

Mémoire présenté le :

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

Par : Narimane Mahmoud

Titre : Effet du choc de la Covid-19 en année *jump-off* sur les estimations des espérances
de vie résiduelles

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

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

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

signature

Entreprise : Seabird Conseil

Nom :

Signature :

Directeur de mémoire en entreprise :

Nom :

Signature :


Invité :

Nom :

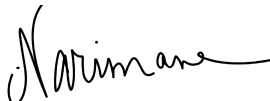
Signature :

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

Signature du responsable entreprise



Signature du candidat



Résumé

La pandémie de Covid-19 a provoqué un choc de mortalité en 2020, pouvant impacter les modèles actuariels utilisés pour l'estimation de la longévité et la valorisation des engagements en assurance vie. Ce mémoire étudie l'effet de ce choc lorsqu'il est pris en compte en année *jump-off*, c'est-à-dire en dernière année de calibration des modèles de mortalité. L'objectif est d'analyser comment l'intégration de la Covid-19 influence les estimations des espérances de vie résiduelles, en comparant plusieurs pays : la France, l'Espagne, le Royaume-Uni, le Japon et le Canada.

L'étude repose sur des modèles de mortalité stochastiques, notamment la famille Lee-Carter et ses extensions telles que Renshaw-Haberman et Age-Period-Cohort. Deux scénarios sont comparés : un scénario Baseline, qui exclut l'effet du choc pandémique et permet de modéliser la situation où la Covid-19 n'aurait jamais eu lieu, et un scénario Covid, qui intègre l'année 2020 dans la calibration.

Nous allons examiner l'impact de ces scénarios sur un portefeuille de rentiers viagers, afin de quantifier comment les projections de mortalité influencent les engagements financiers des assureurs. Une approche mixte est également étudiée, combinant les projections des deux scénarios avec une atténuation progressive de l'effet du choc sur une période choisie, permettant ainsi de revenir vers une tendance normale à long terme.

Mots clés : Lee-Carter, Renshaw-Haberman, Age-Period-Cohort, Année jump-off, AIC, BIC, Espérances de vie résiduelles, Séries temporelles, Rentes viagères

Abstract

The Covid-19 pandemic caused a mortality shock in 2020, potentially impacting actuarial models used for longevity estimation and the valuation of life insurance liabilities. This thesis examines the effect of this shock when considered as a jump-off year, meaning the last calibration year of mortality models. The purpose of this study is to analyze how integrating Covid-19 influences the estimation of residual life expectancies, comparing several countries: France, Spain, the United Kingdom, Japan, and Canada.

The study is based on stochastic mortality models, particularly the Lee-Carter family and its extensions such as Renshaw-Haberman and Age-Period-Cohort. Two scenarios are compared: a Baseline scenario, which excludes the impact of the pandemic and models a situation where Covid-19 never occurred, and a Covid scenario, which incorporates the year 2020 into the calibration.

We analyze the impact of these scenarios on a portfolio of life annuitants to quantify how mortality projections influence insurers' financial liabilities. A hybrid approach is also explored, combining projections from both scenarios with a gradual attenuation of the shock over a chosen period, thus allowing a return to a normal long-term trend.

Keywords: Lee-Carter, Renshaw-Haberman, Age-Period-Cohort, Jump-off Year, AIC, BIC, Residual Life Expectancies, Time Series, Life Annuities

Remerciements

Je souhaite remercier tout particulièrement Quentin EMERY, de m'avoir encadrée sur ce mémoire. Je lui suis extrêmement reconnaissante pour le temps précieux qu'il m'a accordé au cours de mes travaux, ses conseils avisés et ses relectures.

Je remercie également Monsieur Quelin STANISLAS, mon tuteur pédagogique, pour m'avoir suivie tout au long de cette étude.

Enfin, je tiens à exprimer ma gratitude à Jonathan DEVILLARD pour son aide et son soutien dans la réalisation de ce mémoire.

Note de synthèse

Avant la pandémie de la Covid-19 en 2020, les tendances générales de mortalité étaient à la baisse, principalement grâce aux progrès médicaux et aux améliorations des conditions de vie. L'impact de ce phénomène sur les engagements d'assurance est non négligeable et doit être pris en compte, notamment concernant le risque de longévité. L'enjeu pour l'assureur est d'anticiper cette tendance afin de pouvoir respecter ses engagements envers les assurés.

La pandémie a introduit une rupture soudaine dans ces tendances, provoquant un choc de mortalité vers la hausse. L'objectif de ce mémoire est d'analyser l'impact de la prise en compte de ce choc pandémique sur la dernière année de calibration, appelée l'année *jump-off*, sur les projections des espérances de vie résiduelles. Il ne s'agit pas de déterminer si la pandémie a réellement modifié les tendances de mortalité ou non, mais plutôt d'examiner comment sa prise en compte dans les modèles pourraient influencer les estimations actuarielles et de comparer ces effets entre différentes populations dans 5 pays : la France, l'Espagne, le Royaume-Uni, le Japon et le Canada, avec une segmentation entre les hommes et les femmes.

Méthodologie et données

Nous avons obtenu les données historiques à partir de la *Human Mortality Database*, une base de données en libre accès dans laquelle les données sur la mortalité et la population, provenant d'agences nationales autorisées, sont regroupées et standardisées. Nous avons décidé de nous concentrer sur les âges allant de 15 à 110 ans. Les modèles étudiés sont la famille des modèles Lee-Carter, plus précisément le modèle Lee-Carter poissonien (aussi appelé Log-Poisson), le modèle Renshaw-Haberman (RH) et le modèle Age-Period-Cohort (APC).

Une approche intuitive serait de comparer simplement les changements réels de l'espérance de vie entre 2019 et 2020. Cependant, cela ne peut pas saisir pleinement les effets du COVID-19 sur la mortalité. En effet, cette stratégie ne prend pas en compte les variations intrinsèques de la mortalité d'une année sur l'autre en l'absence de COVID-19 : ce qui se serait passé en 2020 sans la pandémie. Ainsi, sans considérer les changements attendus d'une année sur l'autre en mortalité en 2020, la perte totale d'espérance de vie causée par la pandémie de COVID-19 serait probablement sous-estimée. Dans ce mémoire, nous allons alors supposer deux scénarios :

- **Scénario Baseline** (de référence) : Ce scénario représente la situation hypothétique où la pandémie n'a jamais eu lieu. Nous calibrons nos modèles en utilisant les données historiques de mortalité s'arrêtant à 2019 inclus, l'année juste avant l'apparition de la Covid. Ce scénario va nous permettre d'obtenir les projections des mortalités qui ont été attendues en l'absence de la pandémie en tenant compte des variations intrinsèques qui auraient pu se produire dans des conditions normales.
- **Scénario COVID** : Ce scénario intègre le choc de la pandémie de COVID-19. La calibration des modèles est effectuée sur les données historiques s'arrêtant à 2020, prenant donc en

compte ce choc de mortalité. Les projections obtenues avec ce scénario représentent les mortalités attendues dans un modèle intégrant le choc en année *jump-off*.

Cette démarche a déjà été utilisée dans plusieurs études de mortalité liées à la pandémie. L'idée principale est que, en comparant les résultats issus des deux scénarios, nous pouvons isoler l'effet de la pandémie sur les estimations de l'espérance de vie et évaluer la déviation par rapport aux tendances préexistantes.

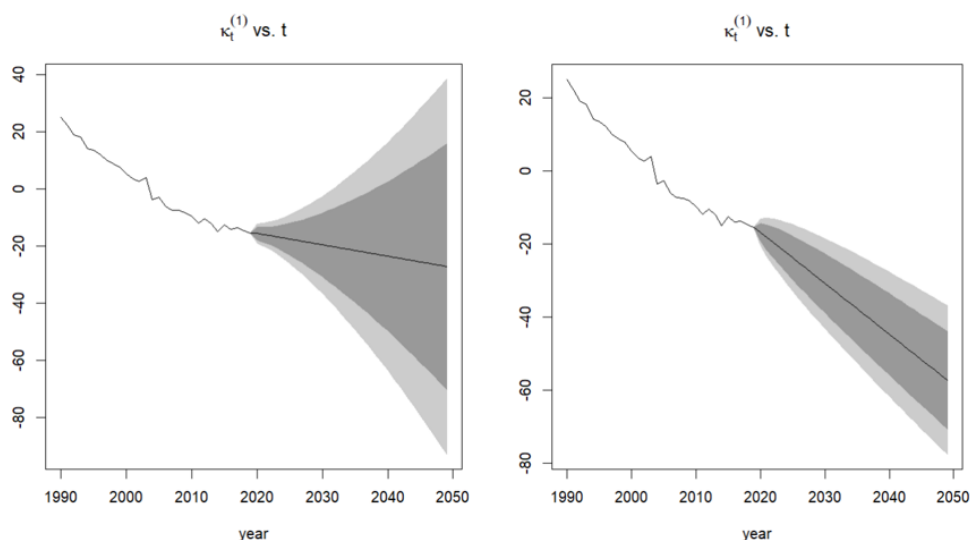
Modélisation de la mortalité

Dans un premier temps, nous avons cherché le modèle de durée optimal à utiliser pour chaque population et sous chaque scénario. L'objectif serait d'évaluer si un modèle qui était jugé pertinent pour une population donnée avant la pandémie conserve sa pertinence après l'intégration du choc initial de la COVID-19 en 2020, et d'examiner l'influence de celui-ci sur les paramètres de calibration de ces modèles.

Les critères d'information d'Akaike (AIC) et Bayésien (BIC) ont été utilisés afin d'évaluer la performance des modèles. Les résultats montrent que, dans nos 5 pays et pour les deux sexes, le modèle optimal est resté identique entre les deux scénarios. Pour les femmes, nous avons retenu le modèle RH en France, au Japon et au Royaume-Uni, et le modèle APC en Espagne et au Canada. Pour les hommes, nous avons retenu le modèle RH en France et au Canada, le modèle APC au Royaume-Uni et en Espagne, et le modèle Lee-Carter poissonien au Japon.

Cependant, en traçant les paramètres de ces modèles, nous avons pu évaluer comment ils ont été impactés entre les deux scénarios pour une même population. Nous avons trouvé que, en général, les paramètres des modèles APC et LC ont présenté un saut de la composante temporelle k_t en 2020, reflétant le choc de la Covid, mais tous les autres paramètres sont restés quasiment identiques entre les deux scénarios. En revanche, les paramètres du modèle RH ont été plus sensibles à ce choc.

Pour le choix de la méthode de projection de la composante temporelle k_t , présente dans nos modèles et permettant la projection des taux dans le temps, nous avons choisi le modèle optimal selon le critère AIC entre les modèles ARIMA et la méthode de marche aléatoire avec dérive. Il est important de noter que la méthode de projection de k_t peut avoir un impact remarquable sur la trajectoire future de la série. Par exemple, dans la figure ci-dessous, nous pouvons constater que la même série k_t diminue plus rapidement d'ici 2050 lorsque la méthode optimale de la marche aléatoire avec dérive est appliquée, comparativement à la méthode optimale ARIMA(0,2,2). Cette observation se maintient également pour le scénario Covid. Cela impactera par conséquent les mortalités projetées. Il est donc crucial de sélectionner la méthode de projection la plus adéquate.



Comparaison des trajectoires de la composante k_t projetée avec la méthode ARIMA(0,2,2) et la marche aléatoire avec dérive optimale respectivement pour le scénario Baseline pour les femmes en France

Nous avons trouvé que pour les hommes au Japon et en France, les modèles optimaux pour projeter la composante temporelle sont restés identiques entre les deux scénarios. Pour les autres cas, nous avons observé soit un changement de méthode, passant d'une marche aléatoire avec drift à un modèle ARIMA plus complexe, soit une modification des paramètres des modèles entre les deux scénarios.

Comparaison des projections

Nous avons ensuite analysé l'effet de ce choc sur les projections de mortalité, en se concentrant plus particulièrement sur l'espérance de vie résiduelle, un facteur essentiel en assurance vie notamment dans le calcul des rentes viagères. En effet, l'espérance de vie résiduelle détermine la durée estimée pendant laquelle une personne continuera de recevoir des paiements de rente, influençant ainsi directement la valorisation des engagements des compagnies d'assurance.

1) Comparaison des taux de mortalité du modèle Baseline avec les taux observés en 2020

Dans un premier temps, nous avons comparé directement les taux de mortalité réels de l'année 2020 avec ceux projetés par les modèles Baseline. L'objectif étant de voir si les taux réels du choc se situent dans la « marge d'erreur » prévue par le modèle non choqué.

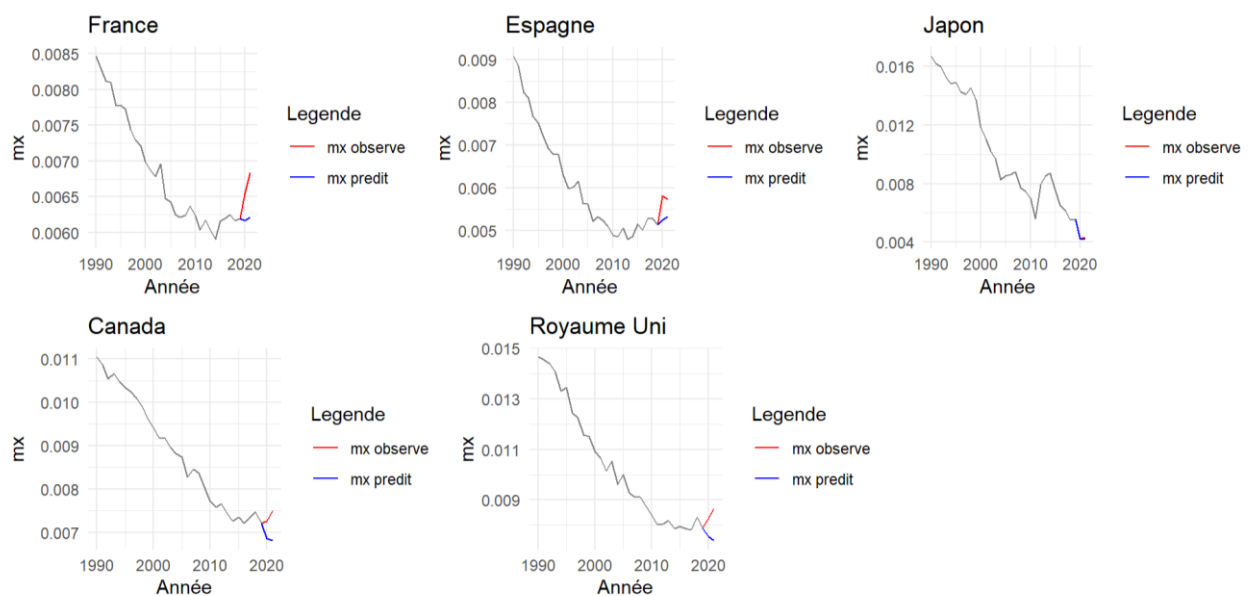
Nous avons réalisé 10 000 simulations à partir desquelles, en calculant les différents quantiles, nous avons pu obtenir des intervalles de confiance à 95 % et 80 %.

En observant, dans le tableau ci-dessous, les proportions des taux observés se situant dans les intervalles de confiance estimés, nous constatons que les résultats ne sont pas très satisfaisants. En effet, à l'exception du Japon, une faible proportion des taux observés se trouve à l'intérieur de nos IC, chez les deux sexes. Par exemple, en Espagne, chez les hommes, seul 1% des taux de mortalité réels en 2020 sont situés à l'intérieur de l'intervalle de confiance prévu par le modèle Baseline.

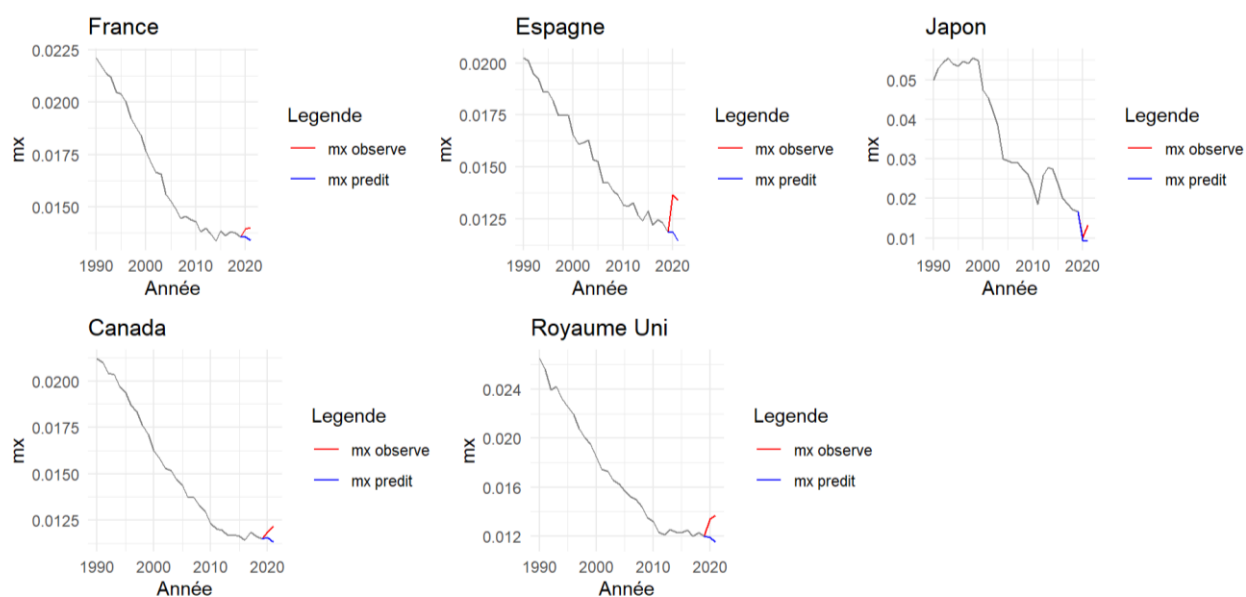
	France	Espagne	Japon	Canada	Royaume-Uni
Femmes	32% (15%)	15% (9%)	47% (17%)	14% (7%)	10% (5%)
Hommes	16% (10%)	1% (1%)	55% (44%)	10% (7%)	4% (3%)

Proportion des taux de mortalité observés en année 2020 se situant dans l'IC 95% (et 80%) du modèle Baseline

Nous avons ensuite tracé, pour chaque pays, les écarts entre les taux de mortalité observés et prédits pour les années 2020 et 2021 chez les femmes et les hommes âgés de 65 ans, correspondant à l'âge légal moyen de départ à la retraite dans la majorité des pays. Nous pouvons voir que dans la majorité des cas, les taux de mortalité observés ont été supérieurs à ceux prédits. Cela était attendu puisque le modèle a été calibré sur la période avant le choc et a ainsi prédit les taux dans un scénario "normal" sans évènement exceptionnel. En revanche, au Japon, chez les deux sexes, nous n'observons presque pas d'écart entre les valeurs prédites et celles observées pour l'année 2020, mais un écart a été observé chez les hommes en 2021.



