement dans la trajectoire projetée et reflétant un éventuel ralentissement de l'amélioration de la mortalité.

En complément, l'utilisation d'une méthode de projection alternative, telle que le modèle PROPHET, capable d'intégrer des ruptures de tendance, reflète un ralentissement encore plus marqué de l'amélioration de la mortalité. Cette méthode offre une approche flexible qui pourrait se révéler particulièrement pertinente si les perturbations récentes traduisent des changements structurels dans les dynamiques de mortalité.

Ces analyses soulèvent des questions fondamentales quant à la validité et à l'applicabilité des différents scénarios de mortalité. Les résultats obtenus montrent clairement l'impact des perturbations récentes sur les projections et, par conséquent, sur les montants des Provisions Mathématiques Théoriques (PMT). Cependant, il reste difficile à ce stade de trancher sur la justesse des scénarios construits. Les incertitudes liées aux choix méthodologiques, aux hypothèses sous-jacentes et à la nature des données perturbées imposent une prudence dans l'interprétation de ces résultats.

Références

- [1] PLANCHET F. [2024] "Modèles de durée", Institut de Sciences Financières et d'Assurances (ISFA)
- [2] PLANCHET F. [2005], "Tables de mortalité d'expérience pour des portefeuilles de rentiers", Note méthodologique
- [3] BOUREAU C. [2014] "Prise en compte de l'hétérogénéité des individu dans le modèle de Lee-Carter", Diplôme Universitaire des Actuaire de Strasbourg (DUAS), Institut des Actuaire
- [4] 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", Euro-Institut d'Actuariat (EURIA), Institut des Actuaire
- [5] WAFO KANKEU C. B. [2023] "Impact du montant de rente sur la longévité au sein d'un portefeuille de rentiers : Une approche par les modèles additifs généralisés", Euro-Institut d'Actuariat (EURIA), Institut des Actuaire
- [6] GUO J. [2018] "Modélisation des dérivés de longévité avec mortalité stochastiques", HEC MONTRÉAL, Sciences de la gestion
- [7] VALLERAND P. [1939] "Origines et usages des tables de mortalité", Assurance, 7-2
- [8] BLANPAIN N., CHARDON O. [2010] "Projections de population 2007-2060 : amélioration de la mesure des quotients de mortalité au-delà de 90 ans et projections du nombre de centenaires", Direction des Statistiques Démographiques et Sociales
- [9] Institut National d'Études Démographiques (INED), Table de mortalité 2020-2022
- [10] Institut National de la Statistique et des Études Économiques (INSEE), Taux de mortalité en 2023 et nombre de décès en 2022, Comparaisons régionales et départementales
- [11] Institut National de la Statistique et des Études Économiques (INSEE) [2024] "Les écart d'espérance de vie entre cadres et ouvriers : 5 ans chez les hommes, 3 ans chez les femmes", INSEE Première, N.2005
- [12] Institut National de la Statistique et des Études Économiques (INSEE) [2018] "L'espérance de vie par niveau de vie : chez les hommes, 13 ans d'écart entre les plus aisés et les plus modestes", INSEE Première, N.1687
- [13] Institut National de la Statistique et des Études Économiques (INSEE), Recensement de la population - Estimations de la population, <https://www.insee.fr/fr/outil-interactif/5014911/pyramide.htm>
- [14] Institut National de la Statistique et des Études Économiques (INSEE), Fichiers des personnes personnes décédées depuis 1970, <https://www.insee.fr/fr/information/4190491>
- [15] FOUILLET A., GHOSN W., RIVERA C., CLANCHE F., COUDIN E. [2023], "Grandes causes de mortalité en France en 2021 et tendances récentes", Santé publique France, Centre d'épidémiologie sur les causes médicales de décès de l'Inserm, Direction de la recherche, des études ((Inserm-CépiDc), de l'évaluation et des statistiques (DRESS)

- [16] LEE R. D., CARTER L. R. [1992] "*Modeling and Forecasting US Mortality*", *Journal of the American Statistical Association*, Vol.87-419-659-671
- [17] CHAN W.S. LI J. S. H., LI J. [2014] "*The CDB Mortality Indexes : Modeling and Applications*", Orlando, Fla
- [18] QUASHIE A., DENUIT M. [2005] "Modèles d'extrapolation de la mortalité aux grands âges", Institut des Sciences Actuarielles & Institut de Statistique
- [19] GIROSI F., KING G. [2007] "*Understanding the Lee-Carter Mortality Forecasting Method*"
- [20] TAYLOR S. J., LETHAM B. [2017] "*Forescasting at Scale*" ,Facebook, Menlo Park, Californie
- [21] BIGOT J. [2017] "Séries chronologiques", Université de Bordeaux
- [22] CHARPENTIER A. "Introduction à la théorie des processus en temps discret", Modèles ARIMA et méthode Box & Jenkins
- [23] PLANCHET F., TOMAS J. "Construction et Validation des références de mortalité de place", Institut des Actuaire, Institut de Science Financière et d'Assurances (ISFA), Laboratoire de Sciences Actuarielles et Financières (SAF)
- [24] PLANCHET F., TOMAS J. "Critères de validation : aspects méthodologiques", Institut des Actuaire, Institut de Science Financière et d'Assurances (ISFA), Laboratoire de Sciences Actuarielles et Financières (SAF)
- [25] PLANCHET F., TOMAS J. "Méthodes de positionnement : aspect méthodologiques", Institut des Actuaire, Institut de Science Financière et d'Assurances (ISFA), Laboratoire de Sciences Actuarielles et Financières (SAF)
- [26] GUO Y., LI J. S. H. [2022] "*Robust Parameter Estimation for the Lee-Carter Family : A Probabilistic Principal Component Approach*", *Department of Statistics and Actuarial Science, University of Waterloo*
- [27] Arrêté du 22/11/2024 relatif à l'article L.911-1 du code de la sécurité sociale, Loi Industrie Verte
- [28] Cour de justice de l'Union européenne [2011] "La prise en compte du sexe de l'assuré en tant que facteur de risques dans les contrats d'assurance constitue une discrimination", Communiqué de presse 12/11

Annexes

A L'Article A132-18 du Code des assurances

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.

B La certification initiale

La certification des tables de mortalité, rendue obligatoire par l'Article A 132-18 du Code des assurances, assure que les tables construites reflètent fidèlement les dynamiques de mortalité des populations assurées. Cette étape est réalisée par un actuinaire agréé par l'Institut des Actuaires, qui doit justifier d'une expérience minimale de 5 ans en tarification et provisionnement dans le domaine de l'assurance de personnes et des engagements viagers. De plus, une solide formation en statistiques appliquées à la construction des tables de mortalité est requise. L'actuinaire doit maintenir une position d'indépendance vis-à-vis des organismes qui le sollicitent et établir un rapport de certification valable pour une période de 5 ans, garantissant que la table de mortalité permet la "constitution de provisions suffisantes et prudentes".

- **Évaluation des données** : Le processus commence par une analyse approfondie des données utilisées pour élaborer la table de mortalité. Cela comprend la vérification de l'exhaustivité, de la précision et de la représentativité des données, en examinant notamment les sources comme les registres de décès ou les bases de données internes de l'assureur, ainsi que les méthodes de collecte et de traitement des informations.
- **Validation et analyse des modèles statistiques** : Le certificateur évalue les techniques statistiques et les modèles utilisés pour estimer les taux de mortalité, s'assurant qu'ils capturent de manière fiable les dynamiques de mortalité observées. Cette validation inclut la calibration des modèles pour qu'ils soient adaptés au contexte spécifique du portefeuille d'assurance, en tenant compte des marges de prudence nécessaires pour gérer les incertitudes futures.
- **Conditions d'utilisation et recommandations** : Le rapport précise les conditions d'utilisation de la table de mortalité, en identifiant les contrats concernés et en proposant des recommandations pour l'amélioration continue de la table. Ces suggestions peuvent inclure l'affinement des méthodes de collecte de données, l'ajustement des hypothèses en fonction des nouvelles connaissances ou l'intégration de techniques statistiques plus avancées pour améliorer la précision des estimations.

Ce travail de validation rigoureux assure que les tables de mortalité sont adaptées au portefeuille spécifique et qu'elles permettent une gestion efficace des engagements de l'assureur, tout en anticipant les évolutions futures.

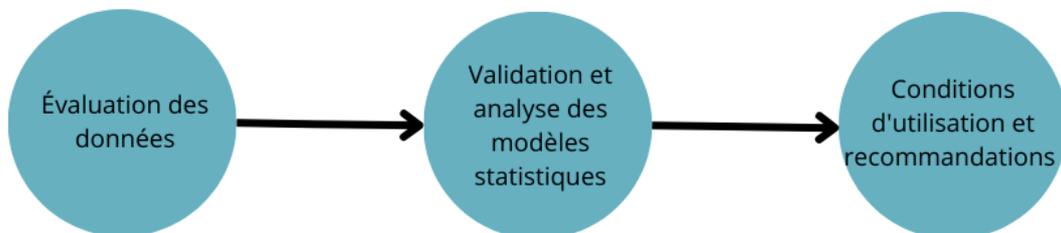


FIGURE A1 – Les tables d'expérience : la certification

C Le suivi annuel :

Après la certification initiale des tables de mortalité d'expérience, les organismes d'assurance doivent réaliser un suivi rigoureux pour s'assurer que ces tables restent alignées avec les caractéristiques évolutives de leur portefeuille. Ce suivi, effectué par un actuinaire agréé et mandaté, est essentiel pour garantir que les tables continuent de refléter fidèlement les dynamiques de mortalité au sein du portefeuille et pour anticiper toute évolution significative des risques. Chaque année, un rapport détaillé est produit à cet effet.