Taux de mortalité observés VS prédits pour les femmes âgées de 65 ans



Taux de mortalité observés VS prédits pour les hommes âgés de 65 ans

2) Comparaison des espérances de vie projetées en 2021

Dans cette partie, nous allons construire les tables de mortalité à partir des taux de mortalité prédits et observés en 2021 afin de pouvoir comparer les estimations des espérances de vie.

❖ Comparaison entre les estimations des deux modèles

Une fois les espérances de vie résiduelles en 2021 calculées pour chaque âge, nous analyserons l'effet du choc survenu à l'année *jump-off* sur ces estimations en comparant celles issues du modèle Baseline avec celles issues du modèle Covid. Nous avons obtenu les écarts suivants entre les deux modèles :

Ecart Covid/Baseline	France	Espagne	Canada	Japon	Royaume-Uni
Femmes	-3,5 mois	-6,1 mois	-2,7 mois	1,2 mois	-7,9 mois
Hommes	-4,9 mois	-5,9 mois	-5,9 mois	0,1 mois	-8,4 mois

Écarts (en mois) entre les estimations d'espérance de vie résiduelles issues du modèle Covid et celles issues du modèle Baseline, tous âges confondus.

Les résultats montrent qu'en France, au Canada et au Royaume-Uni, l'écart entre les estimations des deux modèles a été, en moyenne, plus important en valeur absolue chez les hommes que chez les femmes. Par exemple, au Canada, l'utilisation du modèle Covid a conduit à une diminution moyenne de 5,9 mois de l'espérance de vie des hommes par rapport aux estimations du modèle Baseline.

De plus, nous remarquons que l'écart le plus élevé a été observé au Royaume Uni chez les deux sexes, et le moins élevé au Japon.

❖ Comparaison entre les estimations et les valeurs observées

Maintenant que nous avons vu l'impact qu'a le choix du scénario sur les estimations, nous cherchons à déterminer quel scénario parmi les deux a prédit le mieux la perte d'espérance de vie observée à l'année 2021. Pour cela, nous avons comparé les écarts par âge entre les espérances de vie estimées sous les deux scénarios et celles réellement observées.

Les résultats ont montré que le modèle Covid semble mieux refléter la réalité en 2021 que le modèle Baseline. En effet, dans la majorité des cas, les écarts entre les valeurs prédites par le modèle Baseline et les valeurs réelles sont généralement plus importants que ceux obtenus avec le modèle Covid.

Cependant, ces observations ne sont pas présentes au Japon. En effet, nous avons remarqué que chez les hommes les deux modèles présentent des écarts similaires vis-à-vis de la réalité, tandis que chez les femmes, le modèle Baseline semble même offrir des estimations plus proches de la réalité que le modèle Covid.

Afin d'obtenir une vision plus globale, nous avons calculé les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance estimés sous les deux scénarios. Nous remarquons dans les tableaux ci-dessous que, à l'exception du Japon, chez les deux sexes, les modèles Covid ont été significativement plus précis pour l'année 2021, contenant une plus grande proportion des valeurs réelles dans leurs marges d'erreur.

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	13% (8%)	5% (3%)	16% (9%)	94% (17%)	7% (2%)
Covid	82% (72%)	93% (57%)	76% (32%)	53% (0%)	94% (76%)

Les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance 95% (et 80%) chez les femmes pour l'année 2021

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	99% (33%)	24% (12%)	38% (23%)	98% (97%)	14% (8%)
Covid	96% (95%)	97% (94%)	78% (51%)	96% (39%)	99% (98%)

Les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance 95% (et 80%) chez les hommes pour l'année 2021

Applications sur un portefeuille de rentiers

Nous allons dans cette partie quantifier l'impact du choc pandémique à l'année *jump-off* sur un portefeuille de rentiers viagers. Nous cherchons à calculer les engagements de l'assureur à l'année 2021 sous les deux scénarios dans chaque pays et voir comment ceux-ci ont été modifiées.

Nous avons supposé les hypothèses suivantes dans notre étude :

- L'âge de départ à la retraite est de 65 ans dans les 5 pays.
- Le taux d'actualisation est supposé identique entre les pays et constant dans le temps. Il est égal à 3%.
- Nous utilisons l'Euro comme devise pour tous les pays.

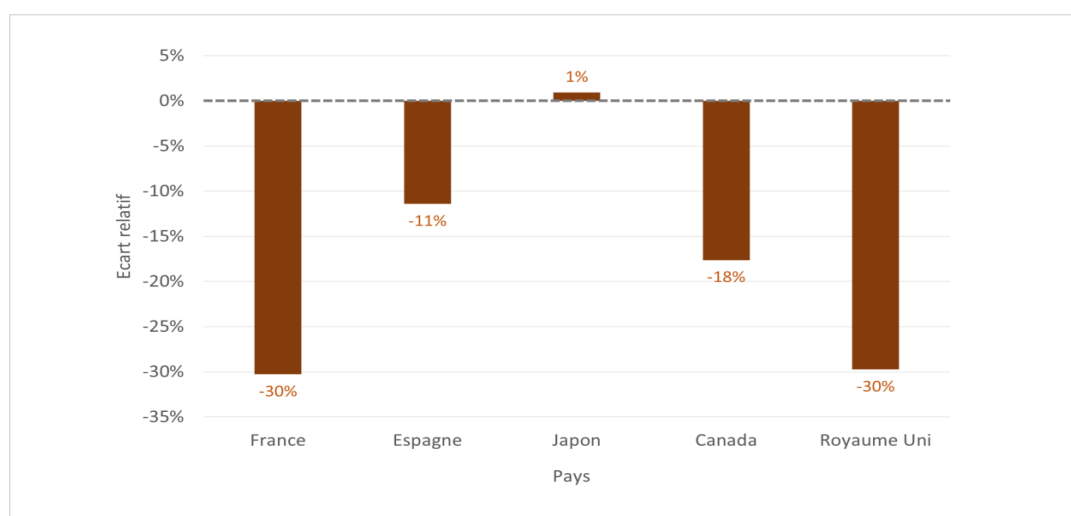
Ces hypothèses ont pour but de mieux isoler l'effet du COVID des autres facteurs externes susceptibles d'influencer notre portefeuille, mais aussi de faciliter les comparaisons entre les pays.

Pour le calcul des engagements, nous avons procédé ainsi :

- 1) Nous avons tout d'abord réalisé des projections des taux de mortalité m_x jusqu'à l'année 2105, qui correspond à l'année où l'ensemble du portefeuille aurait déjà atteint l'âge de sortie de 110 ans. Cela correspond à une matrice de dimension 96 lignes (pour les âges de 15 à 110 ans) et 85 colonnes (représentant les années de 2021 à 2105).
- 2) Ensuite, nous avons calculé les probabilités de survie p_x associées à chaque âge et chaque année projetée.
- 3) Enfin, après avoir calculé le facteur d'annuités des rentes unitaires (a_x) pour chaque individu, nous avons multiplié cette valeur par le montant contractuel de rentes afin d'obtenir le montant des engagements.

Résultats

En analysant les écarts relatifs entre les engagements totaux sur la figure ci-dessous, nous constatons que l'Espagne et la France ont été les plus impactées entre les deux scénarios. En effet, le portefeuille de ces deux pays présente un écart de -30% entre les engagements sous le scénario Baseline et ceux sous le scénario Covid. En d'autres termes, un assureur utilisant le modèle Covid verrait ses engagements diminuer de 30% par rapport au modèle Baseline.



Écarts relatifs entre les engagements totaux sous le scénario Baseline et ceux sous le scénario Covid, par pays.

Cependant, ces écarts observés semblent disproportionnés et se baser sur ces résultats pourrait être peu prudent. En effet, le modèle Covid semble, comme nous l'avons vu, pertinent pour l'année 2021, qui a été directement impactée par le choc pandémique. Il pourrait alors être intéressant et prudent de le prendre en compte pour les engagements vie de court terme, par exemple des contrats de prévoyance collective qui verse un capital en cas de décès sur une période d'un an. Toutefois, pour les engagements ou les produits de long terme, l'application directe de ce modèle pourrait sous-estimer les engagements des assureurs.

C'est pour cette raison que nous proposons une approche plus « réaliste » qui mélange à la fois les projections issues du modèle Baseline et celles issues du modèle Covid. Concrètement, en intégrant les projections COVID pour les premières années de projection, nous allons prendre en compte l'effet immédiat de la pandémie sur la mortalité. Ensuite, en revenant progressivement aux résultats du modèle Baseline, nous supposons un retour progressif à une tendance plus "normale".

Cela permettrait de mieux capter l'impact temporaire de la crise sanitaire tout en conservant une trajectoire de mortalité plus cohérente à long terme.

Approche de combinaison Baseline/Covid

Nous avons choisi une approche exponentielle pour réaliser la transition entre les projections issues du modèle Covid et celles du modèle Baseline. Cette méthode repose sur une décroissance exponentielle de l'influence du modèle COVID : l'effet du choc diminue rapidement au début et plus lentement ensuite.

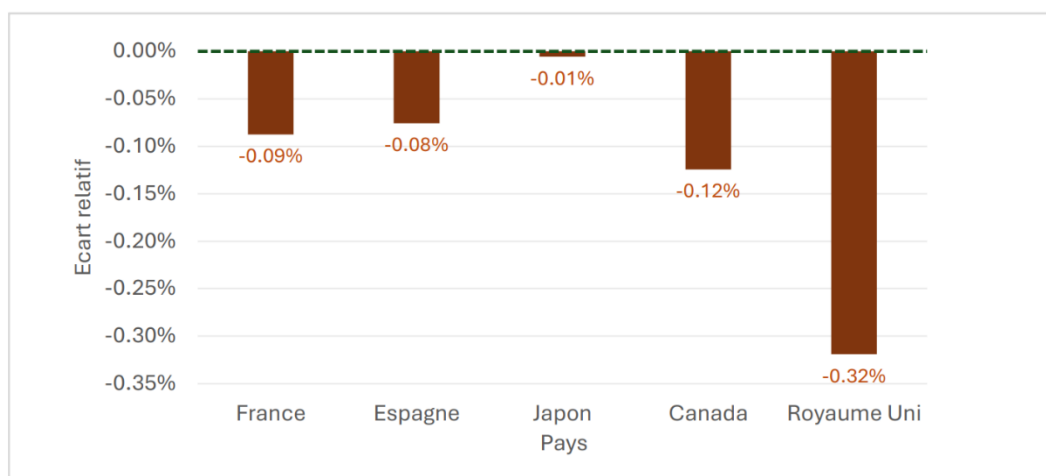
Le poids du modèle Covid à une année t est définie par :

$$r_t = \exp(-\theta \times (t - t_0))$$

Avec :

- $t_0 = 2021$
- θ un paramètre déterminant la vitesse de d'amortissement. Plus θ est grand, plus la décroissance est rapide.

Avec un paramètre de vitesse de décroissance θ égal à 0.5 qui permet de moduler la rapidité de cette diminution, et une période d'amortissement égale à 5 ans nous avons obtenu les écarts relatifs suivants :



Écarts relatifs entre les engagements totaux sous le scénario Baseline et ceux sous le scénario mixte, par pays

Ces résultats nous semblent plus réalistes et prudents pour les engagements sur le long terme. Cependant, une limite de ce modèle est le fait que le choix du paramètre θ et du nombre d'années n est crucial et pourraient influencer les résultats. Une estimation incorrecte de ce paramètre pourrait conduire à une sous-estimation ou une surestimation de la réduction du choc. C'est pour cela nous

allons à la fin de ce mémoire réaliser une étude de sensibilité des écarts relatifs face à la variation des paramètres θ et n .

Executive summary

Before the Covid-19 pandemic in 2020, general mortality trends were declining, primarily due to medical advancements and improved living conditions. The impact of this phenomenon on insurance liabilities is significant and must be considered in assessing longevity risk. Insurers must anticipate this trend to ensure they can meet their commitments to policyholders.

The pandemic introduced a sudden disruption to these trends, causing a mortality shock with an increase in death rates. The objective of this study is to analyze the impact of incorporating this pandemic shock in the last calibration year, known as the jump-off year, on the projections of residual life expectancies. The aim is not to determine whether the pandemic has permanently altered mortality trends but rather to examine how its inclusion in mortality models could influence actuarial estimates. We will compare these effects across different populations: France, Spain, the United Kingdom, Japan, and Canada, with a breakdown by gender.

Methodology and Data

We obtained historical data from the Human Mortality Database, an open-access database where mortality and population data, sourced from authorized national agencies, are compiled and standardized. We decided to focus on ages ranging from 15 to 110 years. The models studied belong to the Lee-Carter family, specifically the Poisson-based Lee-Carter model (also known as Log-Poisson), the Renshaw-Haberman model (RH), and the Age-Period-Cohort (APC) model.

An intuitive approach would be to simply compare the actual changes in life expectancy between 2019 and 2020. However, this may not fully capture the effects of Covid-19 on mortality. Indeed, this strategy does not account for the intrinsic year-to-year variations in mortality that would have occurred in the absence of the pandemic—what would have happened in 2020 without Covid-19. Without considering these expected annual fluctuations, the total loss in life expectancy caused by Covid-19 could be underestimated.

To address this, we consider two scenarios in our study:

- **Baseline Scenario:** This scenario represents a hypothetical situation where the pandemic never occurred. We calibrate our models using historical mortality data up to 2019, the year immediately preceding the emergence of Covid-19. This scenario allows us to obtain mortality projections that would have been expected in the absence of the pandemic, taking into account the natural variations that might have occurred under normal conditions.
- **Covid Scenario:** This scenario incorporates the mortality shock caused by the pandemic. Model calibration is performed using historical mortality data up to 2020, thus including the Covid-19 shock. The projections from this scenario represent mortality expectations under a model that integrates the pandemic's impact in the jump-off year.

This approach has already been used in several mortality studies related to the pandemic. The key idea is that by comparing the results from both scenarios, we can isolate the effect of the pandemic on life expectancy estimates and assess the deviation from pre-existing trends.

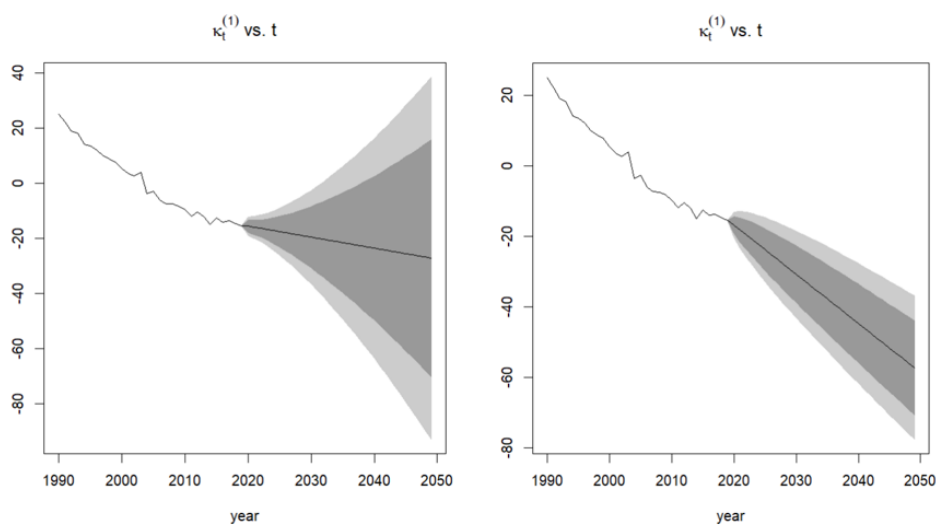
Mortality Modeling

First, we sought to determine the optimal mortality model to use for each population and under each scenario. The objective was to assess whether a model that was considered relevant for a given population before the pandemic remains suitable after integrating the initial Covid-19 shock in 2020, and to examine its influence on the model calibration parameters.

To evaluate model performance, we used Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The results show that, across all five countries and for both genders, the optimal model remained unchanged between the two scenarios. For women, we selected the RH model in France, Japan, and the United Kingdom, and the APC model in Spain and Canada. For men, we retained the RH model in France and Canada, the APC model in the United Kingdom and Spain, and the Poisson-based Lee-Carter model in Japan.

However, by analyzing the estimated parameters of these models, we evaluated how they were affected between the two scenarios for the same population. We found that, in general, the parameters of the APC and Lee-Carter models exhibited a jump in the temporal component k_t in 2020, reflecting the impact of Covid-19, while all other parameters remained almost identical between the two scenarios. In contrast, the parameters of the Renshaw-Haberman model were more sensitive to this shock.

For the selection of the projection method for the temporal component k_t , which is present in our models and allows for the projection of mortality rates over time, we chose the optimal model based on the Akaike Information Criterion (AIC) between ARIMA models and the random walk with drift method. It is important to note that the choice of k_t projection method can have a significant impact on the future trajectory of the series. For example, as shown in the figure below, the same k_t series declines more rapidly by 2050 when using the optimal random walk with drift method, compared to the optimal ARIMA(0,2,2) model. This observation is also valid for the COVID scenario. As a result, it will have an impact on projected mortality rates. Therefore, selecting the most appropriate projection method is crucial to ensure reliable mortality forecasts.



Comparison of the projected k_t component trajectories using the ARIMA(0,2,2) method and the optimal random walk with drift method for the Baseline scenario for women in France

We found that for men in Japan and France, the optimal models for projecting the temporal component remained identical between the two scenarios. However, in other cases, we observed either a change in the projection method, transitioning from a random walk with drift to a more complex ARIMA model, or a modification of the model parameters between the two scenarios.

Comparison of Projections

We then analyzed the impact of this shock on mortality projections, focusing particularly on residual life expectancy, which is a key factor in life insurance, especially for the valuation of annuity payments. Residual life expectancy determines the estimated duration during which an individual will continue to receive annuity payments, directly influencing the valuation of insurance liabilities.

1) Comparison of Baseline Mortality Rates with Observed 2020 Mortality Rates

First, we directly compared the actual mortality rates for 2020 with the rates projected by the Baseline models. The objective was to determine whether the observed rates of the initial shock fell within the expected margin of error of the “unchocked” model.

To achieve this, we performed 10,000 simulations, from which we calculated different quantiles, allowing us to estimate confidence intervals (CIs) at 95% and 80%.

As shown in the table below, the proportion of observed mortality rates falling within the estimated confidence intervals is relatively low, except for Japan. In both genders, we found that in most cases, the actual mortality rates for 2020 were significantly higher than those projected by the Baseline model. For example, in Spain, only 1% of the observed mortality rates for men in 2020 were within the confidence interval predicted by the Baseline model.

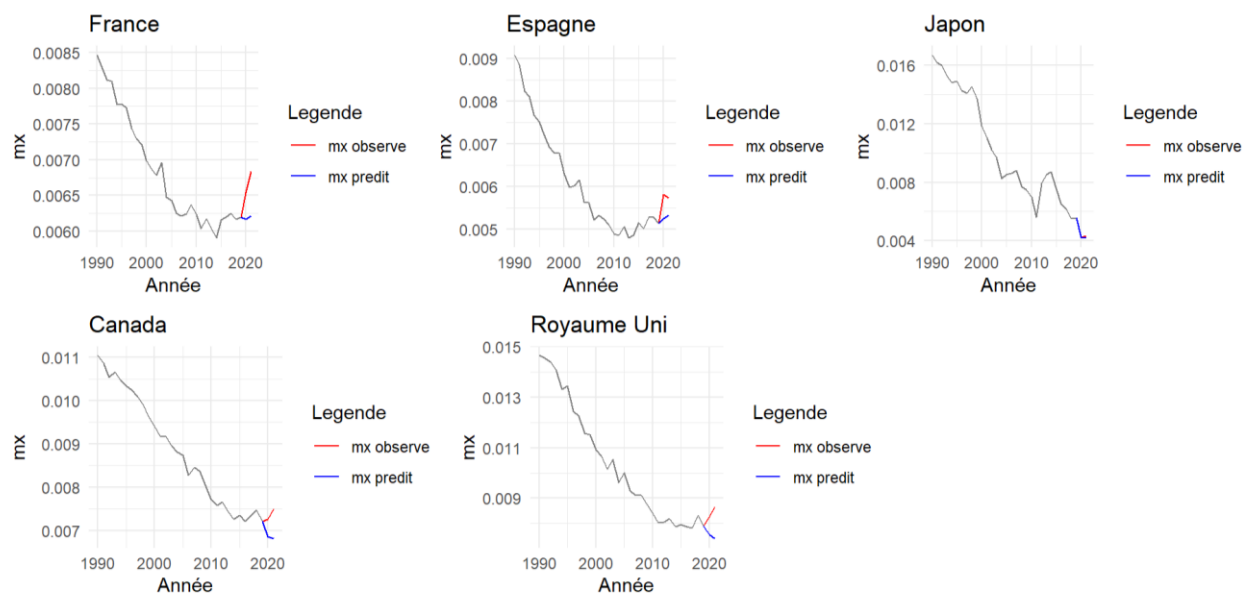
	France	Spain	Japan	Canada	UK
Women	32% (15%)	15% (9%)	47% (17%)	14% (7%)	10% (5%)
Men	16% (10%)	1% (1%)	55% (44%)	10% (7%)	4% (3%)

Proportion of the observed mortality rates in 2020 falling within the 95% (and 80%) confidence interval of the Baseline model.

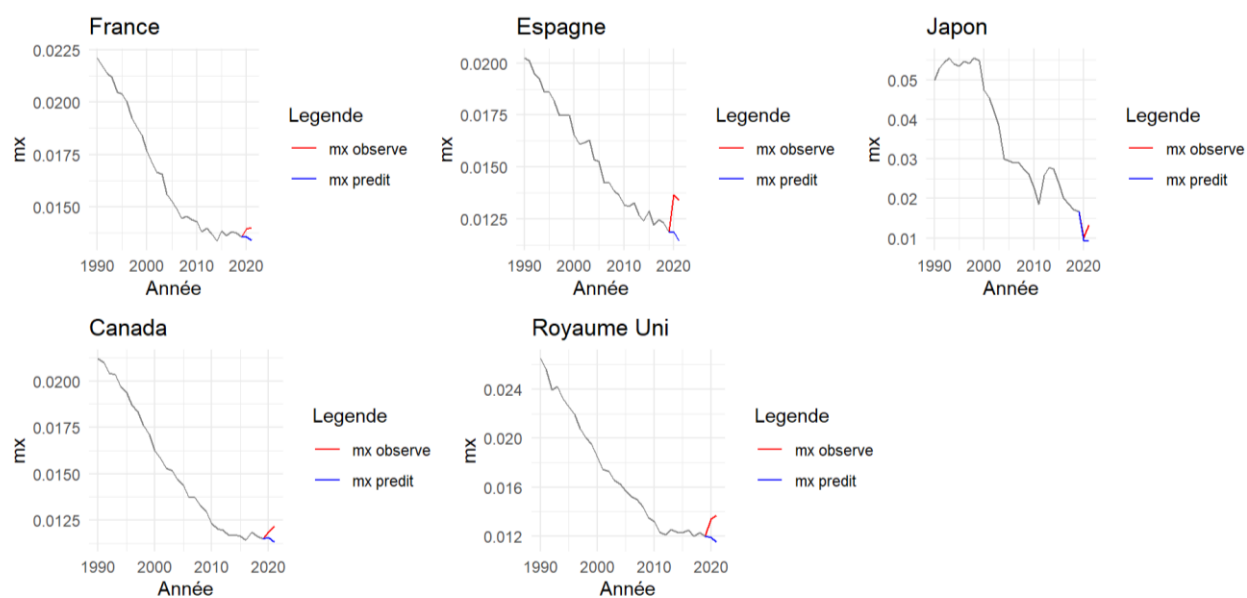
We then plotted, for each country, the differences between the observed and predicted mortality rates for the years 2020 and 2021 for women and men aged 65, which corresponds to the average legal retirement age in most countries.

As shown in the below graphs, in most cases, the observed mortality rates were higher than the predicted ones. This was expected since the model was calibrated on the pre-pandemic period, projecting mortality rates under a “normal” scenario without any exceptional events.

However, in Japan, for both genders, we observed almost no difference between the predicted and observed values for 2020. Nevertheless, a discrepancy was noted for men in 2021.



Observed vs Predicted Mortality Rates for Women Aged 65



Observed vs Predicted Mortality Rates for Men Aged 65

2) Comparison of Projected Life Expectancies in 2021

In this section, we construct mortality tables based on the predicted and observed mortality rates for year 2021 in order to compare the estimated life expectancies.

❖ Comparison Between the Two Model Estimates

Once the residual life expectancies for 2021 are calculated for each age, we analyze the impact of the shock at the jump-off year on these estimates by comparing the results from the Baseline model with those from the Covid model.

The following differences were observed between the two models:

Covid/Baseline gaps	France	Spain	Canada	Japan	UK
Women	-3,5 months	-6,1 months	-2,7 months	1,2 months	-7,9 months
Men	-4,9 months	-5,9 months	-5,9 months	0,1 months	-8,4 months

Differences (in months) between the residual life expectancy estimates from the Covid model and those from the Baseline model, across all ages.

The results show that in France, Canada, and the United Kingdom, the difference between the estimates of the two models was, on average, larger in absolute value for men than for women. For example, in Canada, using the Covid model resulted in an average decrease of 5,9 months in men's life expectancy compared to the estimates from the Baseline model.

Furthermore, the largest gap was observed in the United Kingdom for both genders, while the smallest gap was recorded in Japan.

❖ Comparison Between Estimates and Observed Values

Now that we have analyzed the impact of scenario selection on life expectancy estimates, we seek to determine which scenario better predicted the actual loss in life expectancy observed in 2021. To do so, we compared the differences by age between estimated life expectancies under both scenarios and the actual observed values.

In general, results show that the Covid model appears to better reflect reality in 2021 than the Baseline model. In most cases, the differences between the values predicted by the Baseline model and the actual observed values are generally larger than those obtained with the Covid model.

However, this pattern does not hold in Japan. Among men, both models exhibit similar discrepancies compared to actual values, whereas for women, the Baseline model even appears to provide closer estimates to reality than the Covid model.

To gain a more comprehensive perspective, we calculated the proportion of actual values that fall within the confidence intervals estimated under both scenarios. As shown in the tables below, except for Japan, the Covid models were significantly more accurate for 2021, capturing a higher proportion of actual values within their error margins for both genders.

	France	Spain	Canada	Japan	United-Kingdom
Baseline	13% (8%)	5% (3%)	16% (9%)	94% (17%)	7% (2%)
Covid	82% (72%)	93% (57%)	76% (32%)	53% (0%)	94% (76%)

The proportions of the actual values falling within the 95% (and 80%) confidence intervals for women in the year 2021.

	France	Spain	Canada	Japan	United-Kingdom
Baseline	99% (33%)	24% (12%)	38% (23%)	98% (97%)	14% (8%)
Covid	96% (95%)	97% (94%)	78% (51%)	96% (39%)	99% (98%)

The proportions of the actual values falling within the 95% (and 80%) confidence intervals for men in the year 2021.

Applications on an Annuity Portfolio

In this section, we quantify the impact of the pandemic shock at the jump-off year on an annuity portfolio. Our objective is to calculate the insurer's liabilities in 2021 under both scenarios for each country and assess how they have been affected.

We assume the following hypotheses in our study:

- The retirement age is set at 65 years across all five countries.
- The discount rate is assumed to be identical between countries and constant over time, set at 3%.
- The Euro is used as the currency for all countries.

These assumptions aim to better isolate the effect of Covid-19 from other external factors that could influence the portfolio while also facilitating comparisons between countries.

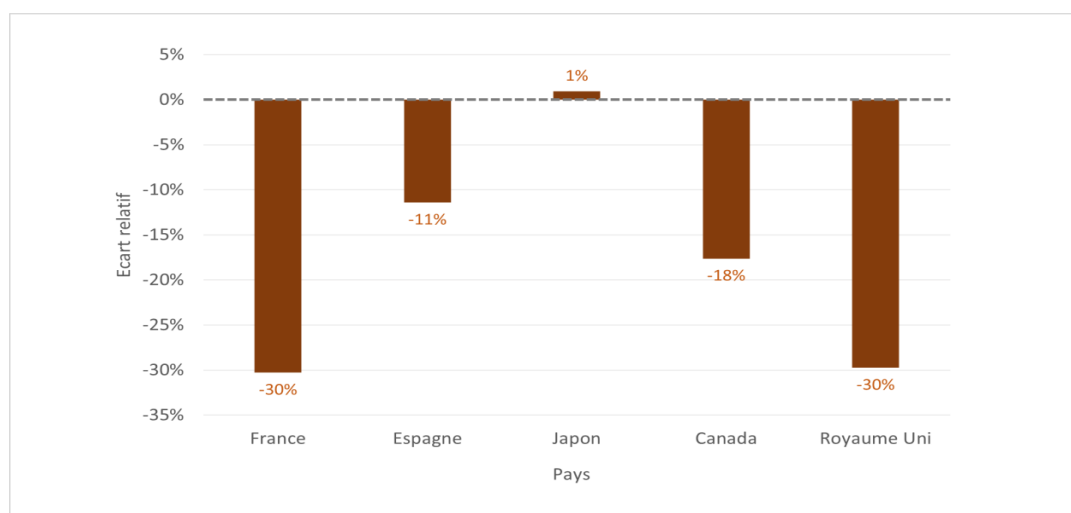
For the calculation of liabilities, we proceeded as follows:

- 1) We projected mortality rates m_x up to 2105, the year by which all individuals in the portfolio would have reached the maximum age of 110 years. This results in a matrix of 96 rows (for ages 15 to 110) and 85 columns (representing the years 2021 to 2105).
- 2) We then computed age- and year-specific survival probabilities p_x from the projected mortality rates.
- 3) Finally, after calculating the annuity factor for unitary annuities (a_x) for each individual, we multiplied this value by the contractual annuity amount to obtain the liabilities.

Results

By analyzing the relative differences between the total commitments in the figure below, we observe that Spain and France have been the most impacted between the two scenarios. Indeed, the portfolio of these two countries shows a gap of -30% between the commitments under the Baseline scenario

and those under the Covid scenario. In other words, an insurer using the Covid model would see its commitments decrease by 30% compared to the Baseline model.



Relative differences between total commitments under the Baseline scenario and those under the Covid scenario, by country.

However, these observed differences seem disproportionate, and basing decisions on these results could be unwise. The Covid model seems, as we have seen, relevant for the year 2021, which was directly impacted by the pandemic shock. It might then be interesting and prudent to consider this model for short-term life commitments, such as Group Life contracts that pay a capital in the event of death over a one-year period. However, for long-term commitments or products, directly applying this model could underestimate insurers' commitments. This is why we suggest a more "realistic" approach that combines both the projections from the Baseline model and those from the Covid model. Specifically, by integrating the Covid projections for the first years of projection, we will account for the immediate effect of the pandemic on mortality. Then, by gradually returning to the results of the Baseline model, we assume a gradual return to a more "normal" trend. This would allow for a better capture of the temporary impact of the health crisis while maintaining a more consistent long-term mortality trajectory.

Approach combining Baseline/Covid projections

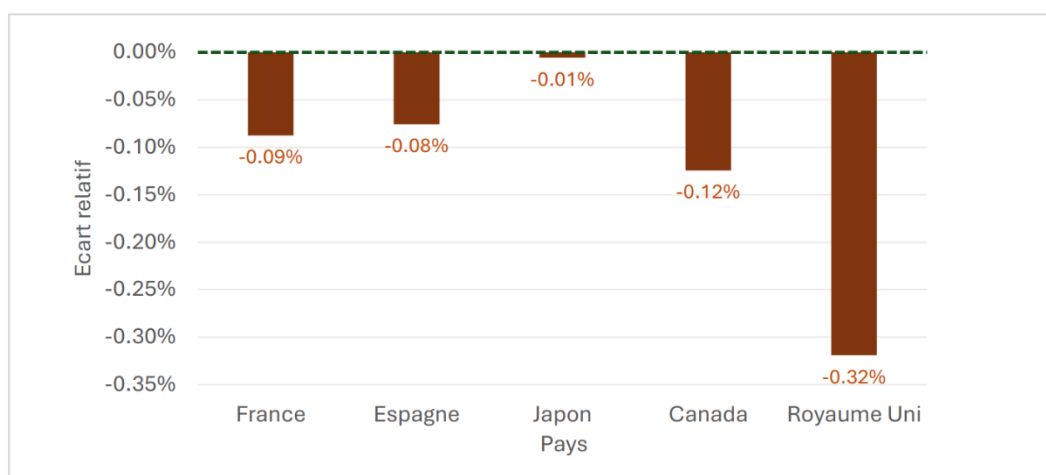
We have chosen an exponential approach to transition between the projections from the Covid model and those from the Baseline model. This method relies on an exponential decay of the influence of the Covid model: the shock's effect diminishes quickly at first and more slowly thereafter. The weight of the Covid model at year t is defined by:

$$r_t = \exp(-\theta \times (t - t_0))$$

With :

- $t_0 = 2021$
- θ a parameter determining the rate of decay. The larger θ is, the faster the decay.

With a decay rate parameter θ equal to 0,5, which allows for modulation of the speed of this decrease, and a smoothing period of 5 years, we obtained the following relative differences:



Relative differences between total commitments under the Baseline scenario and those under the mixed approach, by country.

These results seem more realistic and prudent for long-term commitments. However, one limitation of this model is that the choice of the parameter θ and the number of years n is crucial and could influence the results. An incorrect estimation of this parameter could lead to an underestimation or overestimation of the shock's reduction. This is why, at the end of this report, we will conduct a sensitivity analysis of the relative differences in response to variations in the parameters θ and n .