- **Suivi des éléments clés identifiés lors de la certification :** Lors de la certification initiale, plusieurs points sont souvent identifiés comme nécessitant une attention particulière lors du suivi. Ces points incluent les hypothèses fondamentales sur lesquelles reposent les tables, les méthodes de calcul employées, ainsi que la représentativité et la qualité des données utilisées. L'objectif est de s'assurer que ces éléments critiques demeurent valides et que les tables continuent d'offrir une estimation fiable des risques.
- **Vérification de l'adéquation des tables :** Une des principales responsabilités de l'actuaire dans le cadre du suivi annuel est de comparer les taux de mortalité observés avec ceux prédits par les tables de mortalité. Si des écarts significatifs sont détectés, l'actuaire doit évaluer s'ils résultent de fluctuations normales ou s'ils signalent un changement structurel dans la population assurée. Cette vérification est essentielle pour assurer que les tables de mortalité restent adaptées à la tarification des contrats et à la constitution des provisions.

Selon les résultats de cette vérification, l'actuaire peut recommander de maintenir les tables en l'état, de les ajuster, ou, dans certains cas, de les remplacer par de nouvelles tables plus adaptées. Le rapport de suivi annuel exprime clairement l'avis de l'actuaire sur le maintien ou l'abandon de la certification des tables de mortalité.

Enfin, en l'absence de suivi annuel, la validité des tables de mortalité expire deux ans après leur certification initiale, soulignant l'importance d'un suivi continu pour garantir leur pertinence.

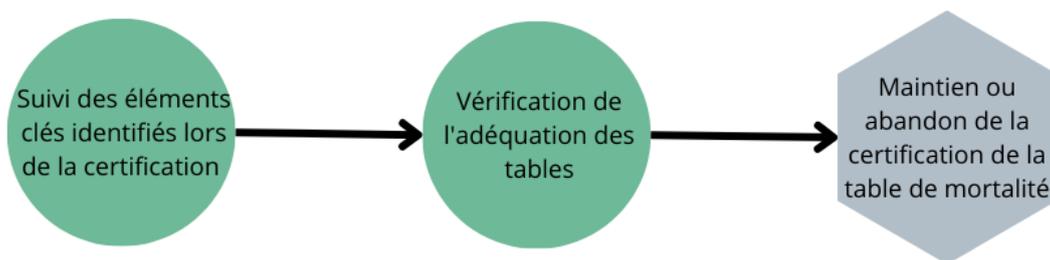


FIGURE A2 – Les tables d'expérience : le suivi

D Les pyramides des âges au 1^{er} janvier

Pour un âge donné, les excédents d'hommes ou de femmes sont représentés en couleur plus foncée, permettant de visualiser les différences de genre dans la distribution de la population.

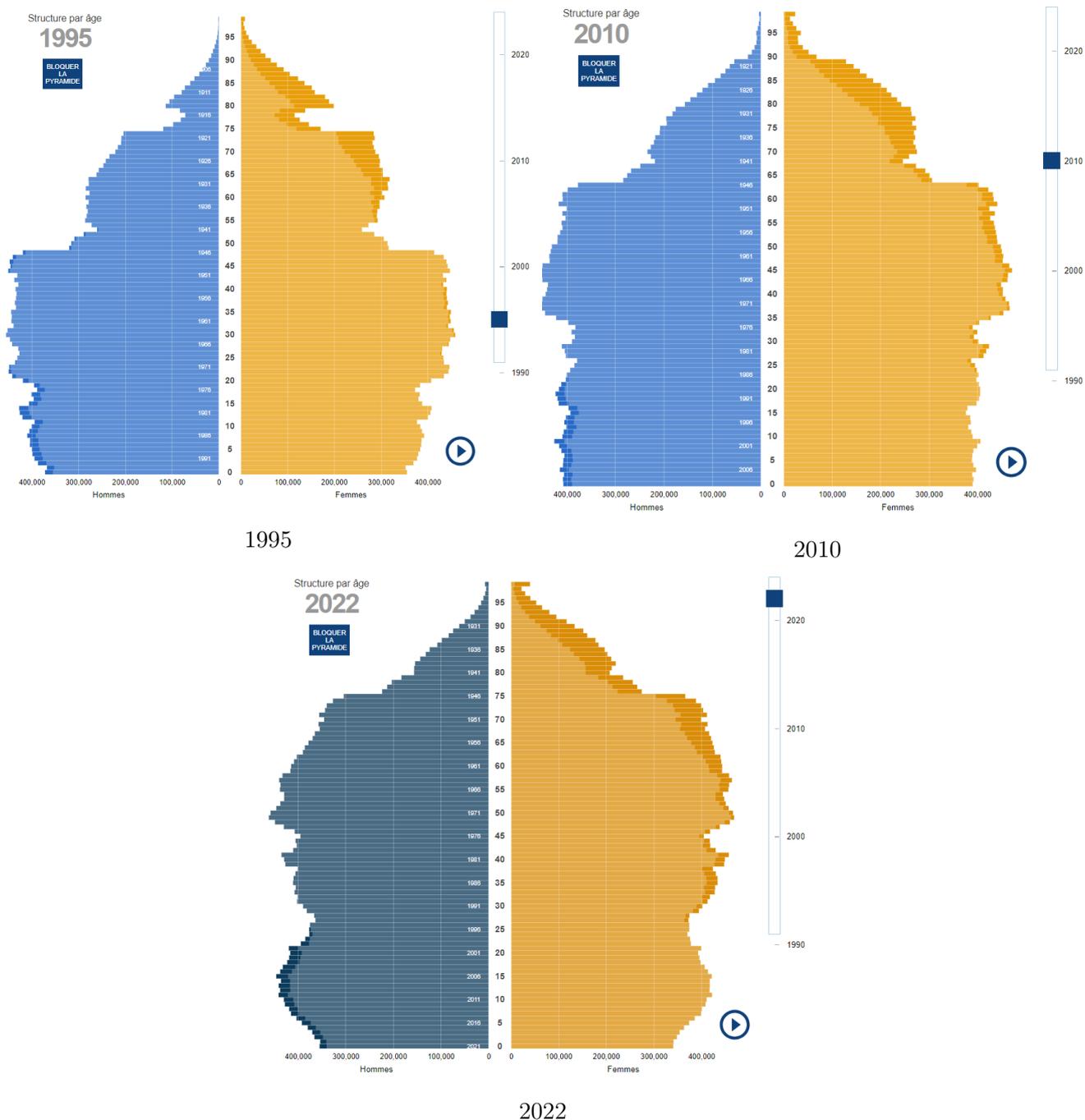


FIGURE A3 – Les pyramides des âges au 1^{er} janvier

Source : Institut National de la Statistique et des Études Économiques (INSEE)

E L'estimateur de HOEM

L'estimateur de HOEM est un estimateur du maximum de vraisemblance qui généralise l'estimateur binomial en tenant compte du temps d'observation spécifique à chaque individu au cours de la période d'étude. Pour formaliser cet estimateur, les notations suivantes sont introduites :

- $\alpha_{i,t}$ représente le début de l'observation pour l'individu i à l'âge x au cours de l'année t ;
- $\beta_{i,t}$ représente la fin de l'observation pour l'individu i à l'âge x au cours de l'année t .

Ainsi la probabilité que l'individu décède à l'âge x et à l'année t dans l'intervalle $[\alpha_{i,t}, \beta_{i,t}] \subset [x, x + 1]$ est égale à $(\beta_{i,t} - \alpha_{i,t})q_{x,t}$, où $q_{x,t}$ est la probabilité de décès à l'âge x pour l'année t .

Le nombre de décès observés sur l'intervalle $[x, x + 1]$ de l'année t , noté $D_{x,t}$, suit une loi Binomiale de paramètres $N_{x,t}$ et $(\beta_i - \alpha_i)q_{x,t}$, c'est-à-dire,

$$D_{x,t} \sim \mathcal{B}(N_{x,t}, (\beta_{i,t} - \alpha_{i,t})q_{x,t})$$

Où $N_{x,t}$ représente le nombre d'individus vivants à l'âge x au cours de l'année t .

L'exposition au risque à l'âge x et au temps t , notée $E_{x,t}$, est alors définie par :

$$E_{x,t} = \sum_{i=1}^{N_{x,t}} (\beta_{i,t} - \alpha_{i,t})$$

L'estimateur de HOEM s'écrit alors :

$$\hat{q}_{x,t} = \frac{d_{x,t}}{E_{x,t}}$$

F Une régression polynomiale

L'équation du modèle s'écrit :

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_m x^m + \epsilon$$

Avec :

- y : la variable dépendante.
- x : la variable indépendante.
- $\beta_0, \beta_1, \dots, \beta_m$: les coefficients du modèle à estimer.
- ϵ : le terme d'erreur.