Table des matières

Introduction	1
Partie I : Le risque de mortalité et la Covid-19	2
1) Mortalité et risque de longévité : Contexte	2
2) Quelques rappels de notions de mortalité	4
3) Revue des modèles de durée existants	11
1.1) Les premiers modèles de mortalité	12
1.1.1) Le modèle de Brass	12
1.1.2) Le modèle de Cox.....	13
1.2) Famille de modèles « Lee-Carter »	16
1.2.1) Le modèle de Lee-Carter	16
1.2.2) Le modèle de Renshaw-Haberman	24
1.2.3) Le modèle Age-Period-Cohort.....	26
1.3) Famille de modèles « Cairns-Blake-Dowd »	27
4) La pandémie de Covid-19 : quels enjeux sur la mortalité ?	29
Partie II : Modélisation de la mortalité	32
1) Les données HMD.....	32
2) Présentation des pays étudiés	33
3) Calibration des modèles.....	35
3.1) Calibration des modèles pour la France	35
3.1.1) Résultat de calibration des modèles	36
3.1.2) Analyse des résidus.....	44
3.1.3) Sélection des modèles	49
3.1.4) Choix de la méthode de projection de la composante temporelle	51
3.2) Calibration et résultats de modélisation pour chaque pays	53
3.2.1) Sélection des modèles par pays.....	53
3.2.2) Résultats de calibration des modèles choisis par pays	56
Partie III : Projections et applications actuarielles	63
1) Comparaison des taux de mortalité du modèle Baseline avec les taux observés en 2020	63
2) Comparaison des espérances de vie projetées en 2021.....	66

2.1) Comparaison entre les estimations des deux modèles	66
2.2) Comparaison entre les estimations et les valeurs observées	69
3) Applications sur un portefeuille de rentiers	73
3.1) Présentation de notre portefeuille	73
3.2) Calcul des engagements	75
3.3) Résultats	76
3.4) Approche de combinaison Baseline/Covid	78
Partie IV : Limites de l'étude	82
Conclusion	84
Bibliographie.....	85

Introduction

L'émergence de la pandémie de COVID-19 a touché la mortalité des populations à l'échelle mondiale. Depuis son apparition fin 2019, le virus a entraîné des millions de décès. Les mesures de confinement, les restrictions de voyage, et la réorientation des ressources de santé vers la lutte contre la pandémie ont eu des effets sur la survie, au-delà des décès directement liés à la maladie.

Une tendance générale vers la baisse de la mortalité était observée dans le monde avant la pandémie, portée par les progrès médicaux et technologiques. Toutefois, son émergence a introduit un choc de mortalité en 2020 par rapport aux années précédentes, ce qui pourrait poser un enjeu pour les assureurs vie. En effet, ces derniers, afin d'évaluer les engagements financiers de leurs portefeuilles, utilisent des modèles de durée en se basant sur les taux historiques de mortalité. Un événement comme la Covid-19 pourrait donc affecter la calibration de ces modèles, influençant ainsi leurs projections et leurs engagements.

L'objectif de ce mémoire est d'étudier l'effet de ce choc pandémique en année *jump-off* sur les estimations de mortalité. Plus précisément, nous cherchons à analyser dans quelle mesure la prise en compte de la Covid-19 en dernière année de calibration pourrait influencer les estimations de l'espérance de vie résiduelle des femmes et des hommes dans 5 pays : la France, l'Espagne, le Royaume-Uni, le Japon et le Canada. Il ne s'agit pas de déterminer si la pandémie a réellement modifié ou pas les tendances de mortalité, mais plutôt d'examiner comment sa prise en compte dans les modèles pourraient influencer les estimations actuarielles et de comparer ces effets entre ces différentes populations.

L'étude vise dans un premier temps à explorer comment ce choc en année *jump-off* pourrait affecter la calibration des modèles de durée utilisés pour analyser la survie. Plus spécifiquement, nous allons évaluer si les modèles qui étaient jugés pertinents pour une population donnée avant la pandémie conservent leur pertinence après l'intégration de la COVID-19, et examiner l'influence de celle-ci sur les paramètres de calibration de ces modèles. Ensuite, nous allons analyser l'effet de ce choc sur les projections de mortalité, en se concentrant plus particulièrement sur l'espérance de vie résiduelle, un facteur essentiel en assurance vie notamment dans le calcul des rentes viagères. En effet, l'espérance de vie résiduelle détermine la durée estimée pendant laquelle une personne continuera de recevoir des paiements de rente, influençant ainsi directement la valorisation des engagements des compagnies d'assurance. Enfin, nous allons quantifier cet effet sur un portefeuille de rentiers et proposer une approche mixte permettant de prendre temporairement en compte le choc pandémique avant de revenir vers une tendance « normale ».

Partie I : Le risque de mortalité et la Covid-19

1) Mortalité et risque de longévité : Contexte

Le contexte de mortalité concerne les variations et les tendances des taux de décès dans une population donnée, qui peuvent être influencées par des facteurs comme les avancées médicales, les changements socio-économiques, de mode de vie, les pandémies, ou d'autres événements tels que les guerres ou les catastrophes naturelles¹. L'allongement de la vie humaine est un phénomène de plus en plus observé ces dernières années. En effet, l'espérance de vie mondiale à la naissance a augmenté de plus de 6 ans entre 2000 et 2019, passant de 66,8 ans en 2000 à 73,1 ans en 2019². Cela résulte principalement de la baisse de la mortalité infantile et également de l'allongement de la durée de vie des personnes âgées.

Nous pouvons observer cette tendance générale vers la hausse sur la figure 1. Par exemple, un homme né en 1965 avait en moyenne une espérance de vie de 52 ans, tandis qu'un homme né en 2019 aura une espérance de vie de 70,1 ans. Cela représente un gain de 18 ans en espérance de vie depuis 1965. On observe également un gain moyen de 19 ans pour les femmes.

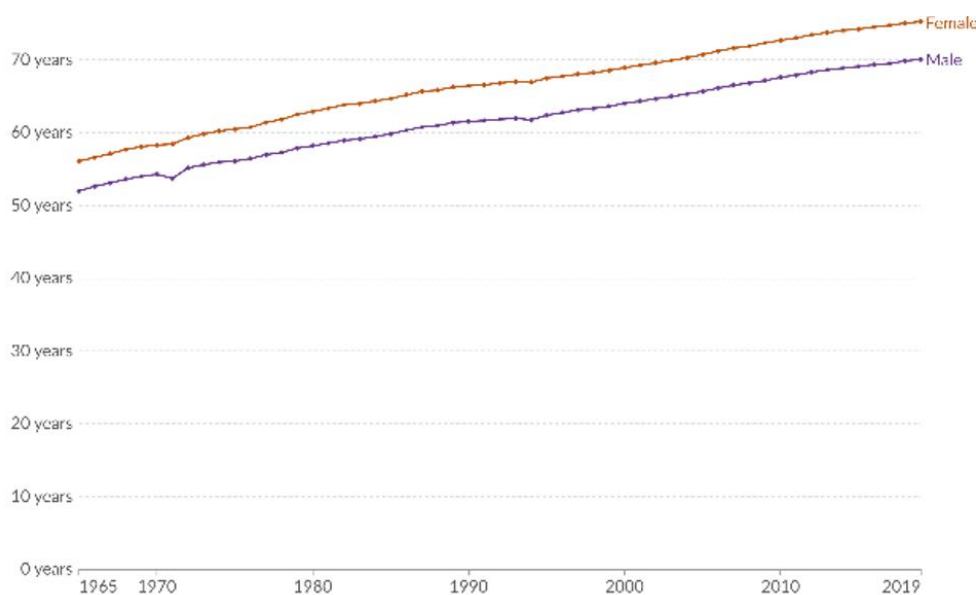


Figure 1 : L'évolution de l'espérance de vie chez les femmes et chez les hommes dans le monde entre 1965 et 2019

De plus, il est important de comprendre qu'il est nécessaire de distinguer les hommes des femmes dans les études de longévité. En effet, les tendances de mortalité diffèrent considérablement entre les deux sexes, en raison de facteurs biologiques, comportementaux, sociaux et économiques. En règle générale, les femmes ont une espérance de vie plus élevée que les hommes, avec en moyenne 5 ans

¹ Source : "The future of life expectancy : Forecasting long-term mortality improvement trends for insurance", [Swiss Re Institute](#)

² Source : [World Health Organisation](#)

de plus dans le monde en 2023³. Cette différence est principalement due à une combinaison de facteurs tels que la biologie (les femmes ont tendance à avoir un système immunitaire plus robuste), les comportements (les hommes adoptent généralement des comportements plus à risque) et l'accès aux soins de santé.⁴ C'est pour cela que nous ne ferons pas exception à la règle en proposant une analyse différenciée en fonction du sexe afin de prendre en compte les disparités entre eux.

Cependant, cet allongement de vie humaine n'est pas sans conséquences sur le marché de l'assurance : on parle du risque de longévité. Il se définit par la possibilité que les individus vivent plus longtemps que ce qui était initialement anticipé. Ce phénomène, bien que positif sur le plan humain, peut poser des défis financiers significatifs, notamment pour les assureurs vie. En effet, l'activité de l'assurance vie s'axe principalement autour de la durée de vie, et donc autour de la mortalité des assurés. Cependant, le risque de longévité représente un défi pour les assureurs puisqu'il est particulièrement complexe à maîtriser. En effet, en raison de son évolution rapide et de sa dépendance à plusieurs disciplines comme la médecine, la démographie et la biologie, ce risque peut faire face à une grande incertitude concernant les tendances d'évolutions futures. De plus, ses conséquences ne sont notables que sur le long terme, et peuvent impacter la santé financière des institutions.

Prenons l'exemple des produits de rentes viagères versées à partir de la retraite, qui garantissent donc un revenu jusqu'au décès. Pour pouvoir faire face à leurs engagements, les assureurs doivent calculer des provisions qui dépendent directement des estimations de l'espérance de vie résiduelle des rentiers : Si l'espérance de vie résiduelle à l'âge de la retraite augmente, cela implique un versement des fonds pendant une durée plus longue que celle initialement prévue dans le calcul des provisions, nécessitant ainsi plus de ressources financières que ce qui avait été estimé. Par ailleurs, Scor a prédit que d'ici à 2050, du fait des avancées de la médecine et de l'accès aux soins, le temps passé en retraite sera presque deux fois plus long que celui qui avait été prévu dans les années 1950. Ainsi, il est crucial pour l'assureur de bien analyser les tendances de mortalité afin d'estimer correctement les espérances de vie de ses rentiers et s'assurer que les provisions constituées sont suffisantes pour couvrir ses engagements.

³ Source : United Nations. World Population Prospects (2024)

⁴ Source : Gicquaud, N. « Espérance de vie : avantage aux femmes ». [INSEE](#)

2) Quelques rappels de notions de mortalité

Nous allons dans cette partie rappeler quelques notations et notions actuarielles utilisées en assurance vie et qui nous seront utiles pour la suite du mémoire.

Notations actuarielles

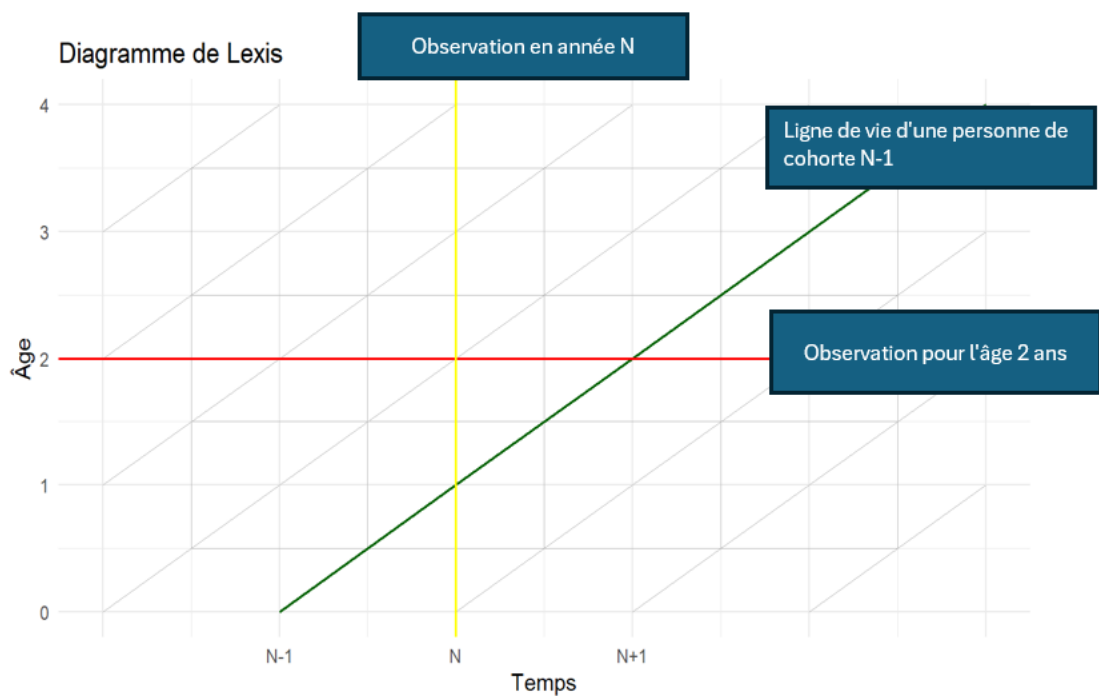
Tables de mortalité

Les tables de mortalité sont des outils essentiels en actuariat et en démographie, permettant d'estimer les probabilités de décès et de survie à différents âges pour une population donnée. Avant de présenter les éléments présents dans celles-ci, il est important de comprendre les différents types de tables qui existent. Nous en distinguons deux :

- 1) **Les tables de mortalité du moment** représentent les mortalités observées à un moment précis pour différentes tranches d'âges. Elles sont construites à partir du nombre de décès par âge sur cette période ainsi que l'exposition correspondante, c'est-à-dire le nombre de personnes encore en vie à chaque âge et susceptibles de décéder au cours de la période étudiée.
- 2) **Les tables de mortalité générationnelles ou prospectives**, contrairement aux tables du moment, tiennent compte de l'évolution de la mortalité au fil du temps. Elles sont construites à partir du nombre de décès ainsi que l'exposition en fonction de l'âge et de l'année de naissance. Ces tables suivent les cohortes de naissance spécifiques et projettent les taux de mortalité en tenant compte des tendances observées et des anticipations des progrès médicaux, sociaux et environnementaux. Prenant en compte l'évolution au fil du temps ainsi que l'effet de cohorte que nous allons détailler plus loin dans le mémoire, ces tables sont bien adaptées à l'étude du risque de longévité.

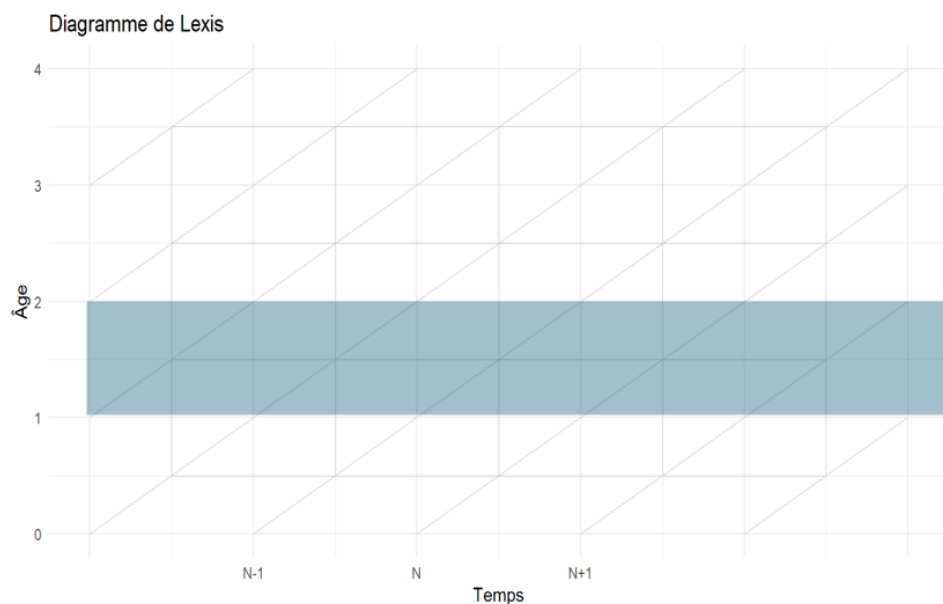
Toutefois l'inconvénient avec ces tables prospectives est que l'on doit attendre l'extinction de toute la génération pour pouvoir les construire. Le besoin des compagnies d'assurance étant immédiat pour tarifier leurs produits, nous devons quand même avoir des tables pour une génération « récente » sans avoir à attendre 40 ou 50 ans le décès de la dernière personne de cet échantillon. C'est pour cela que nous utilisons des techniques de projection de la mortalité en se basant sur des modèles qui feront l'objet de notre étude dans la partie II.

Afin de mieux visualiser cette notion de table générationnelle, nous pouvons présenter le diagramme de Lexis. Il s'agit d'un outil graphique qui permet d'étudier l'évolution de la mortalité en fonction de l'âge et du temps comme ci-dessous :

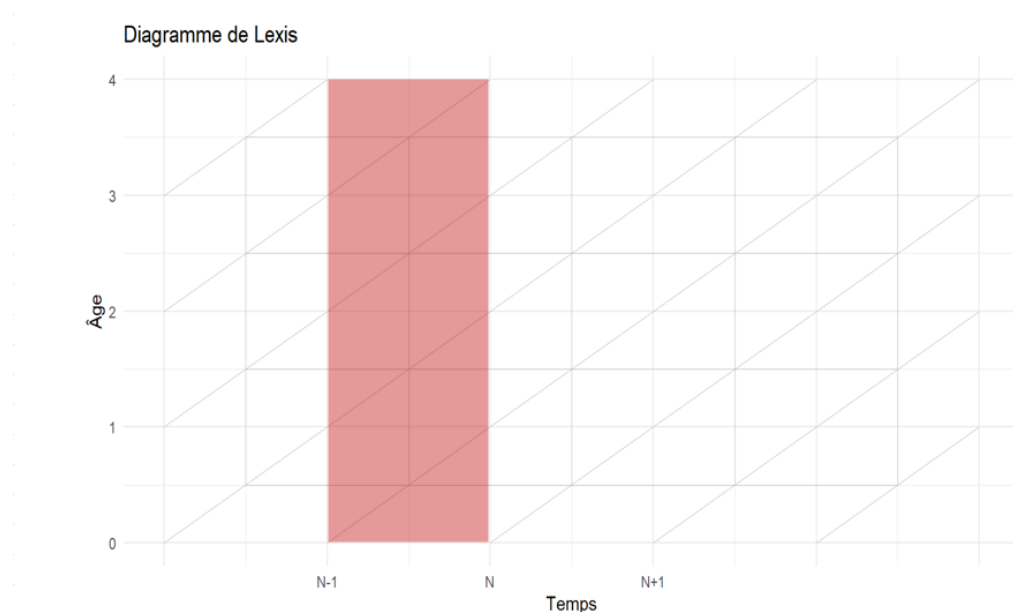


Avec en axe des abscisses le temps, ou bien l'année, et en axe des ordonnées l'âge. Les lignes en diagonales représentent la ligne de vie pour d'une génération, ou bien d'une « cohorte ».

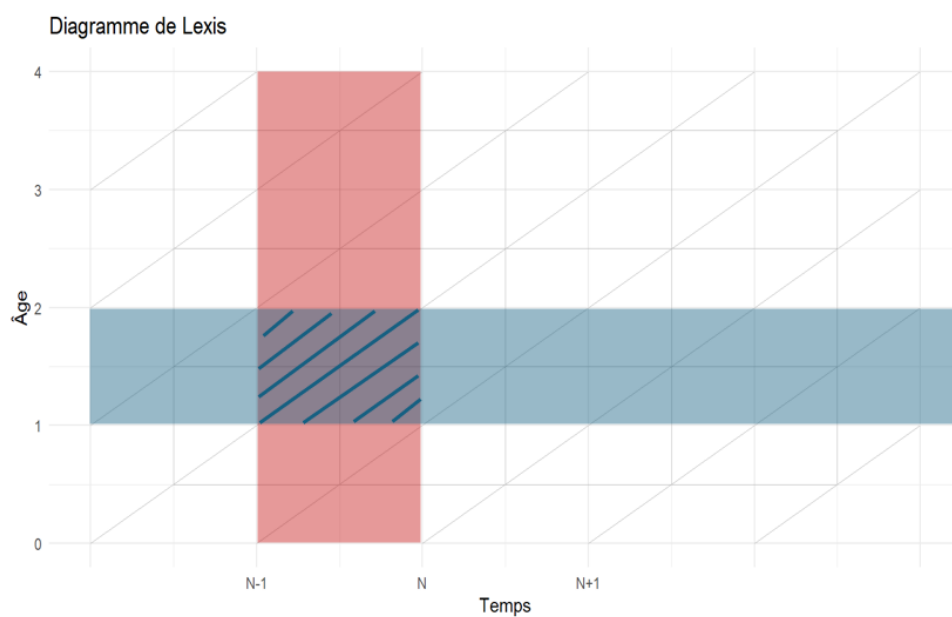
Par exemple, pour regarder les données pour un âge précis, il suffit de lire la ligne de cet âge sur le diagramme comme sur la figure ci-dessous :



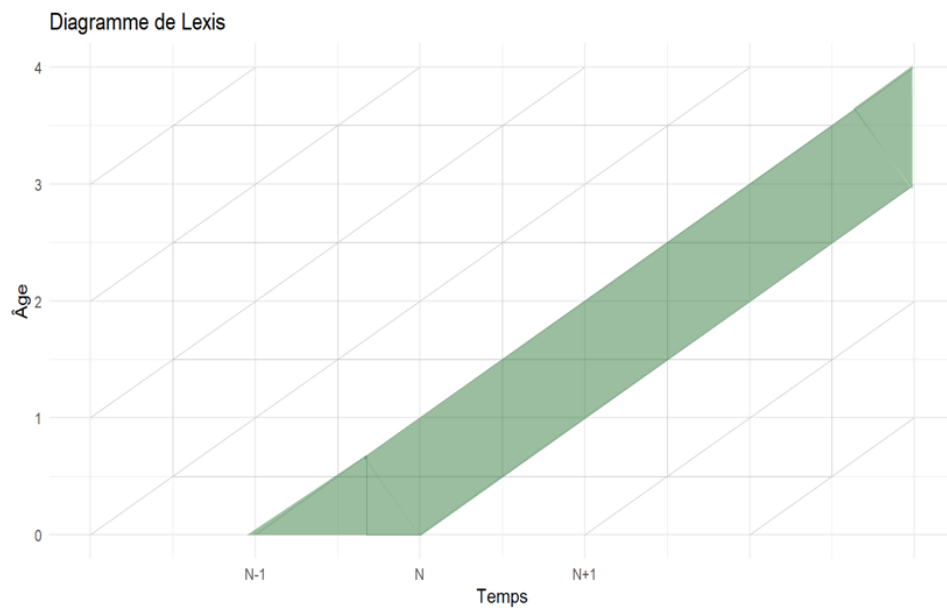
Pour se focaliser sur une année précise, il suffit de prendre la colonne de cette année comme ci-dessous :



Et pour regarder les données pour un âge précis et une année précise, il suffit de regarder l'intersection des deux :



En revanche, si nous souhaitons analyser une génération précise, il faut regarder la ligne oblique correspondante à l'année de naissance de la génération voulue :



Les tables de générationnelles sont alors obtenues à partir des diagonales de ce diagramme.

Nous reviendrons plus en détails dans la suite du mémoire sur cette notion de cohorte, qui est essentielle pour l'étude de mortalité et nous expliquerons pourquoi il est important de la prendre en compte dans notre étude.

Les notations clés des tables

On retrouve dans toutes les tables de mortalité les éléments suivants :

- l_x : Le nombre de survivants à l'âge x dans une population initiale, souvent standardisée à 100 000 à l'âge 0.
- d_x : Le nombre de décès entre les âges x et $x + 1$, calculé comme $d_x = l_x - l_{x+1}$

Ces l_x sont la base permettant de fournir les autres éléments clés de mortalité telles que les probabilités de survie, de décès ainsi que l'espérance de vie.

Probabilités de survie et de décès (les quotients de mortalité)

Soit T_x une variable aléatoire continue à valeurs dans $[0 ; +\infty[$, représentant la durée de vie résiduelle d'un individu ayant l'âge x . On peut alors définir :

- ❖ ${}_tq_x$: la probabilité de mortalité à l'âge x pour t années, c'est-à-dire la probabilité qu'une personne âgée de x ans décède avant d'atteindre $x+t$ ans.

En d'autres termes :

$${}_tq_x = P(T_x \leq t) = P(T \leq t + x \mid T > x)$$

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

- ❖ ${}_tp_x$: la probabilité de survie à l'âge x pour t années, définie comme ${}_tp_x = 1 - {}_tq_x$, c'est-à-dire la probabilité qu'une personne âgée de x ans survive jusqu'à l'âge de $x+t$ ans. Nous avons :

$${}_tp_x = P(T_x > t) = P(T > t + x \mid T > x)$$

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

Par convention, lorsque $t = 1$, c'est-à-dire lorsque l'on cherche la probabilité qu'une personne âgée de x ans décède ou survive dans l'année, on utilise les notations q_x et p_x .

Force de mortalité et taux de mortalité

- ❖ $\mu_{x,t}$: Elle représente la force de mortalité, c'est-à-dire le taux de mortalité instantané à un âge x à l'année t .

La relation entre la force de mortalité et la probabilité de survie est donnée par :

$${}_tp_x = \exp\left(-\int_{s=0}^{s=t} \mu_{x+s} ds\right)$$

— $m_{x,t}$: Elle représente le taux de mortalité à l'année t pour la population âgée de x ans. Il s'agit du rapport entre le nombre de décès et l'effectif moyen sur la période.

$$m_{x,t} = \frac{d_{x,t}}{E_{x,t}}$$

Avec $d_{x,t}$ le nombre de décès à l'année t pour la population âgée de x ans et $E_{x,t}$ l'effectif moyen de cette population sur la période t .

Espérance de vie résiduelle

L'espérance de vie résiduelle est une notion clé en analyse de mortalité. Il s'agit de la moyenne des années de vie supplémentaires qu'une personne peut s'attendre à vivre à partir d'un certain âge. L'espérance de vie résiduelle à l'âge x pourrait être définie par :

$$e_x = \frac{l_{x+1} + l_{x+2} + \dots + l_{\omega}}{l_x}$$

Avec ω l'âge limite de vie envisageable.

Fonction de survie et du hasard

Revenons à la variable T , nous nous intéressons maintenant à la fonction de survie $S(t)$ qui représente la probabilité d'avoir survécu au-delà d'un instant t . Son expression est liée à la fonction de répartition $F(t)$ de T par :

$$S(t) = P(T > t) = 1 - F(t)$$

La fonction $S(t)$ est toujours décroissante. À la date d'origine ($t = 0$), toute la population a « survécu » puisque personne n'a connu l'évènement, $S(0)$ est alors égale à 1. Plus le temps passe, plus le nombre d'individus qui ont connu l'évènement augmente et la valeur de $S(t)$ diminue.

Une autre fonction est également essentielle pour l'analyse des durées de vie, il s'agit de la fonction du hasard $h(t)$. En effet, cette fonction évalue le risque que l'évènement survienne à un instant précis, sachant qu'il n'a jamais été observé auparavant. Elle peut être définie par :

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

Avec $f(t)$ la densité de T définie par :

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt)}{dt}$$

Notions de base

Modèles stochastiques et déterministes de mortalité

Il existe deux types de modèles selon que l'on prenne en compte ou non l'incertitude des changements futurs de la mortalité : le modèle de mortalité déterministe (modèle de mortalité statique) et le modèle de mortalité aléatoire (modèle de mortalité stochastique).

Le modèle de mortalité déterministe suppose que les taux de mortalité évoluent de manière prévisible et sans variations aléatoires. En effet, les projections de mortalité sont basées sur des tendances historiques et des hypothèses fixes, ce qui simplifie l'analyse et la compréhension des résultats, mais ne tient pas compte des incertitudes futures et des variations potentielles. Par exemple, le modèle de Brass, que nous présenterons plus tard, peut être considéré comme un modèle déterministe.

En revanche, le modèle stochastique prend en compte l'incertitude et la variabilité dans ses projections. Il utilise des processus aléatoires pour simuler différentes trajectoires possibles des taux de mortalité futurs, ce qui permet de prendre en compte les risques et les imprévus. C'est ce type de modèle qui va nous intéresser dans ce mémoire.

Rentes viagères et risque de longévité

En assurance vie, les rentes viagères constituent un engagement majeur pour les assureurs. Il s'agit d'un contrat permettant au rentier de recevoir un revenu régulier tant qu'il est en vie. Cet engagement de l'assureur est en contrepartie d'un capital (unique ou périodique) versé avant que la rente ne prenne cours, aussi appelé "le capital constitutif de la rente". Les rentes peuvent être payées annuellement ou fractionnées (par exemple mensuellement).

Il existe plusieurs types de rentes viagères :

- ❖ Les rentes viagères différées : Les paiements commencent à une date future
- ❖ Les rentes viagères immédiates : Les paiements commencent immédiatement après le paiement d'une prime unique

Une rente viagère, peu importe son type, nécessite la constitution par les assureurs de provisions reflétant cet engagement. Ces provisions sont calculées à l'aide de facteurs d'annuités. Ces derniers représentent la somme théorique nécessaire pour assurer le versement d'une rente de 1 euro. Il est calculé à partir du taux technique (rémunération des sommes placées) et des probabilités de survie déterminées à partir des tables de mortalité. Nous allons revenir plus en détails au calcul des rentes dans la partie 3 du mémoire.

Le risque de longévité est étroitement lié à ce type de contrat. En effet, comme nous l'avons expliqué précédemment, ce risque provient de l'incertitude liée à l'allongement de la durée de vie des rentiers, c'est-à-dire le fait qu'ils vivent plus longtemps que prévu, ce qui, en conséquence, entraîne des paiements de rentes plus longs et donc plus coûteux que ceux initialement estimés par les hypothèses.

Par exemple, imaginons que l'assureur versant des rentes viagères annuelles ait estimé une espérance de vie de x années pour un rentier et que ce dernier vive $x + 20$ ans. Cela signifie que l'assureur devra

verser 20 années de rentes supplémentaires par rapport aux prévisions, ce qui entraîne des charges financières additionnelles pour l'assureur.

3) Revue des modèles de durée existants

L'étude de la mortalité a longtemps occupé une place importante dans les recherches médicales, démographiques, de planification sociale et économique ainsi que dans le domaine de l'assurance. Les modèles utilisés avant les années 1980 pour prédire directement les taux de mortalité et l'espérance de vie résiduelle étaient relativement simples et subjectifs, reposant sur un nombre important d'hypothèses souvent basées sur les observations et expériences personnelles des experts plutôt que sur des modèles statistiques plus rigoureux. Ces premiers modèles ont néanmoins permis de poser des fondations importantes pour la compréhension des tendances de mortalité et ont ouvert la voie à la construction de méthodes plus élaborées et complexes dans l'analyse démographique et actuarielle.

Cette section ne consiste pas en une présentation exhaustive de la littérature ; elle a pour but de présenter les principaux modèles de mortalité utilisés jusqu'à aujourd'hui y compris les modèles que nous avons choisi d'appliquer dans notre étude. Nous pouvons distinguer plusieurs grandes familles de modèles que nous présenterons dans cette partie :

Famille	Modèle	Année	Description
Les premières approches	Modèle de Brass	1971	Transforme une table de mortalité à partir d'une autre en utilisant une relation logistique.
	Modèle de Cox	1972	Permet d'examiner l'impact de diverses variables explicatives sur la survie et le risque de décès.
La famille Lee-Carter	Modèle de Lee-Carter (LC) et Log-Poisson	1992 et 2002	Approche basée sur des facteurs liés à l'âge et la période pour modéliser les taux de mortalité à travers le temps.
	Modèle de Renshaw-Haberman (RH)	2006	Extension du modèle de Lee-Carter prenant en compte les effets par cohorte.
	Modèle Age-Period-Cohort (APC)	2006	Sous-structure simplifiée du modèle de Renshaw-Haberman, prenant également en compte les effets de l'âge, la période et cohorte sur la mortalité.
La famille Cairns, Blake and Dowd	Modèle Cairns, Blake et Dowd et extensions	2006	Modèle basé sur deux facteurs : l'un affecte tous les âges de manière uniforme, l'autre impacte davantage les âges avancés. Plusieurs extensions de ce modèle ont été créées, dont plusieurs prenant en compte l'effet de cohorte.

1.1) Les premiers modèles de mortalité

Autrefois, les actuaires et démographes utilisaient principalement des tables de mortalité statiques basées sur des observations empiriques, comme les tables de vie de John Graunt en 1662 et Edmond Halley en 1693.

Au 19^{ème} siècle, les premiers modèles paramétriques de mortalité ont été créés. En 1825, Gompertz a proposé un modèle exponentiel pour décrire le taux de mortalité. Son modèle suggère que le taux de mortalité augmente de manière exponentielle avec l'âge. C'est l'un des premiers modèles paramétriques pour la mortalité. William Makeham a ensuite élargi, en 1862, le modèle de Gompertz en ajoutant un terme constant pour tenir compte des décès prématurés dus à des causes non liées à l'âge.

Avec le temps, les besoins en matière de modélisation de la mortalité ont évolué, et de nouvelles approches ont vu le jour pour répondre à des questions plus complexes telles que l'effet des facteurs de risque autres que l'âge sur la survie ainsi que l'analyse de la mortalité entre des populations différentes. C'est dans ce contexte qu'apparaissent les modèles de Cox et de Brass, qui marquent une avancée significative dans la manière dont les actuaires et les démographes abordent la modélisation de la mortalité.

1.1.1) Le modèle de Brass

Présentation du modèle

Le modèle de Brass a apporté une avancée notable en introduisant une méthode pour transformer les tables de mortalité en utilisant des relations logistiques. L'idée sous-jacente au modèle est qu'il est possible de décrire la mortalité observée par l'intermédiaire d'une table de mortalité de référence existante. Généralement, cette dernière est la table de mortalité de la population générale. Le modèle repose sur l'idée que les courbes de mortalité de différentes populations peuvent être alignées en utilisant une transformation logistique simple.

Les logits des probabilités de décès de la population observée sont linéairement liés dans ce modèle aux logits des probabilités de mortalité de référence par la relation suivante :

$$\text{logit}(q_x) = \alpha + \beta \times \text{logit}(q_x^{\text{ref}})$$

Le logit est défini par :

$$\text{logit}(q_x) = \ln \left(\frac{q_x}{1 - q_x} \right)$$

Le paramètre α est un indicateur de mortalité qui affecte l'ensemble du portefeuille indépendamment de l'âge, tandis que le paramètre β mesure la contribution du vieillissement à la mortalité.