Cette équation peut être écrite sous forme matricielle :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Ainsi, cette équation peut être simplifiée en :

$$Y = X\beta + \epsilon$$

L'estimation des coefficients du modèle s'effectue par moindres carrés, conduisant à :

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

G Un lissage par splines

Un lissage par splines consiste à diviser la plage de la fonction à ajuster en plusieurs segments. Une fonction polynomiale est ajustée sur chaque segment, en veillant à ce que les fonctions soient continues aux points de jonction, appelés nœuds. Supposant l'utilisation d'un polynôme de degré 3 et un découpage de la plage en 2, l'équation s'écrit :

$$q_x = \begin{cases} p_0(x) & \text{si } x_0 \leq x \leq x_1 \\ p_1(x) & \text{si } x_1 \leq x \leq x_2 \end{cases}$$

Avec :

- $p_i(x)$: un polynôme de degré 3.
- Les contraintes de continuité suivantes :

$$\begin{aligned} p_0(x_1) &= p_1(x_1), \\ p'_0(x_1) &= p'_1(x_1), \\ p''_0(x_1) &= p''_1(x_1). \end{aligned}$$

Ainsi, les polynômes peuvent être définis comme suit :

$$p_0(x) = c_1 + c_2x + c_3x^2 + c_4x^3 \quad \text{et} \quad p_1(x) = p_0(x_1) + c_5(x - x_1)^3$$

Pour ajuster les splines aux données, un critère de moindres carrés pondérés est utilisé :

$$M = \sum_{x=x_0}^{x=x_2} w_x (q_x - \hat{q}_x)$$

Avec les w_i représentant les poids associés aux observations.

Le problème de minimisation se résout en utilisant l'équation suivante :

$$(X^t W X)c = X^t W \hat{q}$$

où X est la matrice de taille $(m, 5)$ pour m valeurs de \hat{q}_x , W est la matrice des poids, et c le vecteur des coefficients des polynômes.

Pour n nœuds, la matrice X devient de taille $(m, n + 4)$ et les coefficients sont obtenus par :

$$c = (X^t W X)^{-1} X^t W \hat{q}$$

H La méthode de WHITTAKER - HENDERSON

Cette méthode consiste à retenir des valeurs ajustées qui minimisent la combinaison linéaire des critères suivants :

- La fidélité des données.
- La régularité des données.

En fixant les poids w_i , les critères de fidélité et de régularité sont respectivement définis comme suit :

$$\begin{cases} F = \sum_{i=0}^p w_i (q_i - \hat{q}_i)^2 \\ S = \sum_{i=0}^{p-z} (\Delta^z q_i)^2 \end{cases}$$

avec z un paramètre du modèle.

Le critère à minimiser est la combinaison linéaire de la fidélité et de la régularité, le poids de chacun de ces termes étant contrôlé par le paramètre h tel que :

$$M = F + h \times S$$

En écrivant :

$$F = (q - \hat{q})^T w (q - \hat{q}), \quad S = (\Delta^z q)^T \Delta^z q,$$

avec $\hat{q} = (\hat{q}_i)_{i=1, \dots, p}$, $q = (q_i)_{i=1, \dots, p}$, $w = (w_i)_{i=1, \dots, p}$ et $\Delta^z = K_z$, une matrice de taille $(p - z, p)$ dont les termes sont les coefficients binomiaux d'ordre z et dont le signe commence positivement pour z pair.

La fonction à minimiser s'écrit alors :

$$M = q^T w q - 2q^T w \hat{q} + \hat{q}^T w \hat{q} + h K_z^T K_z$$

I Lissage des paramètres du modèle LOG POISSON

- **Le paramètre α_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre α_x des femmes et des hommes.

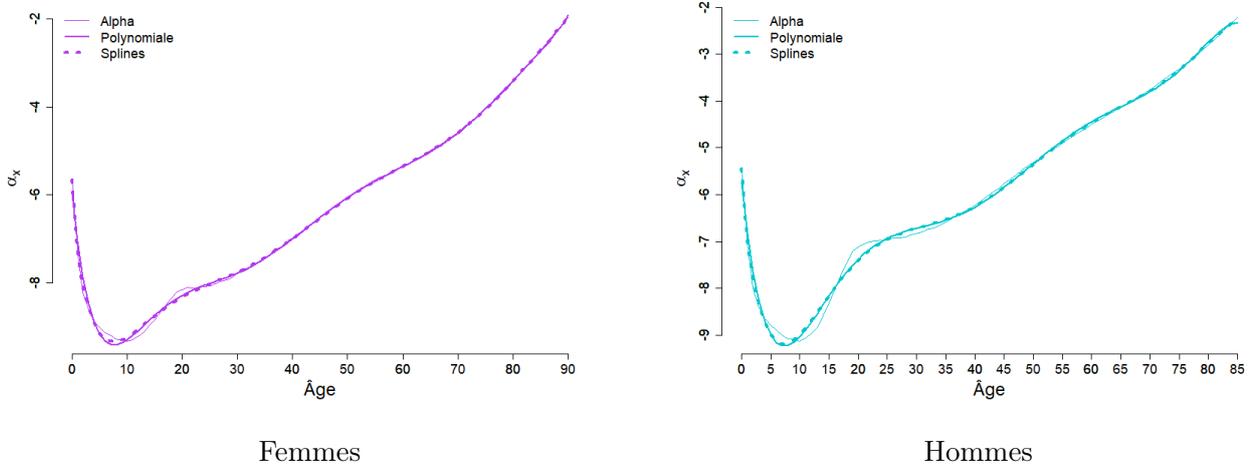


FIGURE A4 – Le paramètre $\hat{\alpha}_x$ du modèle LOG POISSON lissé

Les lissages par polynômes et par splines montrent tous les deux une bonne adaptation à la forme générale des données. L'analyse visuelle révèle une cohérence dans la captation des grandes tendances avec une légère divergence dans les zones où les fluctuations sont plus marquées.

	Femmes	Hommes
Polynomiale	0,9985	0,9962
Splines	0,9992	0,9969

TABLE A1 – Le lissage du paramètre α_x de la méthode LOG POISSON : Les valeurs de R^2

Les valeurs élevées du R^2 montrent une excellente capacité de ces deux méthodes à capturer les variations de α_x , confirmant que les tendances globales sont bien reproduites par les modèles. La méthode des splines est retenue dans la suite de ce mémoire.

- **Le paramètre β_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre β_x des femmes et des hommes.

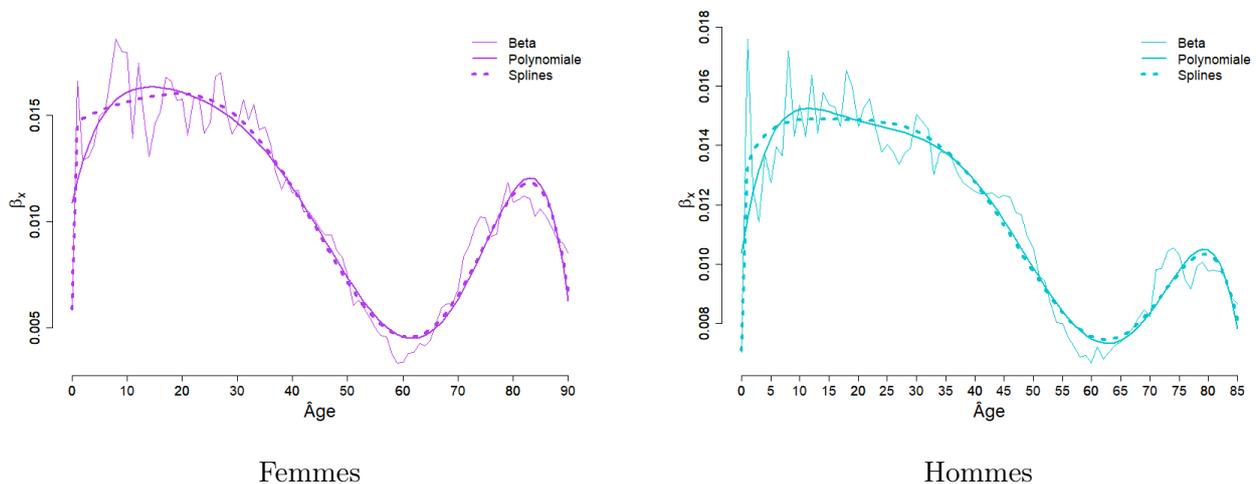


FIGURE A5 – Le paramètre $\hat{\beta}_x$ du modèle LOG POISSON lissé

Les graphiques montrent que les lissages réduisent les fluctuations des données tout en préservant leur tendance générale.

	Femmes	Hommes
Polynomiale	0,9072	0,8771
Splines	0,9337	0,9023

TABLE A2 – Le lissage du paramètre β_x de la méthode LOG POISSON : Les valeurs de R^2

Les R^2 sont légèrement inférieurs pour β_x par rapport à α_x ce qui suggère que les lissages capturent bien les variations globales, mais qu'il reste des fluctuations locales difficiles à modéliser parfaitement. La méthode des splines est retenue pour la suite de ce mémoire.

— **Le paramètre κ_t** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre κ_t des femmes et des hommes.

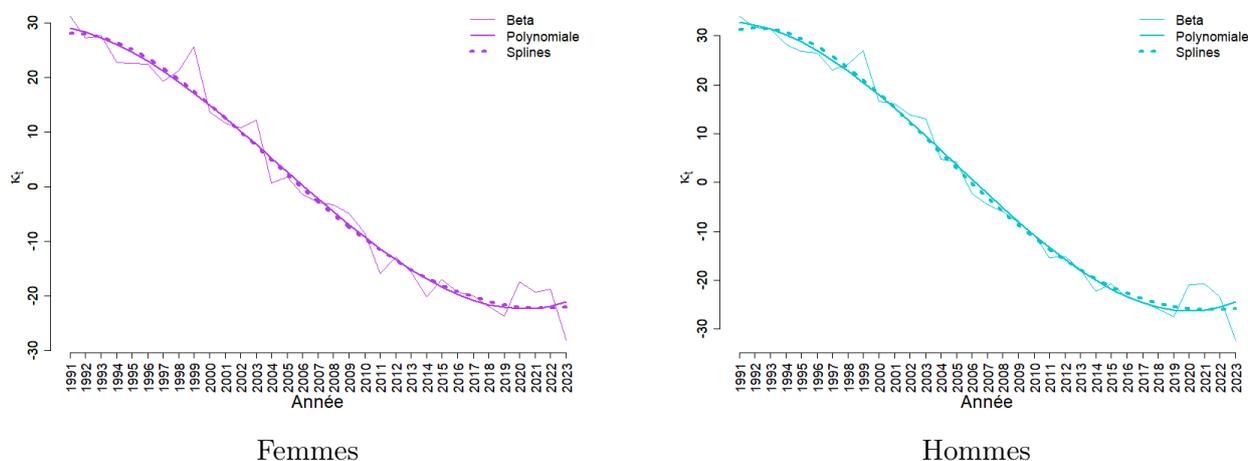


FIGURE A6 – Le paramètre $\hat{\kappa}_t$ du modèle LOG POISSON lissé

Les tendances temporelles sont bien capturées par les deux méthodes de lissage.

	Femmes	Hommes
Polynomiale	0,9745	0,985
Splines	0,9751	0,9859

TABLE A3 – Le lissage du paramètre κ_t de la méthode LOG POISSON : Les valeurs de R^2

Les valeurs élevées de R^2 pour κ_t indiquent que les lissages par polynômes et splines reproduisent efficacement les tendances temporelles. La méthode des splines est retenue pour lisser les surfaces de mortalité. La série brute sera utilisée pour l'extrapolation des taux de mortalité afin de tenir compte de la bosse de 2019.

J Lissage des paramètres du modèle de RENSHAW & HABERMAN

- **Le paramètre α_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre α_x des femmes et des hommes.

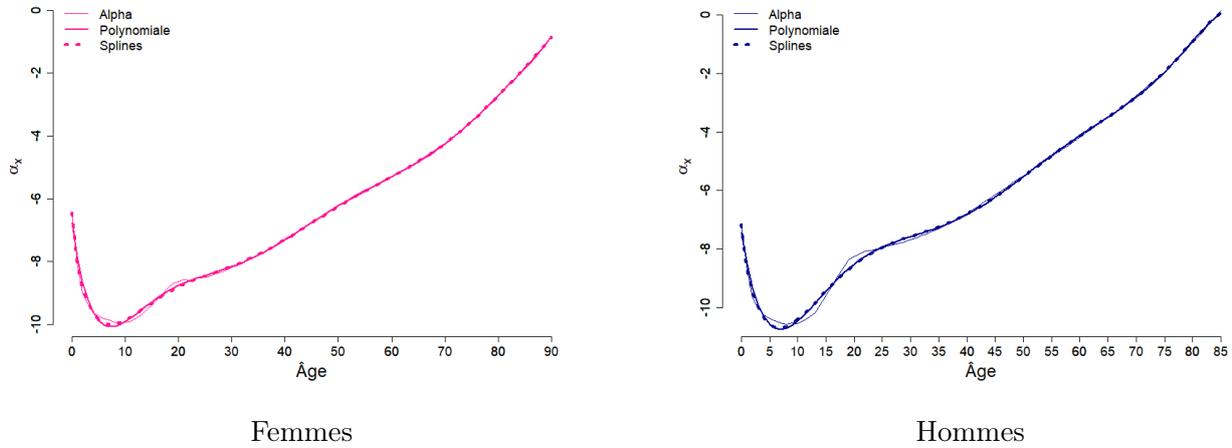


FIGURE A7 – Le paramètre $\hat{\alpha}_x$ du modèle de RENSHAW & HABERMAN lissé

L'ajustement visuel du paramètre α_x montre une correspondance presque parfaite entre les courbes lissées et les données brutes, avec une capture très précise de la tendance générale de la mortalité en fonction de l'âge.

	Femmes	Hommes
Polynomiale	0,999	0,9984
Splines	0,9994	0,9987

TABLE A4 – Le lissage du paramètre α_x de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2

Le lissage des paramètres a abouti à des R^2 très élevés pour les hommes et les femmes, indiquant une excellente correspondance entre les valeurs lissées et les données brutes. La méthode des splines est retenue pour la suite du mémoire.

- **Le paramètre β_x** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre β_x des femmes et des hommes.

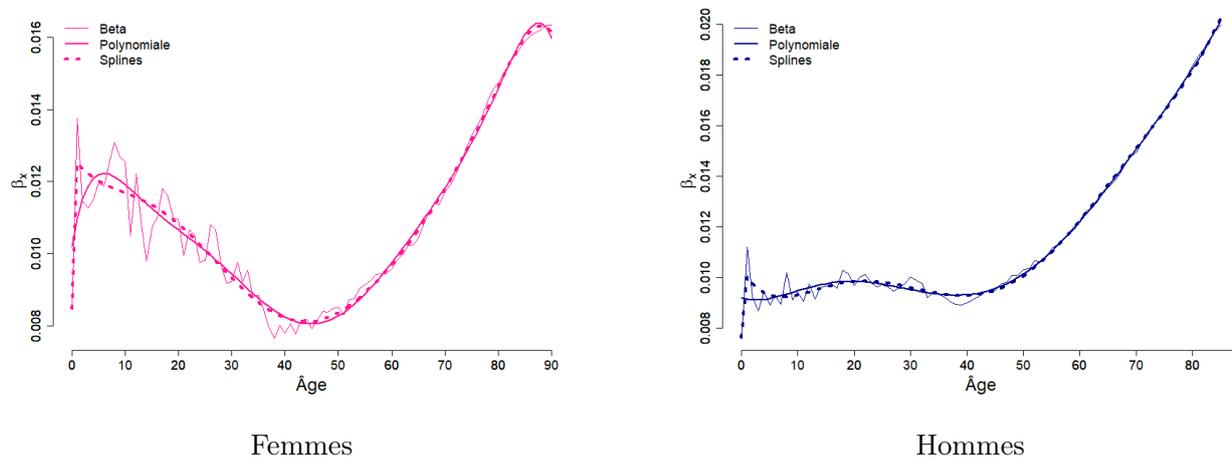


FIGURE A8 – Le paramètre $\hat{\beta}_x$ du modèle de RENSHAW & HABERMAN lissé

Visuellement, les courbes lissées pour le paramètre β_x suivent de près les fluctuations des données brutes.

	Femmes	Hommes
Polynomiale	0,9506	0,9878
Splines	0,9645	0,9929

TABLE A5 – Le lissage du paramètre β_x de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2

Ces valeurs montrent que les deux méthodes capturent la variabilité du paramètre β_x . La méthode des splines présente des valeurs de R^2 légèrement plus élevées. Ainsi, cette méthode est retenue dans la suite de ce mémoire pour les femmes et les hommes.

— **Le paramètre κ_t** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre κ_t des femmes et des hommes.

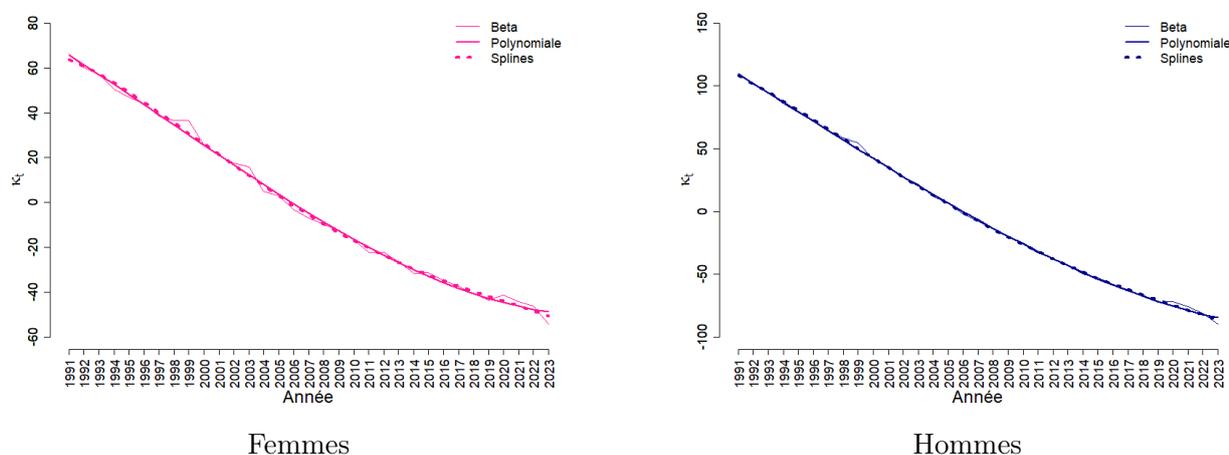


FIGURE A9 – Le paramètre $\hat{\kappa}_t$ du modèle de RENSHAW & HABERMAN lissé

L'ajustement visuel du paramètre κ_t révèle une excellente capture des tendances temporelles, avec des courbes lissées qui suivent étroitement la trajectoire descendante des données brutes, témoignant d'une modélisation efficace des changements dans la mortalité au fil du temps.