Les paramètres du modèle sont généralement estimés par la méthode des moindres carrés ordinaires. Pour rappel cette méthode consiste à estimer les paramètres du modèle en minimisant la somme des carrés des différences entre les valeurs observées et les valeurs prédites. Il s'agit alors de minimiser :

$$\sum_{i=1}^n \left(\text{logit}(q_x) - (\alpha + \beta \times \text{logit}(q_x^{\text{ref}})) \right)^2$$

Conclusion sur le modèle de Brass

Le modèle de Brass est un outil efficace pour ajuster et comparer les courbes de mortalité entre différentes populations, particulièrement utile lorsque les données sont incomplètes ou peu fiables. En utilisant des transformations logistiques simples, il permet d'obtenir des estimations assez robustes des taux de mortalité. Cependant, ce modèle présente certaines limites, notamment le fait qu'il soit conçu uniquement pour ajuster les taux de mortalité et n'inclut pas la possibilité d'analyser l'impact de multiples facteurs ou covariables sur la mortalité. C'est pour cela que le modèle de Cox a été introduit.

1.1.2) Le modèle de Cox

Le modèle de Cox est un modèle de régression semi-paramétrique dit « à hasard proportionnel », c'est-à-dire que le modèle suppose que les variables explicatives telles que l'âge, le sexe ou la catégorie socio-professionnelle ont un effet multiplicatif sur la mortalité de base. Cette hypothèse, bien qu'elle soit forte, permet d'offrir un équilibre entre la simplicité mathématique du modèle, sa flexibilité et sa facilité d'interprétation.

Présentation du modèle

On dispose de n covariables de dimension p : Z_1, \dots, Z_p . En notant : $Z_i = (Z_{i1}, \dots, Z_{ip})^T$. On cherche alors à estimer le risque instantané conditionnellement à ces covariables :

$$h(t|Z_{i1}, \dots, Z_{ip}) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t, Z_{i1}, \dots, Z_{ip})}{dt}$$

Posons les paramètres $\theta_0 = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$. Le modèle de Cox s'écrit :

$$\begin{aligned} h(t|Z_{i1}, \dots, Z_{ip}) &= h_0(t) \times \exp(\theta_1 Z_{i1} + \dots + \theta_p Z_{ip}) \\ &= h_0(t) \times \exp(\theta_0 Z_i) \end{aligned}$$

Avec :

$$\left\{ \begin{array}{l} h_0(t) : \text{la fonction de hasard de base qui dépend uniquement du temps} \\ \theta_0 = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p, \text{ ce sont les paramètres du modèle de dimension } p \text{ représentant} \\ \text{l'effet des covariables sur le risque instantané, ils ne dépendent pas du temps} \\ \exp(\theta_0 Z_i) : \text{est le risque relatif} \end{array} \right.$$

Le modèle de Cox peut être vu comme le produit d'une fonction de hasard de référence, qui représente le risque instantané de décès et un risque relatif, représenté par le facteur exponentiel. Deux propriétés importantes ressortent de la définition du modèle : la log-linéarité et l'hypothèse des risques proportionnels.

Hypothèse de la Log-linéarité

La propriété de la log-linéarité implique, comme son nom l'indique, que le logarithme du risque instantané (le hasard) est une fonction linéaire des variables explicatives Z :

$$\ln\left(\frac{h(t|Z=z)}{h_0(t)}\right) = \theta^T z$$

Cette égalité peut aussi se réécrire comme :

$$\ln(h(t|Z=z)) = \ln(h_0(t)) + \theta^T z$$

Cette propriété implique que l'effet d'une augmentation d'une unité d'une covariable z sur le risque est multiplicatif et constant dans le temps. Cela peut être une hypothèse forte dans certains cas. Par exemple, si on souhaite étudier la survie des patients atteints d'un cancer en fonction de leur âge, cette hypothèse suppose que l'effet d'augmentation d'un an sur le risque est proportionnel, ce qui n'est pas réaliste puisque l'effet est, en général, plus prononcé chez un patient de 75 ans qu'un patient de 30 ans.

Hypothèse des risques proportionnels

La deuxième hypothèse du modèle de Cox implique que le rapport des risques instantanés de deux patients est indépendant du temps. C'est l'hypothèse des risques proportionnels :

$$\ln\left(\frac{h_i(t|Z=z_i)}{h_j(t|Z=z_j)}\right) = \theta^T (z_i - z_j)$$

En d'autres termes, cette hypothèse suggère que si un patient a un risque de décès deux fois plus élevé qu'un autre à un moment donné, ce rapport restera identique au fil du temps, ce qui reste également une hypothèse forte.

Conclusion sur le modèle de Cox

Le modèle de Cox, ou modèle de régression des risques proportionnels, est utilisé pour examiner l'impact des covariables sur le temps jusqu'à un événement particulier (le décès dans notre cas). Ce modèle a été un outil précieux dans divers domaines, notamment l'épidémiologie et les études cliniques. Il permet d'examiner comment des facteurs comme l'âge, le sexe, les habitudes de vie, ou les traitements médicaux influencent la survie et les résultats des patients.

Toutefois, bien que le modèle de Cox soit puissant pour des analyses de survie à court terme et spécifiques à des covariables, il peut montrer des limites lorsqu'il s'agit de modéliser la mortalité à long terme et de projeter les tendances futures. En effet, lorsque l'objectif est d'analyser la mortalité sur de longues périodes, nous avons vu que le modèle repose sur plusieurs hypothèses fortes et il devient nécessaire d'utiliser des modèles qui capturent non seulement les effets des covariables à un moment donné, mais aussi les variations de ceux-ci au fil du temps.

1.2) Famille de modèles « Lee-Carter »

1.2.1) Le modèle de Lee-Carter

Présentation du modèle

Le modèle de Lee-Carter a été introduit en 1992 par Ronald Lee et Lawrence Carter. Leur étude était construite sur des données de mortalité américaines et est rapidement devenue un standard pour la modélisation et la prévision de la mortalité. Le modèle de Lee-Carter se distingue par sa capacité à décomposer les taux de mortalité en trois composantes principales : un effet **âge**, un effet **temps** et un paramètre de **dévi**ation. L'effet âge permet de prendre en compte comment les taux de mortalité varient selon l'âge des individus. En règle générale, les taux de mortalité sont relativement élevés chez les nourrissons, diminuent chez les jeunes enfants et adolescents, restent relativement bas jusqu'à un certain âge adulte, puis augmentent de façon exponentielle quand les individus vieillissent. L'effet temps, quant à lui, permet de capturer comment la mortalité évolue au fil du temps en raison de facteurs tels que les progrès médicaux ou l'amélioration des conditions de vie. Enfin, le terme de déviation permet de prendre en compte d'autres effets aléatoires ou imprévus, il capte donc les fluctuations autour des tendances principales.

Le modèle Lee-Carter est donc adapté pour étudier l'évolution de la mortalité dans le temps, car il permet de modéliser les tendances globales sur le long terme tout en tenant compte des variations aléatoires qui peuvent survenir d'une année à l'autre. Cette approche est pertinente pour évaluer les risques liés à des produits comme les rentes viagères et les assurances vie puisque ces produits impliquent généralement des engagements financiers dépendant de la mortalité des assurés qui s'étendent sur de longues périodes, souvent de plusieurs décennies.

Le modèle de Lee-Carter propose la modélisation suivante du taux de mortalité instantané :

$$\ln(\mu_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

Avec :

{	μ_{xt} :	La force de mortalité à l'âge x et au temps t
	α_x :	Valeur moyenne des $\ln(\mu_{xt})$ au cours du temps
	β_x :	Sensibilité de la mortalité instantanée à l'âge x par rapport à l'évolution générale κ_t
	κ_t :	Indice temporel représentant l'évolution de la mortalité dans le temps, il est utilisé pour extrapoler les taux futurs
	ϵ_{xt} :	Terme d'erreur qui représente les fluctuations non expliquées ou aléatoires autour du modèle

Le modèle permet alors d'expliquer le taux de mortalité instantané à un âge x et à un temps t à partir de trois éléments principaux α_x , β_x et κ_t .

Le paramètre α_x permet de capturer l'effet de l'âge sur la mortalité, sans tenir compte des variations temporelles. En générale, en dehors des décès prématurés, ce paramètre croît avec l'âge, ce qui signifie que la mortalité augmente avec le vieillissement des individus.

La composante temporelle κ_t permet, quant à elle, de capturer l'effet temporelle sur la mortalité, c'est-à-dire les variations générales qui peuvent arriver au fil du temps, par exemple l'amélioration des conditions de vie.

Enfin, le paramètre β_x permet d'expliquer la sensibilité de la mortalité à un âge x par rapport à l'évolution temporelle générale. En d'autres termes, cela correspond à la manière dont les différents âges réagissent face aux variations temporelles générales. Par exemple, un β_x élevé pour un âge donné signifie que la mortalité à cet âge x est fortement sensible à l'évolution générale telles que les avancées médicales.

Il est important de noter qu'il s'agit d'un modèle formulé en temps continu. Cependant, pour des applications numériques, nous faisons souvent une hypothèse classique mais forte qui est la constance de la force de mortalité par morceaux entre les âges et les années. Sous cette hypothèse, μ_x devient par définition égale à $-\ln(p_x)$. De plus, pour des raisons pratiques, il est courant de réaliser l'approximation de $\mu_{x,t}$ par $m_{x,t}$. Nous allons supposer cette hypothèse tout au long du mémoire.

Estimation des paramètres

Nous pouvons remarquer que le modèle est non identifiable, c'est à dire qu'il existe plusieurs solutions de paramètres qui peuvent s'appliquer et donner des résultats équivalents. En d'autres termes, les paramètres estimés du modèle ne sont alors pas obtenus de manière unique. Par exemple, avec la transformation suivante des paramètres (c une constante), le modèle reste invariant :

$$(\alpha_x, \beta_x, \kappa_t) \rightarrow (\alpha_x, \frac{\beta_x}{c}, c\kappa_t)$$

$$(\alpha_x, \beta_x, \kappa_t) \rightarrow (\alpha_x - c\beta_x, \beta_x, c + \kappa_t)$$

C'est pourquoi, afin de rendre le modèle identifiable, des contraintes ont été imposées sur les paramètres. En général, pour $x \in [x_1, \dots, x_p]$ et $t \in [t_1, \dots, t_n]$, on retiendra les contraintes :

$$\sum_{x=x_1}^{x_p} \beta_x = 1 \quad \text{et} \quad \sum_{t=t_1}^{t_n} \kappa_t = 0$$

Ensuite, pour estimer les paramètres, Lee et Carter ont proposé la méthode des moindres carrés ordinaires :

$$(\widehat{\alpha}_x, \widehat{\beta}_x, \widehat{\kappa}_t) = \underset{x,t}{\operatorname{argmin}} \sum (\ln \mu_{x,t} - \alpha_x - \beta_x \kappa_t)^2$$

Estimation du paramètre α_x

On a :

$$\frac{\partial}{\partial \alpha_x} \sum_{y,t} (\ln \mu_{x,t} - \alpha_y - \beta_y \kappa_t)^2 = 2(t_n - t_1 + 1) \alpha_x - 2 \sum_{t=t_1}^{t_n} (\ln \mu_{x,t} - \beta_x \kappa_t)$$

On trouve en tenant compte de la contrainte $\sum_{t=t_1}^{t_n} \kappa_t = 0$ que :

$$\widehat{\alpha}_x = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln \mu_{x,t}$$

En d'autres termes, $\widehat{\alpha}_x$ est la moyenne temporelle, à l'âge x , des logarithmes des taux instantanés de décès.

Estimation des paramètres β_x et κ_t : Décomposition en valeurs singulières

On considère la matrice Z des taux centrés z_{xt} :

$$z_{xt} = \ln \mu_{x,t} - \widehat{\alpha}_x$$

Ainsi, Z est une matrice de dimension $(x_p - x_1 + 1, t_n - t_1 + 1)$

On cherche alors une approximation de Z en produit de deux vecteurs de la forme $Z = \widehat{\beta} \widehat{\kappa}^T$ avec $\widehat{\beta} = (\widehat{\beta}_{x_{min}}, \dots, \widehat{\beta}_{x_{max}})^T$ et $\widehat{\kappa} = (\widehat{\kappa}_{t_{min}}, \dots, \widehat{\kappa}_{t_{max}})^T$, afin que la décomposition soit optimale au sens du critère des moindres carrés et donc qu'elle explique la plus grande part possible de la variance totale.

Pour résoudre cela, il faut décomposer en valeurs propres la matrice Z , que l'on met sous la forme :

$$Z = \sum_{i \geq 1} \sqrt{\lambda_i} v_i u_i^T$$

avec $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ les valeurs propres de $Z^T Z$, u_i le vecteur propre normé de $Z^T Z$ associé à λ_i , et v_i le vecteur propre associé à la même valeur propre pour ZZ^T .

Ceci est justifié par le fait que, comme $Z^T Z u_i = \lambda_i u_i$, on a $ZZ^T (Z u_i) = \lambda_i Z u_i$, et donc les deux matrices transposées ont les mêmes valeurs propres avec des ordres de multiplicité identiques. De

plus, si u_i est un vecteur propre de $Z^T Z$, alors Zu_i est un vecteur propre de ZZ^T associé à la même valeur propre.

Ainsi, si on pose : $v_i = \frac{1}{\sqrt{\lambda_i}} Zu_i$, on voit que : $Zu_i u_i^T = \sqrt{\lambda_i} v_i u_i^T$

Ce qui, en sommant et en tenant compte de l'orthonormalité des vecteurs propres, on peut approximer Z par la combinaison des vecteurs propres associés à la première valeur propre λ_1 :

$$Z \approx \sqrt{\lambda_1} v_1 u_1^T$$

On peut mesurer la qualité de l'approximation par le taux d'inertie (ou pourcentage de variance expliquée) qui indique la part de la "variance" capturée par la première valeur propre λ_1 par rapport à la somme de toutes les valeurs propres, donné par $\frac{\lambda_1}{\sum \lambda_i}$.

En supposant $\sum v_{1j} \neq 0$, on obtient finalement les estimateurs de β et κ .

$$\hat{\beta} = \frac{1}{\sum v_{1j}} v_1 \text{ et } \hat{\kappa} = \sqrt{\lambda_1} \sum v_{1j} u_1$$

Le calcul de la décomposition en valeurs singulières (SVD) d'une matrice peut être très coûteux en ressources, surtout pour les grandes matrices. En effet, le temps nécessaire pour effectuer une SVD augmente rapidement avec la taille de la matrice, avec une complexité de l'ordre de $O(n^3)$ pour une matrice de taille $n \times n$. Cela peut rendre le calcul lent et difficile à réaliser, notamment pour les applications en temps réel. De plus, l'exigence que les vecteurs propres soient orthogonaux peut poser des problèmes, surtout lorsque les données ne suivent pas naturellement ce type de structure.

Ajustement des paramètres

Ceci nous donne une première estimation des paramètres du modèle. Nous cherchons maintenant à crédibiliser les paramètres du modèle pour que le nombre de décès prévus par le modèle chaque année soit égal au nombre de décès observés.

Nous avons la relation suivante de l'exposition E_{xt} au risque :

$$E_{xt} = \frac{D_{xt}}{\mu_{x,t}}$$

La contrainte d'ajustement s'écrit donc :

$$\sum_{x=x_1}^{x_p} D_{xt} = \sum_{x=x_1}^{x_p} E_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$$

On introduit la variable : $F(k) = \sum_{x=x_1}^{x_p} E_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t) - \sum_{x=x_1}^{x_p} D_{xt}$

Nous cherchons alors : $F(k) = 0$

La recherche de la racine de la fonction F peut se faire ensuite par un algorithme de type Newton-Raphson.

Extrapolation de la composante temporelle

Une fois que les paramètres sont estimés, il reste à modéliser la série temporelle k_t . De nombreuses modélisations temporelles peuvent être utilisées (ARIMA, marche aléatoire avec dérive, régression linéaire, etc.) pour la modélisation de ce paramètre, mais le modèle retenu le plus souvent est le modèle ARIMA. Nous allons dans ce mémoire étudier les modélisations ARIMA et marche aléatoire avec dérive pour nos séries.

Modélisation par marche aléatoire multivariée avec dérive

Une marche aléatoire multivariée avec dérive est un processus stochastique qui décrit l'évolution d'une variable aléatoire dans le temps en supposant une tendance systématique (la "dérive") et un terme contenant les fluctuations aléatoires. Un processus de marche aléatoire avec dérive peut s'exprimer ainsi :

$$k_t = \delta + k_{t-1} + \varepsilon_t^k$$

Avec :

- k_t la valeur de la variable à l'instant t
- δ la dérive, représentant la tendance moyenne du processus qui peut être positive ou négative
- k_{t-1} la valeur de la variable à l'instant t-1
- ε_t^k un bruit blanc gaussien, modélisant les fluctuations aléatoires

Modélisation par ARIMA

Avant de présenter le processus ARIMA, il est essentiel de comprendre les différents processus qui le composent :

Les processus AR

Le processus AR fait référence à un processus autorégressif, c'est-à-dire un processus où la valeur actuelle d'une série dépend de ses valeurs passées. On considère qu'un processus stationnaire X_t est autorégressif d'ordre p si l'on peut expliquer sa valeur à l'instant t en utilisant ses p termes précédents :

$$\forall t : X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \epsilon_t$$

Avec ϵ un bruit blanc et les α_i des réels.

Cela veut dire que la valeur actuelle de la série temporelle est exprimée comme une fonction linéaire de ses valeurs passées plus un terme d'erreur, avec p le nombre de valeurs passées qui doivent être prises en compte.

Les processus MA

Un processus Moyenne Mobile (*Moving Average*) d'ordre q est un autre type de modèle utilisé pour expliquer une série temporelle. Contrairement aux processus auto-régressifs (AR), où la valeur actuelle d'une variable dépend de ses valeurs passées, dans un processus à moyenne mobile, la valeur actuelle dépend des valeurs passées des termes d'erreur.

Un processus stationnaire X_t d'ordre q peut donc être exprimé ainsi :

$$\forall t : X_t = \sum_{i=1}^q \beta_i \epsilon_{t-i}$$

Avec q le nombre de termes d'erreur antérieurs à prendre en compte.

Les processus ARMA

Le processus ARMA est tout simplement une combinaison d'un processus AR et d'un processus MA. Cela permet de modéliser des séries temporelles plus complexes. Un modèle ARMA d'ordre (p, q) s'écrit donc :

$$\forall t : X_t = \sum_{i=1}^p \alpha_i X_{t-i} + \epsilon_t + \sum_{i=1}^q \beta_i \epsilon_{t-i}$$

Cependant, l'une des limitations de ces modèles est qu'ils ne peuvent modéliser que des séries temporelles stationnaires. C'est pour cela que les processus ARIMA interviennent.

Les processus ARIMA

Le I de ARIMA signifie '*integrated*', intégré. L'idée de cette étape en plus est que, en différenciant successivement les séries temporelles, il est possible de retirer les tendances qu'elles présentent afin de les stationnariser.

Prenons l'exemple d'une série temporelle avec une tendance linéaire de la forme : $X_t = \alpha + \beta t + \epsilon_t$

Ainsi :

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

On observe qu'en différenciant la série une fois, la dépendance temporelle linéaire est éliminée et la différence est stationnaire. De la même façon, une tendance quadratique peut être éliminée en différenciant la série deux fois. Une fois la série stationnarisée, il est alors possible d'appliquer le modèle ARMA.

Un modèle ARIMA est composé de 3 paramètres : les paramètres p et q du modèle ARMA, ainsi que le paramètre d qui représente le nombre de fois que la série doit être différenciée pour la rendre stationnaire.

Choix du modèle

Pour choisir entre les différents modèles, il est essentiel d'examiner les propriétés de la série temporelle étudiée en se posant plusieurs questions clés.

D'abord, la série est-elle stationnaire ? Il s'agit d'une condition nécessaire pour l'utilisation des modèles AR, MA et ARMA. Dans le cas où la série ne l'est pas, il faudra se diriger vers un modèle ARIMA afin de la rendre stationnaire.

Aussi, nous pouvons nous demander s'il existe-t-il une corrélation entre chaque valeur avec les valeurs passées. Cette question se vérifie à l'aide des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) qui mesurent la corrélation entre les valeurs de la série à différents décalages (lags) et peuvent être tracées facilement sur R ou Python. Ces fonctions, non seulement nous donnent une idée de la nature du modèle à choisir, mais elles peuvent également nous indiquer l'ordre de ce modèle.

Une fois que nous avons choisi quelques modèles potentiels, plusieurs critères utilisant la fonction de vraisemblance comme le critère AIC (*Akaike Information Criterion*) ou le critère BIC (*Bayesian Information Criterion*) peuvent être utilisés afin de choisir le modèle plus adéquat. Il existe aussi des critères d'erreur permettant d'analyser la modélisation effectuée : erreur racine quadratique

moyenne (RMSE), erreur quadratique moyenne (MSE), erreur moyenne absolue. Nous allons détailler plusieurs de ces critères plus tard dans ce mémoire.

Conclusion sur le modèle de Lee-Carter

Le modèle de Lee-Carter permet une approche simple mais puissante pour décomposer et projeter les taux de mortalité. Cependant, il présente certaines limites importantes :

D'abord, le modèle Lee-Carter ne prend pas en compte l'effet de cohorte qui pourrait être présent entre les générations. En effet, l'effet de cohorte, ou effet de génération, se réfère aux variations dans les comportements, attitudes ou conditions de santé et donc la mortalité attribuée à un groupe de personnes nées pendant la même année ou période. Le raisonnement est que chaque cohorte traverse différentes étapes de la vie dans des contextes sociaux, économiques et culturels spécifiques, influençant leurs caractéristiques de manière unique par rapport à d'autres cohortes. Par exemple, une génération ayant vécu une guerre ou une crise économique peut avoir des comportements distincts comparée à une génération ayant grandi en période de prospérité. Il s'agit donc d'un effet non négligeable pouvant affecter la mortalité. Une étude réalisée par Nathalie Blanpain à l'INSEE montre un effet de cohorte existant sur la mortalité pour les générations nées entre 1941 et 1955. Cette recherche montre comment plusieurs facteurs liés spécifiquement à ces générations tels que l'augmentation des accidents de la route, l'apparition du Sida ainsi que des modes de vie liés à la consommation du tabac et de l'alcool ont créé des disparités en matière de santé et de mortalité créant ainsi un effet de cohorte.⁵

Ensuite, le modèle repose sur l'hypothèse d'homoscédasticité des taux de mortalité, ce qui constitue une hypothèse forte et peu réaliste, car en générale la variance des taux de décès croît aux âges élevés, ce qui est notamment dû à la baisse des effectifs de survivants.

De plus, l'hypothèse de constance de force de mortalité par morceaux pourrait être considérée comme une hypothèse forte et discutable notamment pour les âges élevés.

Enfin, nous pouvons observer que la méthode d'estimation des paramètres proposée par Lee et Carter ne repose pas sur une distribution de probabilité. C'est pour cela que Brouhns et al. (2002) ont cherché à trouver une adaptation du modèle Lee-Carter classique : le modèle Log-Poisson, qui associe une loi de probabilité à la variable de décès $D_{x,t}$. L'idée du modèle Log-Poisson est de supposer que la variable de décès $D_{x,t}$ suit une loi de poisson de paramètre $E_{x,t}\mu_{x,t}$. Le taux de décès instantané $\mu_{x,t}$ est modélisé par la même formule que le modèle Lee-Carter classique, et les contraintes d'indentifiabilité restent les mêmes. Ainsi, grâce à ce modèle, il devient possible de réaliser l'estimation des paramètres selon le critère de maximum de vraisemblance. Nous allons alors adopter cette approche pour le reste du mémoire.

⁵ Source : Blanpain, N. "La mortalité stagne à l'âge adulte pour les générations nées entre 1941 et 1955 ». [INSEE](#)

Afin de pallier plusieurs de ces limites, des généralisations et adaptations du modèle de Lee-Carter ont été proposées.

1.2.2) Le modèle de Renshaw-Haberman

Présentation du modèle

Le modèle de Renshaw-Haberman est une généralisation du modèle de Lee-Carter, permettant de capturer les effets de cohorte et les tendances spécifiques à chaque génération dans l'analyse de la mortalité.

Renshaw et Haberman ont proposé une généralisation du modèle de Lee-Carter en incorporant l'année de naissance dans le modèle et donc l'effet de cohorte. La modélisation du taux de mortalité proposée est la suivante :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(0)} \gamma_{t-x} + \epsilon_{xt}$$

Avec :

$$\left\{ \begin{array}{l} \mu_{x,t} : \text{Le taux de mortalité instantané à l'âge } x \text{ et à l'instant } t \\ \alpha_x : \text{La valeur moyenne des } \ln(\mu_{xt}) \text{ au cours du temps} \\ \beta_x^{(1)} : \text{La sensibilité de la mortalité instantanée à l'âge } x \text{ par rapport à l'évolution générale } \kappa_t \\ \kappa_t : \text{La composante temporelle qui représente l'évolution générale de la mortalité au cours du} \\ \text{temps. Il s'agit du paramètre qui servira à l'extrapolation des taux dans le futur} \\ \beta_x^{(0)} : \text{La sensibilité de la mortalité instantanée à l'âge } t-x \text{ par rapport à l'évolution } \gamma_{t-x} \\ \gamma_{t-x} : \text{La composante prenant en compte l'effet de cohorte} \\ \epsilon_{xt} : \text{Le terme d'erreur} \end{array} \right.$$

Estimation des paramètres

De la même manière que pour le modèle de Lee-Carter, il convient d'imposer des contraintes sur les paramètres afin de pouvoir les estimer de manière unique. Ainsi, nous ajoutons les contraintes suivantes aux contraintes que nous avons imposées pour estimer les paramètres du modèle de Lee-Carter :

$$\sum_{x=x_1}^{x_p} \beta_x^{(0)} = 1 \text{ et } \sum_{t-x=t_1-x_p} \gamma_{t-x} = 0$$

Il est possible d'estimer les paramètres du modèle de la même manière que pour le modèle de Lee-Carter en utilisant la méthode des moindres carrés ordinaires ou le maximum de vraisemblance.

Conclusion sur le modèle de Renshaw-Haberman

Le modèle de Renshaw-Haberman est une généralisation du modèle de Lee-Carter, permettant de capturer les effets de cohorte et les tendances spécifiques à chaque génération dans l'analyse de la mortalité. Cependant, la structure du modèle entraîne un grand nombre de paramètres à estimer, estimation qui en outre est rendue difficile par la multicollinéarité entre les variables. Cela peut ainsi complexifier et poser des défis lors de l'estimation des paramètres, en particulier en termes de convergence des algorithmes d'estimation. Ce problème est noté par un certain nombre de chercheurs, y compris Cairns et al. [2009] [2011], Haberman et Renshaw [2009], [2011] et Currie [2016].

Ces problèmes de convergence peuvent entraîner des estimations instables ou non fiables. Pour réduire certaines de ces limitations du modèle de Renshaw-Haberman, le modèle Age-Period-Cohort (APC) constitue une alternative intéressante.

1.2.3) Le modèle Age-Period-Cohort

Présentation du modèle

Le modèle Age-Period-Cohort correspond à une sous-structure du modèle de Renshaw-Haberman où on a imposé les contraintes suivantes :

$$\beta_x^{(1)} = 1 \quad \beta_x^{(0)} = 1$$

On a alors :

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

Ce modèle, grâce à sa structure simplifiée réduisant le nombre de paramètres à estimer, permet de surmonter les problèmes de convergence rencontrés avec le modèle de Renshaw-Haberman. Il a été couramment utilisé en études de démographie et médecine et devient connu dans les études actuarielles grâce aux travaux de Currie [2016].

Le modèle APC est moins flexible que le modèle RH à cause de sa structure purement additive : chaque effet contribue indépendamment à l'estimation des taux de mortalité.

En effet, contrairement au modèle RH, l'effet de cohorte et de temps ne sont pas modulés par des composantes d'âge (β_x), réduisant ainsi le nombre de paramètres à estimer. Cependant, cette simplification implique que le modèle ne prend pas en compte les variations de la mortalité au fil du temps, spécifiques à chaque groupe d'âge.

1.3) Famille de modèles « Cairns-Blake-Dowd »

En 2006, Cairns et al. ont introduit un nouveau modèle connu sous le nom de modèle Cairns-Blake-Dowd (CBD), ou « modèle à deux facteurs ». Ce modèle a pour spécificité le pouvoir de capturer davantage les dynamiques de mortalité de manière plus adaptée et sensible aux âges avancés que les modèles de la famille Lee-Carter en supposant que le logit de la probabilité de décès est une fonction linéaire de l'âge.

Présentation du modèle

Cairns Blake et Dowd ont proposé un modèle contenant deux facteurs de temps. La modélisation du quotient de mortalité proposée est la suivante :

$$\text{logit}(q_x(t)) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(0)}$$

On rappelle que le logit est défini par :

$$\text{logit}(q_x(t)) = \ln\left(\frac{q_x(t)}{1 - q_x(t)}\right)$$

Avec :

$$\left\{ \begin{array}{l} q_x(t) \text{ la probabilité de décès à l'âge } x \text{ pendant l'année } t \\ x \text{ l'âge} \\ \bar{x} \text{ l'âge moyen dans la population étudiée} \\ \mathbf{k}_t^{(1)} \text{ et } \mathbf{k}_t^{(0)} \text{ des paramètres qui varient dans le temps} \end{array} \right.$$

Nous pouvons dire que le modèle est formé de deux composantes : La première composante $\mathbf{k}_t^{(1)}$ capture l'effet global du temps sur la mortalité indépendamment de l'âge. Elle affecte le taux de mortalité de la même manière à tous les âges. La seconde composante $\mathbf{k}_t^{(0)}(x - \bar{x})$ quant à elle, influence davantage les taux de mortalité à un âge élevé qu'à un jeune âge.

Conclusion sur le modèle de Cairns-Blake-Dowd

Le modèle de Cairns-Blake-Dowd constitue une évolution importante dans la modélisation stochastique de la mortalité. Comparé aux modèles précédents comme celui de Lee-Carter, il offre plusieurs changements clés, notamment la prise en compte explicite de la dépendance à l'âge, à travers le terme x , offrant une meilleure adaptabilité pour les âges élevés et l'utilisation de deux facteurs stochastiques pour modéliser de manière plus flexible les taux de mortalité. Plusieurs variantes du modèle de Cairn-Blake-Dowd ont été créés mais que nous n'allons pas détailler dans ce mémoire.

Certes la famille des modèles CBD pourrait offrir une approche intéressante pour modéliser la mortalité, nous avons choisi, dans le cadre de notre étude de nous concentrer sur la famille de modèles « Lee-Carter » afin de modéliser la mortalité. En effet, il s'agit de la famille de modèles la plus largement utilisée dans la littérature et offre une simplicité de mise en place. Nous examinerons le modèle Lee-Carter poissonien simple ainsi que ses variantes, notamment le modèle Renshaw-Haberman et le modèle Age-Period-Cohort, qui prennent en compte l'effet de cohorte.

4) La pandémie de Covid-19 : quels enjeux sur la mortalité ?

Depuis son apparition à la fin de 2019, la pandémie de COVID-19 est devenue la plus grande crise de santé publique à laquelle le monde a été confronté en plus d'un siècle. D'après Worldometer, une source qui regroupe ses données auprès des organisations nationales et internationales les plus fiables et qui a pour but de rendre les statistiques mondiales disponibles au grand public, la COVID-19 a causé plus de 7 millions de morts et 700 millions d'infections. De plus, depuis son émergence, les mesures de politique de santé employées pour minimiser ses effets ont varié considérablement d'un pays à un autre. Ces mesures ont affecté de nombreux facteurs sociaux et économiques, notamment l'accès aux services de santé.

Afin d'évaluer les conséquences de cette pandémie sur la mortalité, les données sur la mortalité toutes causes confondues sont considérées comme les indicateurs les plus fiables. En effet, d'après l'OCDE⁶, utiliser le nombre total de décès peut aider à surmonter les différences entre les pays dans les pratiques de codage et prendre en compte non seulement la possible sous-déclaration des décès liés à la COVID-19, mais aussi la mortalité indirecte causée par l'incapacité des systèmes de santé à gérer d'autres maladies durant cette crise. Cela permet alors d'effectuer des comparaisons moins biaisées entre les pays et donner une meilleure vue d'ensemble de l'effet global de la COVID-19. C'est pour cela que, dans cette étude, nous avons choisi d'utiliser les données historiques de mortalité toutes causes confondues provenant de la base de données *Human Mortality Database*, disponible en libre accès.

Cette base de données HMS a été utilisée dans de nombreuses études portant sur la mortalité liée à la pandémie. Par exemple, un grand nombre de recherches ont exploité ces données pour estimer l'effet du COVID-19 sur la mortalité en mesurant directement l'espérance de vie en 2020 et en la comparant avec celle en 2019. L'idée de cette approche est que, en supposant qu'aucun autre événement majeur n'ait eu un effet significatif sur la mortalité en 2020, les différences de niveaux de mortalité entre 2019 et 2020 peuvent être attribuées à la pandémie de COVID-19. Les études existantes utilisant cette stratégie montrent que la pandémie a provoqué une mortalité considérablement élevée en 2020, réduisant notablement l'espérance de vie observée entre 2019 et 2020 dans de nombreux pays. De plus, la diminution de l'espérance de vie variait considérablement selon le sexe, étant disproportionnellement plus élevée parmi les hommes.

Nous pouvons observer ces résultats sur la figure 2, tirée d'une étude publiée dans *'International Journal of Epidemiology*, qui utilise cette approche pour quantifier les variations d'espérance de vie dues à la pandémie dans 29 pays. Nous pouvons remarquer que la plupart des pays sur le graphique ont eu une perte d'espérance de vie entre 2019 et 2020, et que ces pertes ont été généralement plus élevées chez les hommes que chez les femmes. En France, les femmes ont perdu 0,6 ans d'espérance de vie tandis que les hommes ont perdu 0,67 ans. Aux Etats-Unis, les pertes sont encore plus marquées, avec les femmes ayant perdu 1,65 ans et les hommes 2,23 ans.

⁶ L'organisation de coopération et de développement économiques est une organisation internationale qui regroupe 38 pays membres. Elle vise à promouvoir les politiques publiques favorisant la croissance économique, la stabilité financière et le bien-être social.