	Femmes	Hommes
Polynomiale	0,9964	0,9992
Splines	0,997	0,9993

TABLE A6 – Le lissage du paramètre κ_t de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2

Les R^2 sont extrêmement élevés, soulignant que les tendances temporelles sont bien capturées par les deux méthodes. La très légère différence entre les méthodes indique que les deux approches sont presque équivalentes pour ce paramètre, avec une infime supériorité des splines. La méthode des splines est retenue pour le lissage des taux de mortalité. Il est important de noter que la série temporelle brut sera utilisée pour la projection future.

— **Le paramètre γ_{t-x}** : Les graphiques ci-dessous illustrent les résultats du lissage pour le paramètre γ_{t-x} des femmes et des hommes.

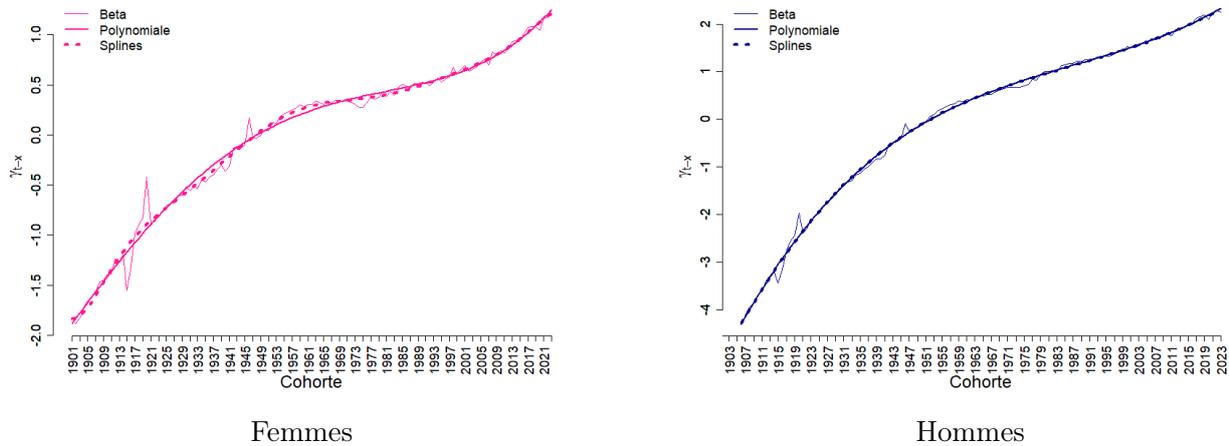


FIGURE A10 – Le paramètre $\hat{\gamma}_{t-x}$ du modèle de RENSHAW & HABERMAN lissé

Le paramètre γ_{t-x} , représentant l'effet de cohorte, présente un ajustement visuel très précis, où les courbes lissées reproduisent fidèlement les effets de génération observés dans les données brutes, mettant en évidence des tendances cohérentes et bien modélisées à travers les cohortes.

	Femmes	Hommes
Polynomiale	0,989	0,9976
Splines	0,9907	0,9976

TABLE A7 – Le lissage du paramètre γ_{t-x} de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2

L'excellent R^2 pour ce paramètre montre que le modèle est capable de capturer efficacement l'effet de cohorte. La méthode des splines est retenue pour l'obtention de taux de mortalité bruts. La série brute sera utilisée dans la projection future des taux de mortalité.

K Application de la méthode de BOX - JENKINS sur la série de 1991 à 2019

1. La détermination de la stationnarité :

Les autocorrélogrammes de la série κ_t montrent des signes de non-stationnarité, tant chez les hommes que chez les femmes, caractérisés par une décroissance lente des coefficients d'autocorrélation.

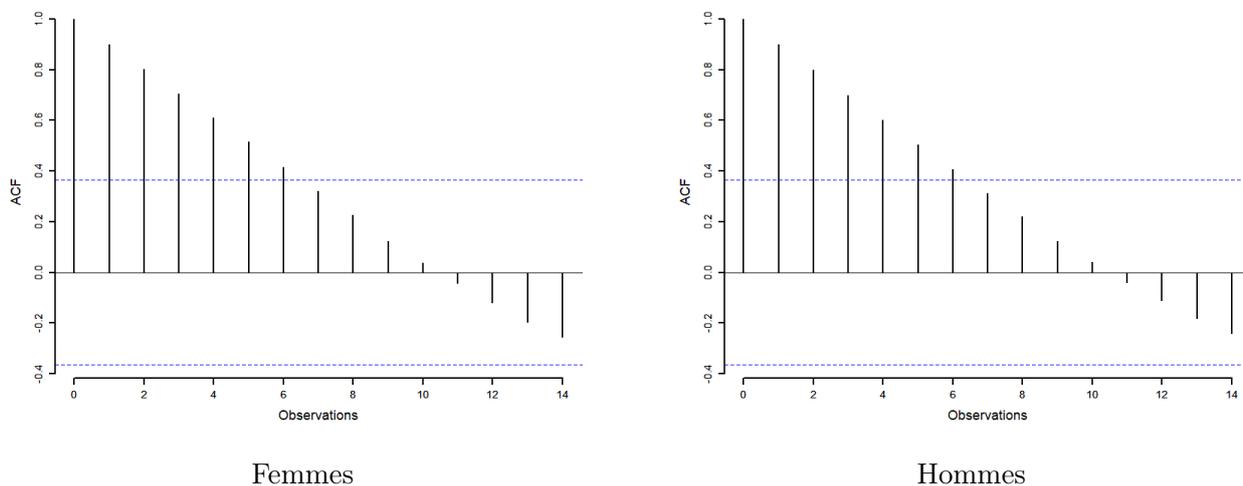


FIGURE A11 – Les autocorrélogrammes (ACF) des séries κ_t de 1991 à 2019

Après avoir différencié la série κ_t une fois, une décroissance rapide des coefficients d'autocorrélation est observée, indiquant que la série est devenue stationnaire.

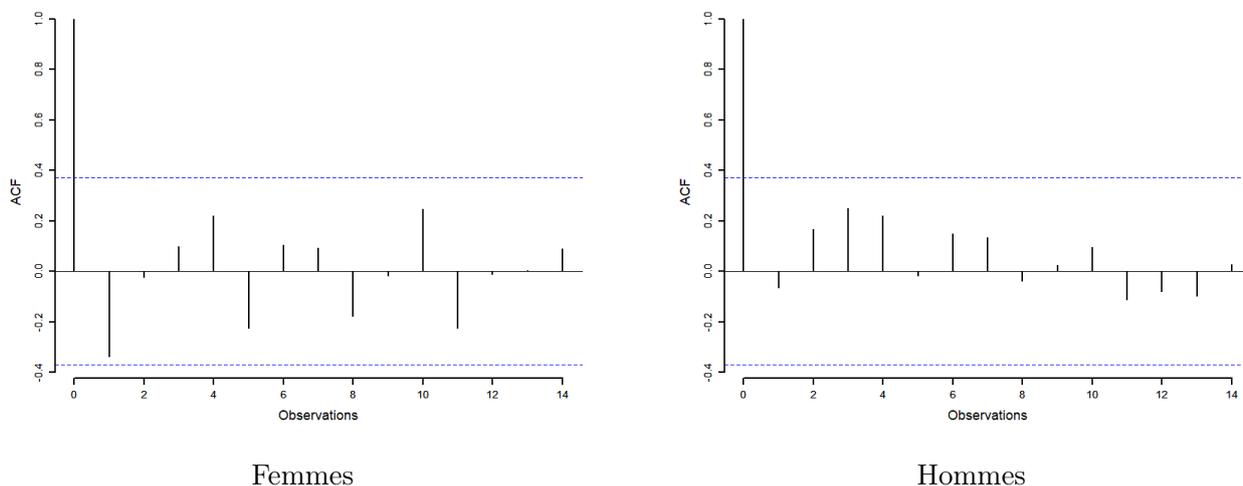


FIGURE A12 – Les autocorrélogrammes (ACF) des séries κ_t différenciées une fois de 1991 à 2019

→ Pour les hommes et pour les femmes, le nombre de différenciations retenu est $d = 1$.

2. L'identification des ordres p, q :

Les ACF des séries κ_t différenciées, présentés ci-dessus, pour les hommes et pour les femmes, montrent une chute notable après le premier lag, suggérant un ordre q probablement faible pour le modèle *MA*.

Les PACF des séries κ_t différenciées montrent des valeurs proches de zéro, suggérant un ordre p très faible ou nul pour les hommes et les femmes.

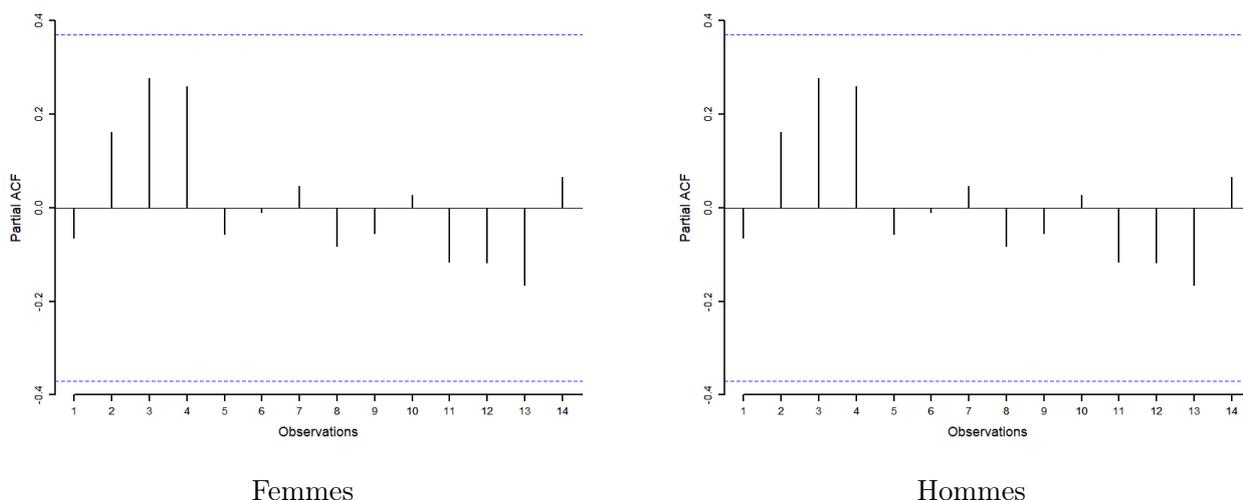


FIGURE A13 – Les autocorrélogrammes partiels (PACF) des séries κ_t différenciées une fois de 1991 à 2019

→ Pour les hommes et pour les femmes, les ordres q et p sont fixés égale à 0, $p = 0, q = 0$.

3. L'estimation des paramètres :

Le modèle retenu ne comportant pas de composante autorégressive ni de moyenne mobile ($p = 0, q = 0$), aucune estimation de paramètre n'est nécessaire.

4. L'analyse de résidus :

— La centralité :

Les résidus du modèle ARIMA apparaissent centrés autour de zéro sur les graphiques ci-dessous.

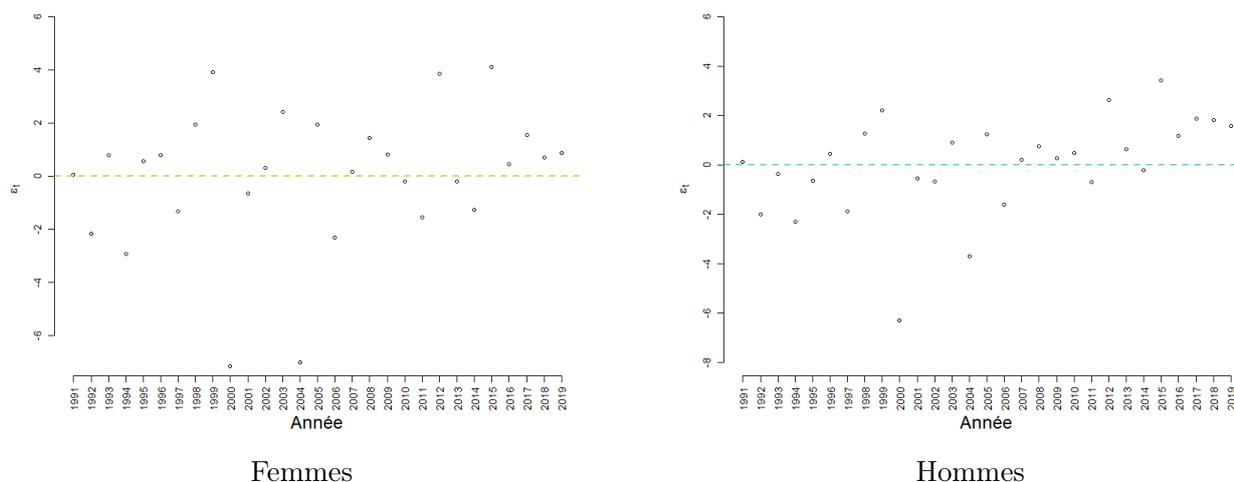


FIGURE A14 – Les nuages des points des résidus de la série 1991-2019

Pour les femmes, la moyenne des résidus est très proche de zéro, 0,024, ce qui suggère une bonne centralité. Pour les hommes, la moyenne des résidus est égale à 0,004, indiquant également une centralité proche de zéro.

Les valeurs élevées des p-values, du test de Student, indique que l'hypothèse nulle, selon laquelle la moyenne des résidus est égale à zéro, ne peut être rejetée. Ainsi, les résidus sont bien centrés autour de zéro pour les hommes et pour les femmes.

	Femmes	Hommes
p-value	0,9961	0,9914

TABLE A8 – Les p-values du test de Student pour la série de 1991 à 2019

— **L'indépendance :**

Les autocorrélogrammes ci-dessous montrent que la majorité des coefficients d'autocorrélation sont proches de zéro et se trouvent à l'intérieur de l'intervalle de confiance, tant pour les femmes que pour les hommes. Cette distribution suggère qu'il n'y a pas d'autocorrélation résiduelle significative parmi les résidus.

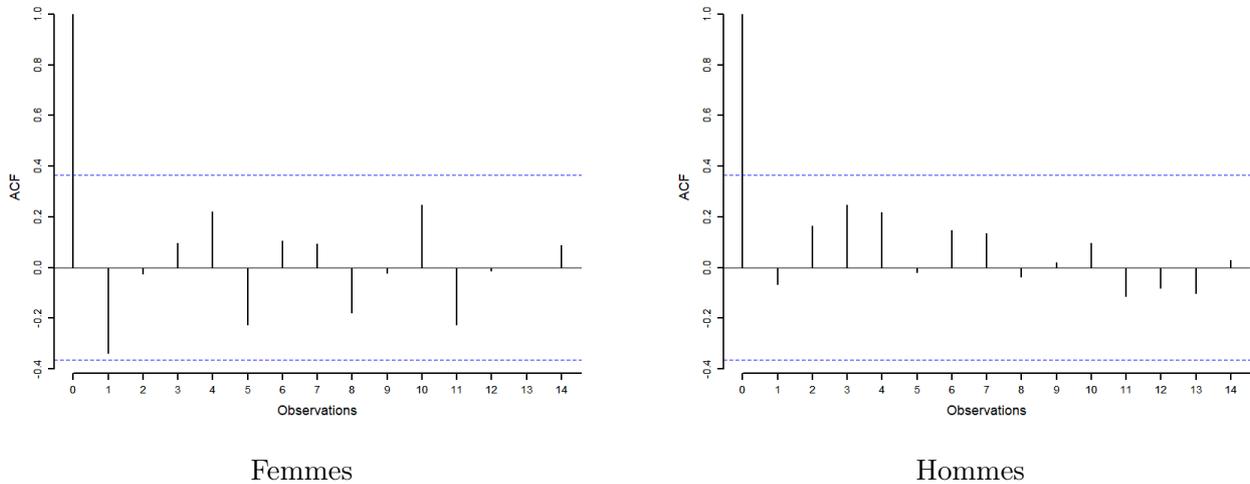


FIGURE A15 – Les autocorrélogrammes des résidus de la série 1991-2019

Pour les femmes, le test de BOX-PIERCE donne une valeur de p-value supérieure au seuil de 5%, indiquant qu'il n'y a pas d'autocorrélation significative dans les résidus. Pour les hommes, le test confirme également l'absence d'autocorrélation significative. Cela suggère que les résidus du modèle ARIMA retenu sont bien indépendants.

	Femmes	Hommes
p-value	0,06822	0,7227

TABLE A9 – Les p-values du test de BOX PIERCE pour la série de 1991 à 2019

— **La distribution normale :**

Les histogrammes ci-dessous montrent que les distributions semblent bien ajustées à la courbe normale, suggérant que les résidus sont proches d'une distribution normale.

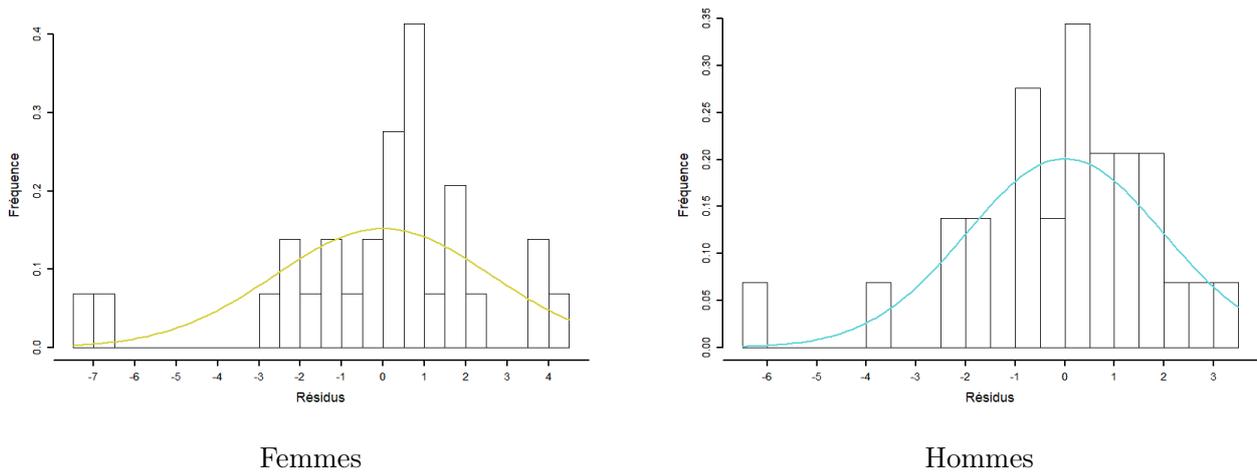


FIGURE A16 – Les histogrammes des résidus de la série 1991-2019

Les points des QQ-plot suivent globalement la diagonale, ce qui indique que les résidus sont proches d’une distribution normale, bien que quelques valeurs s’écartent légèrement de la droite théorique.

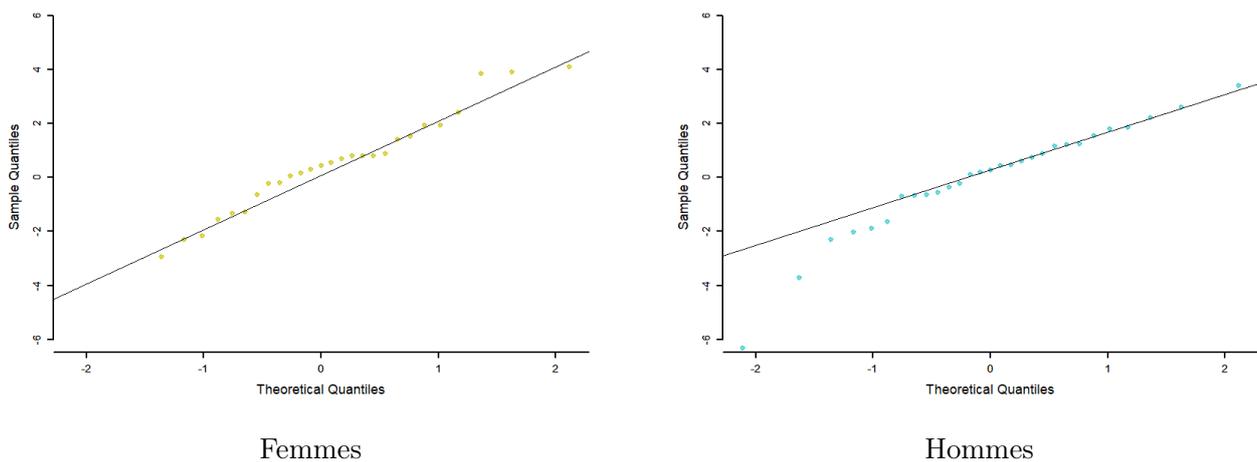


FIGURE A17 – Les QQ-plots des résidus de la série 1991-2019

Tant pour les femmes que pour les hommes, la p-value du test de SHAPIRO est supérieure au seuil de 5%. L’hypothèse nulle, selon laquelle les résidus suivent une distribution normale, ne peut pas être rejetée.

	Femmes	Hommes
p-value	0,357	0,05792

TABLE A10 – Les p-values du test de Shapiro pour la série de 1991 à 2019

L Les modèles ETS

Erreur additive :

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

Erreur multiplicative :

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha (\ell_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$

FIGURE A18 – Les modèles ETS

M Des méthodes de positionnement

L'application d'un coefficient de majoration ou de réduction de façon multiplicative

Dans la pratique, le coefficient de majoration ou de réduction appliqué est souvent un ratio de mortalité standardisé (SMR). Ce dernier sert d'indicateur de la mortalité du portefeuille par rapport à celle de la population de référence et se calcule comme suit :

$$SMR = \frac{\sum_{x,t} D_{x,t}}{\sum_{x,t} E_{x,t} \times q_{x,t}^{ref}}$$

où :

- $D_{x,t}$: le nombre de décès observés à l'âge x et au temps t .
- $E_{x,t}$: l'exposition du portefeuille à l'âge x et au temps t .
- $q_{x,t}^{ref}$: le taux de mortalité à l'âge x et au temps t de la table de mortalité de référence.

Ainsi, les taux de décès spécifiques au portefeuille, notés $\tilde{q}_{x,t}$, peuvent être déterminés comme suit :

$$\tilde{q}_{x,t} = q_{x,t}^{ref} \times SMR$$

L'utilisation d'un modèle relationnel

Une autre méthode couramment utilisée en actuariat est l'application d'une variante du modèle semi-paramétrique de BRASS.

Dans ce cas, une transformation logistique est utilisée pour modéliser les taux de mortalité de la manière suivante :

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

où :

- α et β sont des paramètres d'ajustement spécifiques au modèle.
- $\epsilon_{x,t}$ représente le terme d'erreur pour l'âge x et l'année t .

Ainsi, les taux de décès spécifiques au portefeuille, notés $\tilde{q}_{x,t}$, peuvent être déterminés comme suit :

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

Table des figures

1	Les perspectives du mémoire	8
2	Heatmap des résidus de Pearson du modèle RENSHAW & HABERMAN	9
3	Les projections de la série κ_t avec le modèle <i>ARIMA</i>	10
4	Le schéma synthétique de la construction classique	11
5	Les projections avec les modèles ETS et PROPHET	12
6	Les projections avec les modèles ARIMA et PROPHET	13
7	Le schéma synthétique de construction alternative	14
8	Les sommes des PMT	15
9	The thesis perspectives	18
10	Heatmap of Pearson residuals for the RENSHAW & HABERMAN model	19
11	Projections of the κ_t series using the <i>ARIMA</i> model	20
12	The synthetic framework of the classical construction	21
13	Projections with the ETS and PROPHET models	22
14	Projections with the ARIMA and PROPHET models	23
15	The synthetic framework of alternative construction	24
16	The totals of TMR	25
1.1	Les risques de mortalité et de longévité	29
1.2	Les tables d'expérience : la construction	32
1.3	Synthèse sur l'élaboration d'une table d'expérience	33
2.1	Le diagramme de Lexis - Décès à l'âge x et à l'année t	40
2.2	Le diagramme de Lexis - Décès de la génération $t - x$ à l'âge x	40
2.3	Le diagramme de Lexis - Décès de la génération $t - x$ au temps t	40
2.4	Le diagramme de Lexis - Décès de la génération $t - x$ à l'âge x et au temps t	40
3.1	L'évolution de l'espérance de vie à la naissance en France de 1740 à 2020	45
3.2	L'espérance de vie	45
3.3	L'écart d'espérance de vie entre les femmes et les hommes en années	46
3.4	Le schéma synthétique des facteurs	47
3.5	Le schéma synthétique des objectifs du mémoire	48
4.1	Le nombre d'individus et l'âge moyen des individus par année	50
4.2	Le nombre de décès et l'âge moyen des décès par année	51
4.3	Les logarithmes des taux de mortalité bruts selon le sexe	53
4.4	Les taux de mortalité bruts	54
4.5	Les taux de mortalité bruts selon l'âge par année	55
4.6	Les logarithmes des taux de mortalité bruts selon l'âge par année	55
4.7	Les pyramides âges - Population au 1 ^{er} janvier en 2023 et en 2070	56

4.8	Les logarithmes des taux de mortalité bruts selon l'année pour les 15-30 ans	57
4.9	Les logarithmes des taux de mortalité bruts selon l'année pour les 60-75 ans	57
4.10	Les taux de mortalité et population	58
4.11	Les nombres de décès	59
4.12	Le schéma synthétique des données	60
1.1	Le paramètre $\hat{\alpha}_x$ du modèle de LEE & CARTER	65
1.2	Le paramètre $\hat{\beta}_x$ du modèle de LEE & CARTER	65
1.3	Le paramètre $\hat{\kappa}_t$ du modèle de LEE & CARTER	66
1.4	Le logarithme des taux de mortalité estimés par le modèle de LEE & CARTER	66
1.5	Le paramètre $\hat{\alpha}_x$ du modèle de LEE & CARTER lissé	67
1.6	Le paramètre $\hat{\beta}_x$ du modèle de LEE & CARTER lissé	68
1.7	Le paramètre $\hat{\kappa}_t$ du modèle de LEE & CARTER lissé	69
1.8	Le logarithme des taux de mortalité estimés (lissés) par le modèle de LEE & CARTER . .	70
1.9	Le paramètre $\hat{\alpha}_x$ du modèle LOG POISSON	72
1.10	Le paramètre $\hat{\beta}_x$ du modèle de LOG POISSON	73
1.11	Le paramètre $\hat{\kappa}_t$ du modèle de LOG POISSON	73
1.12	Le logarithme des taux de mortalité estimés (lissés) par le modèle LOG POISSON	74
1.13	Le paramètre $\hat{\alpha}_x$ du modèle de RENSHAW & HABERMAN	76
1.14	Le paramètre $\hat{\beta}_x$ du modèle de RENSHAW & HABERMAN	76
1.15	Le paramètre $\hat{\kappa}_t$ du modèle de RENSHAW & HABERMAN	76
1.16	Paramètres $\hat{\gamma}_{t-x}$ du modèle de RENSHAW & HABERMAN	77
1.17	Le logarithme des taux de mortalité estimés (lissés) par le modèle de RENSHAW & HA- BERMAN	77
1.18	La comparaison du paramètre α_x	78
1.19	La comparaison du paramètre β_x	79
1.20	La comparaison du paramètre κ_t	79
2.1	Les taux de décès prédits et observés en 2015	81
2.2	Les intervalles de confiance de 0 à 40 ans en 2015 du modèle de LEE & CARTER	82
2.3	Les intervalles de confiance de 40 à 60 ans en 2023 pour les femmes	83
2.4	Les intervalles de confiance de 60 à 75 ans en 2018 pour le modèle de RENSHAW & HA- BERMAN	83
2.5	Les résidus de la réponse par âge en 2023 pour le modèle de LEE & CARTER	84
2.6	Les résidus de la réponse par âge en 2018 pour le modèle de RENSHAW & HABERMAN	85
2.7	Heatmap des résidus de Pearson du modèle de LEE & CARTER	86
2.8	Heatmap des résidus de Pearson du modèle LOG POISSON	86
2.9	Heatmap des résidus de Pearson du modèle RENSHAW & HABERMAN	87
2.10	Les SMR par âge, toutes années confondues	88
2.11	Les SMR par année, tous âges confondus	89
2.12	Les valeurs du χ^2 par âge, toutes années confondues	90
2.13	Les valeurs du χ^2 par année, tous âges confondus	90
2.14	Les valeurs du MAPE par âge, toutes années confondues	92
2.15	Les valeurs du MAPE par année, tous âges confondus	92
2.16	Les valeurs du R^2 par âge, toutes années confondues	93
2.17	Les valeurs du R^2 par année, tous âges confondus	94

3.1	La méthodologie BOX & JENKINS	98
3.2	Les autocorrélogrammes (ACF) des séries κ_t de 1991 à 2023	101
3.3	Les autocorrélogrammes (ACF) des séries κ_t différenciées une fois de 1991 à 2023	102
3.4	Les autocorrélogrammes partiels (PACF) des séries κ_t différenciées une fois de 1991 à 2023	102
3.5	Les nuages des points des résidus de la série 1991-2023	103
3.6	Les autocorrélogrammes des résidus de la série 1991-2023	104
3.7	Les histogrammes des résidus de la série 1991-2023	105
3.8	Les QQ-plots des résidus de la série 1991-2023	105
3.9	Les projections de la série κ_t avec le modèle <i>ARIMA</i>	106
3.10	Le logarithme des taux de mortalité projetés avec le modèle ARIMA (1991-2023)	108
3.11	Les taux de mortalité projetés à 30 ans	108
4.1	Les taux de mortalité fermés avec COALE & KISKER (scénario 1991-2023)	111
4.2	Les taux de mortalité fermés avec DENUIT & GODERNIAUX (scénario 1991-2023)	113
4.3	Les logarithmes des taux de mortalité fermés pour l'année 2023	113
4.4	Le schéma synthétique de la partie	115
1.1	La validation croisée dans le cas de données non ordonnées	121
1.2	La validation croisée dans le cas de données ordonnées	121
1.3	Les projections avec différentes combinaisons ETS	122
1.4	<i>L'analyst-in-the-loop</i>	123
1.5	Les projections avec le modèle de PROPHET (Approche 1)	128
1.6	Les projections avec le modèle de PROPHET (Approche 2)	129
1.7	Les projections avec le modèle de PROPHET (Comparaison approches 2 et 3)	130
1.8	Les projections avec le modèle de PROPHET (Comparaison approches 3 et 4)	131
1.9	Les projections avec les modèles ETS et PROPHET	132
1.10	Les projections avec les modèles ARIMA et PROPHET	133
2.1	Le logarithme des taux de mortalité projetés avec le modèle PROPHET	136
2.2	Les taux de mortalité projetés à 30 ans avec le modèle classique et le modèle PROPHET	137
2.3	Les taux de mortalité fermés par DENUIT & GODERNIAUX (PROPHET)	138
2.4	Le schéma synthétique de la partie	139
1.1	La pyramide des âges des assurés actifs ou radiés	143
1.2	La pyramide des âges des bénéficiaires de droits directs	143
1.3	La pyramide des âges des bénéficiaires de droits indirects	144
2.1	Les probabilités de survie du participant (actifs-radiés)	147
2.2	Les sommes des PMT	151
A1	Les tables d'expérience : la certification	158
A2	Les tables d'expérience : le suivi	159
A3	Les pyramides des âges au 1 ^{er} janvier	160
A4	Le paramètre $\hat{\alpha}_x$ du modèle LOG POISSON lissé	165
A5	Le paramètre $\hat{\beta}_x$ du modèle LOG POISSON lissé	165
A6	Le paramètre $\hat{\kappa}_t$ du modèle LOG POISSON lissé	166
A7	Le paramètre $\hat{\alpha}_x$ du modèle de RENSHAW & HABERMAN lissé	167
A8	Le paramètre $\hat{\beta}_x$ du modèle de RENSHAW & HABERMAN lissé	167

A9	Le paramètre $\hat{\kappa}_t$ du modèle de RENSHAW & HABERMAN lissé	168
A10	Le paramètre $\hat{\gamma}_{t-x}$ du modèle de RENSHAW & HABERMAN lissé	169
A11	Les autocorrélogrammes (ACF) des séries κ_t de 1991 à 2019	170
A12	Les autocorrélogrammes (ACF) des séries κ_t différenciées une fois de 1991 à 2019	170
A13	Les autocorrélogrammes partiels (PACF) des séries κ_t différenciées une fois de 1991 à 2019	171
A14	Les nuages des points des résidus de la série 1991-2019	171
A15	Les autocorrélogrammes des résidus de la série 1991-2019	172
A16	Les histogrammes des résidus de la série 1991-2019	173
A17	Les QQ-plots des résidus de la série 1991-2019	173
A18	Les modèles ETS	174

Liste des tableaux

1	Les valeurs des BIC	9
2	Les valeurs du MSE des modèles ETS et PROPHET	13
3	The values of the BIC	19
4	The MSE values of the ETS and PROPHET models	23
1.1	Le lissage du paramètre α_x de la méthode de LEE & CARTER : Les valeurs de R^2	68
1.2	Le lissage du paramètre β_x de la méthode de LEE & CARTER : Les valeurs de R^2	68
1.3	Le lissage du paramètre κ_t de la méthode de LEE & CARTER : Les valeurs de R^2	69
2.1	Les SMR globaux, tous âges et toutes années confondus	89
2.2	Les valeurs globales du χ^2	91
2.3	Les valeurs globales du MAPE	92
2.4	Les valeurs globales du R^2	94
2.5	Les valeurs du BIC	95
3.1	Les p-values du test de STUDENT pour la série de 1991 à 2023	104
3.2	Les p-values du test de BOX PIERCE pour la série de 1991 à 2023	104
3.3	Les p-values du test de SHAPIRO pour la série de 1991 à 2023	105
1.1	Les combinaisons des modèles ETS	118
1.2	Les critères d'information du modèle ETS	123
1.3	Les valeurs du MSE selon la flexibilité du modèle	128
1.4	Les valeurs du MSE selon la saisonnalité	130
1.5	Les valeurs du MSE des modèles ETS et PROPHET	132
2.1	Exemple actif-radié : Les coefficients actuariels	148
2.2	Exemple actif-radié : Les montants de PMT	149
2.3	Les montants de PMT avec et sans actualisation	151
A1	Le lissage du paramètre α_x de la méthode LOG POISSON : Les valeurs de R^2	165
A2	Le lissage du paramètre β_x de la méthode LOG POISSON : Les valeurs de R^2	166
A3	Le lissage du paramètre κ_t de la méthode LOG POISSON : Les valeurs de R^2	166
A4	Le lissage du paramètre α_x de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2	167
A5	Le lissage du paramètre β_x de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2	168
A6	Le lissage du paramètre κ_t de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2	168
A7	Le lissage du paramètre γ_{t-x} de la méthode de RENSHAW & HABERMAN : Les valeurs de R^2	169
A8	Les p-values du test de Student pour la série de 1991 à 2019	172
A9	Les p-values du test de BOX PIERCE pour la série de 1991 à 2019	172

A10 Les p-values du test de Shapiro pour la série de 1991 à 2019 173

Liste des abréviations

- INSEE : Institut National de la Statistique et des Études Économiques.
- INED : Institut National d'Études Démographiques.
- CJUE : Cour de Justice de l'Union Européenne.
- SMR : *Standardised Mortality Ratio*.
- ARIMA : *AutoRegressive Integrated Moving Average*.
- MSE : Erreur quadratique moyenne.
- MAPE : *Mean Absolute Percentage Error*.
- BIC : Critère d'Information Bayésien.
- AIC : Critère d'Information d'Akaike.
- AICc : Critère d'Information d'Akaike corrigé.
- ACF : autocorrélogrammes.
- PACF : autocorrélogrammes partiels.
- ETS : Error, Trend, Seasonal.
- PMT : Provisions Mathématiques Théoriques.
- PERO : Plans d'Épargne Retraite Obligatoires.
- OMS : Organisation Mondiale de la Santé.
- LSTM : *Long Short-Term Memory*.