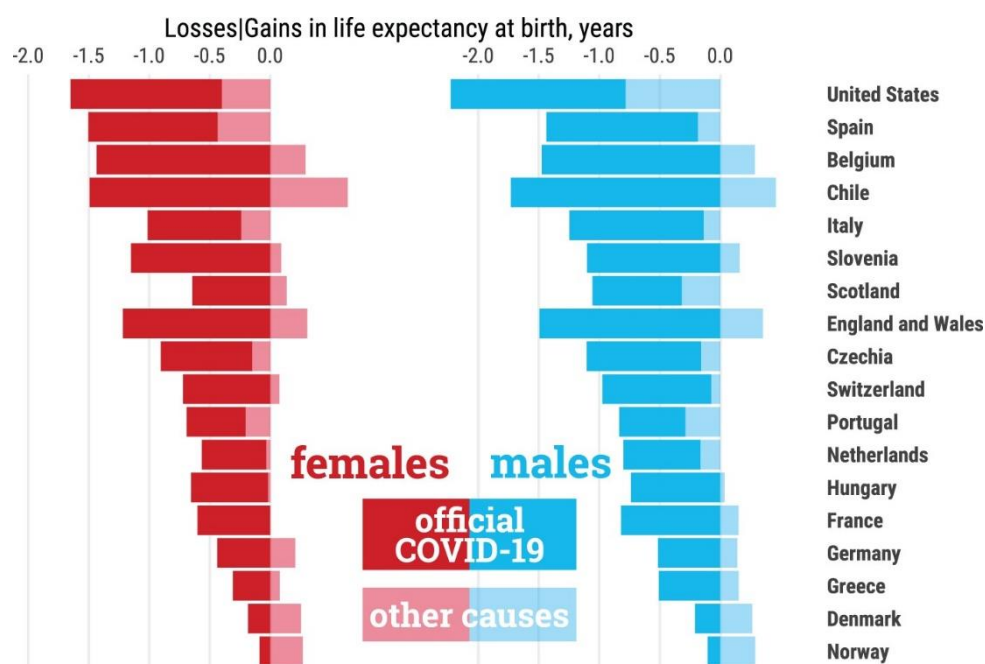


Figure 2 : Contributions (en années) aux changements dans l'espérance de vie à la naissance de 2019 à 2020 attribuables aux décès officiels dus au COVID-19 et aux autres causes de décès.

Cependant, simplement analyser les changements réels de l'espérance de vie entre 2019 et 2020 peut ne pas saisir pleinement les effets du COVID-19 sur la mortalité. En effet, cette approche, certes est intuitive, mais ne prend pas en compte les variations intrinsèques de la mortalité d'une année sur l'autre en l'absence de COVID-19 : ce qui se serait passé en 2020 sans l'épidémie de COVID-19.

Depuis la fin de la Seconde Guerre mondiale, la plupart des pays ont connu des transitions démographiques, avec une espérance de vie généralement en augmentation pendant plus de sept décennies. Cependant, les trajectoires croissantes de l'espérance de vie n'ont pas été linéaires, avec des rythmes variables selon les nations et même des reculs occasionnels pendant des périodes spécifiques pour certains pays. En l'absence de pandémie, l'augmentation générale de l'espérance de vie était censée se poursuivre en 2020 dans presque tous les pays.

Ainsi, sans considérer les changements attendus d'une année sur l'autre en mortalité en 2020, la perte totale d'espérance de vie causée par la pandémie de COVID-19 serait probablement sous-estimée.

C'est pour cela que, dans ce mémoire, pour étudier l'effet de la pandémie en année *jump-off* sur les estimations des espérances de vie résiduelles, nous n'allons pas comparer directement les données projetées avec les données 2019 réelles, mais nous allons supposer deux scénarios :

- **Scénario Baseline** (de référence) : Ce scénario représente la situation hypothétique où la pandémie n'a jamais eu lieu. Nous calibrons nos modèles en utilisant les données historiques de mortalité s'arrêtant en 2019 inclus, l'année juste avant l'apparition de la Covid. Ce scénario va nous permettre d'obtenir les projections des mortalités qui ont été attendues en l'absence de la pandémie, en tenant compte des variations intrinsèques qui auraient pu se produire dans des conditions normales.

- **Scénario COVID** : Ce scénario intègre le choc de la pandémie de COVID-19. La calibration des modèles est effectuée sur les données historiques s'arrêtant à 2020, prenant donc en compte ce choc de mortalité. Les projections obtenues avec ce scénario représentent les mortalités attendues dans un modèle intégrant le choc en année *jump-off*.

Cette démarche a déjà été utilisée dans plusieurs études de mortalité liée à la pandémie. L'idée principale est que, en comparant les résultats issus des deux scénarios, nous pouvons isoler l'effet de la pandémie sur les estimations de l'espérance de vie et évaluer la déviation par rapport aux tendances préexistantes.

Pour cela, nous allons, dans ce mémoire, calibrer nos modèles de longévité sous chaque scénario, en choisissant le modèle le plus adapté pour chaque population étudiée, que nous présenterons dans la partie suivante. Nous allons également analyser comment le choc de la pandémie influence la calibration des modèles en regardant, pour une même population, si un modèle pertinent dans le scénario Baseline reste adéquat après le choc dans le scénario Covid. Si c'est le cas, nous allons observer comment la prise en compte de ce choc a influencé les paramètres de ce modèle.

Ensuite, nous allons réaliser les projections issues des modèles choisis afin de construire nos tables de mortalité générationnelles. Cela nous permettra par la suite de comparer les estimations d'espérances de vie entre les deux scénarios, ainsi qu'entre les différentes populations. Ainsi, nous allons pouvoir évaluer non seulement l'effet de la COVID, mais aussi si cet effet a varié d'une population à l'autre.

Enfin, nous allons finir par réaliser ces comparaisons sur les engagements des rentes viagères de notre portefeuille.

Partie II : Modélisation de la mortalité

1) Les données HMD

Nous avons obtenu les données historiques à partir de la *Human Mortality Database*, une base de données en libre accès dans laquelle les données sur la mortalité et la population, provenant d'agences nationales autorisées, sont regroupées et standardisées. La base de données du HMD est devenue depuis plusieurs années une référence en matière de diffusion de tables de mortalités nationales. Elle est maintenue par le département de démographie de l'Université de Californie à Berkeley, aux États-Unis, et par l'Institut Max Planck pour la recherche démographique à Rostock, en Allemagne. Elle rassemble des données pour 45 pays, incluant des informations telles que le nombre de naissances, le nombre de morts, la taille de la population et l'exposition au risque, couvrant des années qui remontent au 19e siècle pour certains pays. De plus, la base de données HMD offre des données distinctes pour les hommes et les femmes, permettant ainsi de prendre en compte les différences de tendances de mortalité entre les sexes.

Ainsi, pour chaque sexe, à partir nombre de décès $D_{x,t}$ et l'exposition au risque de décès $E_{x,t}$, nous pouvons obtenir une approximation du taux central de mortalité pour les personnes âgées de x ans pendant l'année t comme suit :

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$$

Pour notre étude, nous avons décidé de prendre en compte les données à partir de 1990 afin d'analyser les tendances sur trois décennies complètes. Ce choix nous permet d'éviter les perturbations liées aux changements territoriaux et aux effets des guerres mondiales sur les frontières et les données. En effet, à partir de 1990, la plupart des pays ont déjà atteint une stabilité territoriale, sans changements significatifs de frontières, ce qui nous permet de travailler avec des données plus cohérentes et représentatives des pays actuels.

Concernant les âges, nous avons choisi de nous concentrer sur la tranche d'âge allant de 15 à 110 ans. Le choix de commencer vers 15 ans se base sur le fait que c'est autour de cet âge que les jeunes peuvent légalement entrer sur le marché du travail dans de nombreux pays. De plus, inclure les personnes jusqu'à 110 semble pertinent pour les études de longévité et les rentes viagères, puisque cela permet d'étudier les tendances jusqu'à des âges très avancés où le risque de longévité est plus présent.

2) Présentation des pays étudiés

Nous avons choisi de réaliser cette étude pour 5 pays :

- La France
- L'Espagne
- Le Royaume-Uni
- Le Japon
- Le Canada

Le choix de ces pays a été réalisé suivant la disponibilité des données mais également par la diversité géographique et de tendances qu'ils représentent. Pour avoir une idée de la tendance globale de chaque pays, on peut observer sur la figure 3 ci-dessous le taux moyen de décès par 1000 personnes par pays de 1950 jusqu'à 2023.

Évolution de la mortalité par pays depuis 1950

Taux de mortalité par 1000 habitants

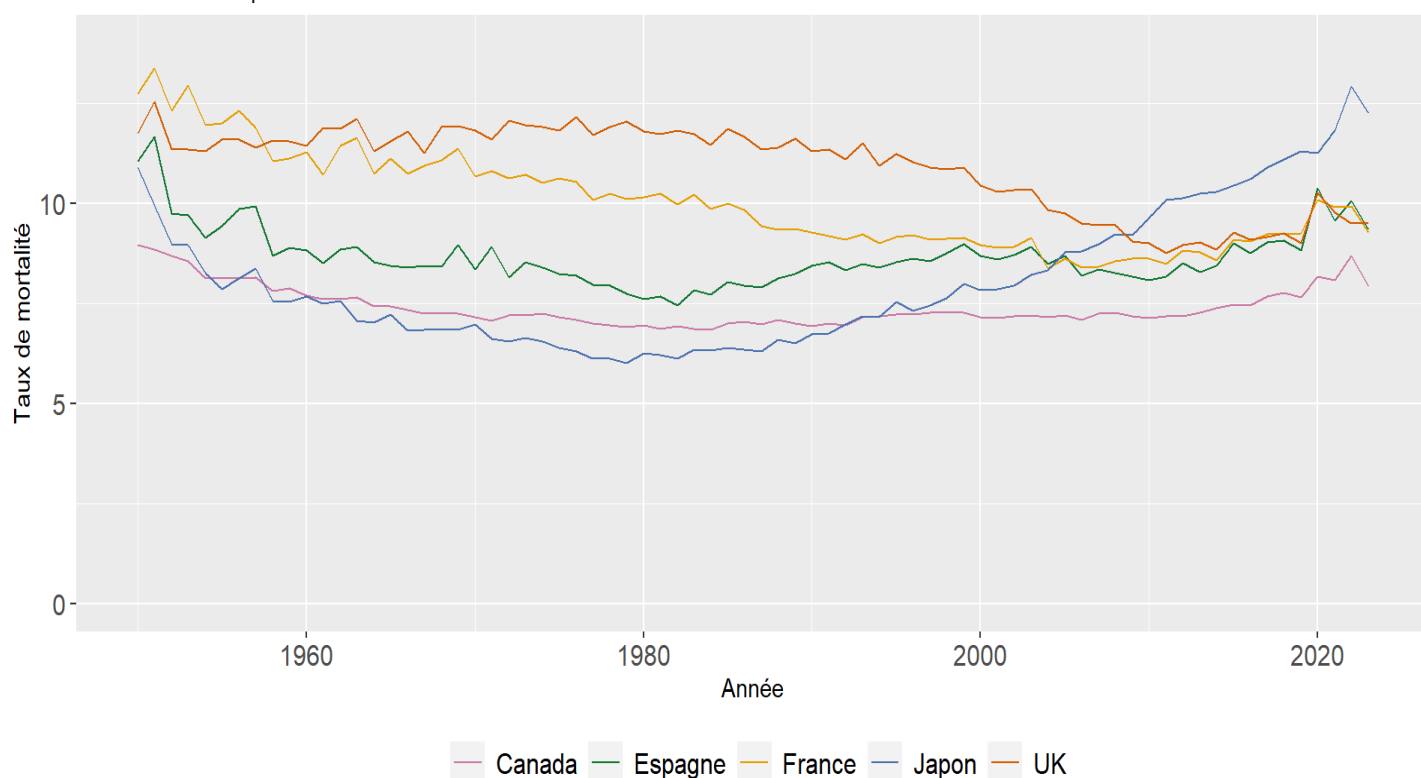


Figure 3 : Évolution des taux de mortalité entre 1950 et 2023 dans les 5 pays. Source : UN, World Population Prospect (2024)

Nous pouvons observer plusieurs différences entre les pays :

Le Japon présente une baisse de la mortalité entre l'année 1950 à 1980 puis une augmentation notable du taux de mortalité depuis les années 1980. Cela est en grande partie dû au vieillissement ainsi qu'au déclin démographique de sa population. En effet, bien que les Japonais bénéficient depuis des années d'une des espérances de vie les plus élevées au monde, le pays fait face à un taux de natalité très bas, qui n'a pas excédé 2 enfants par femme depuis 1975, atteignant même 1,26 en 2005⁷. Cette combinaison de faible natalité et de vieillissement de la population explique donc les tendances observées. De plus, nous pouvons observer une augmentation marquée après 2020, qui pourrait être attribuée en partie à l'effet de la pandémie.

Nous observons une tendance similaire en Espagne. En effet, le taux de mortalité a considérablement baissé depuis 1950 jusqu'aux années 1980, passant de 11,1 par mille en 1950 jusqu'à atteindre 7,5 en 1982. Nous observons ensuite une tendance vers la hausse en raison du vieillissement général de la population ainsi que le ralentissement de la croissance démographique. En effet, l'Espagne présente un taux de fécondité particulièrement faible (<1,2) et qui stagne quasiment depuis 1990.

La France et le Royaume-Uni montrent des tendances relativement identiques. Leurs taux de mortalité ont plutôt diminué progressivement depuis les années 1950. Une légère augmentation est visible en 2020, ce qui pourrait également être attribué à la COVID-19.

Le Canada montre une tendance plutôt constante du taux de mortalité suivie d'une légère augmentation plus récente. Comme pour les autres pays, nous pouvons supposer que l'augmentation après 2020 est liée à la pandémie. Nous pouvons aussi noter que le taux de mortalité du Canada est resté relativement bas par rapport aux autres pays, reflétant potentiellement des facteurs avantageux tels qu'un système de santé efficace et une population relativement jeune.

Tous ces pays, à l'exception du cas particulier du Japon et de l'Espagne, ont connu une tendance générale à la baisse ou constante de la mortalité depuis 1950. Cependant, une augmentation a été observée dans les 5 pays avec l'émergence de la pandémie en 2020 et qui n'aurait probablement pas eu lieu dans le scénario Baseline, c'est-à-dire en absence de la Covid-19.

⁷ Source : [World Bank](#)

3) Calibration des modèles

Dans cette section, nous allons procéder à la calibration des modèles de durée pour chaque pays et chaque sexe. L'objectif est de comparer, pour chaque cas, différents modèles parmi le modèle Lee-Carter poissonien, Renshaw-Haberman et Age-Period-Cohort, puis de sélectionner celui qui est le plus approprié pour la population étudiée. Le modèle choisi pour une population servira ensuite de base pour les projections futures de cette population.

Pour des raisons de simplification et afin d'éviter une longueur excessive, nous détaillerons d'abord la démarche utilisée pour les femmes en France. Ensuite, nous présenterons les résultats obtenus en appliquant cette même démarche aux autres populations, c'est à dire les hommes en France et les autres pays.

3.1) Calibration des modèles pour la France

Dans cette partie, notre objectif est de calibrer les modèles de mortalité pour la France en considérant deux scénarios : le scénario Baseline, pour les âges allant de 15 à 110 ans, calibré sur la période allant de 1990 à 2019, et le scénario Covid, pour les âges allant de 15 à 110 ans, calibré sur la période de 1990 à 2020, intégrant ainsi le choc de la pandémie de Covid-19 en année *jump-off*. Pour cela, nous avons à disposition les données HMD sur la taille de la population ainsi que les taux de mortalité observés par âge et par sexe de l'année 1816 à 2021 et pour les âges allant de 0 à 110.

À l'issue de cette partie, nous serons prêts à réaliser les projections des taux pour chaque scénario. Nous allons procéder comme suit :

- L'idée consiste d'abord à calibrer les différents modèles sur notre population sur les années choisies en fonction du scénario.
- Une fois les paramètres ajustés pour chaque modèle, nous allons réaliser une comparaison entre eux afin de choisir le plus adéquat. En effet, un modèle de durée qui est optimal pour une population particulière peut ne pas l'être pour une autre. Cela s'explique par les différences démographiques, culturelles, sociales et économiques existantes entre les populations et qui peuvent influencer leurs dynamiques de mortalité.
- Après avoir choisi le modèle optimal, nous allons déterminer la méthode de projection pour sa composante temporelle.

3.1.1) Résultat de calibration des modèles

Dans la littérature sur la mortalité, comme évoqué dans la partie précédente, l'ajustement des paramètres peut être réalisé soit par la méthode des moindres carrés, soit par la méthode du maximum de vraisemblance en utilisant l'approche log-poisson que nous avons décidé de retenir. En observant la structure mathématique de nos modèles, nous pouvons remarquer que ceux-ci peuvent être assimilés à des modèles linéaires généralisés : une variable réponse (la mortalité) expliquée linéairement par des variables explicatives, telle que l'âge, avec une fonction lien logarithmique. Cette observation a été formulée pour la première fois par Currie [2016], qui a montré que de nombreux modèles stochastiques de mortalité peuvent être considérés comme des modèles linéaires généralisés (GLM) ou des modèles non linéaires généralisés (GNM). Cette analogie a pour avantage de simplifier l'ajustement de ces modèles en utilisant les logiciels statistiques standard. En effet, Currie a démontré cette approche en ajustant plusieurs modèles stochastiques de mortalité dans R en utilisant la fonction standard `glm` ou la fonction `gnm` présentes dans le package **gnm** par Turner et Firth [2015].

Nous avons dans ce mémoire décidé d'utiliser le package **StMoMo** (*Stochastic Mortality Modelling*) sur R pour la calibration des modèles. Développé par Andres Villegas, il permet de lire et d'importer facilement les données HMD et propose une fonction générique `fit()` pour estimer les paramètres des modèles de mortalité. Cette fonction s'appuie sur le package **gnm** pour l'ajustement des paramètres en se basant sur les travaux de Currie mentionnés précédemment. En effet, il construit les formules équivalentes des modèles adaptées pour le package **gnm**, en les transformant sous forme de GLM lisibles par le package.

Un exemple de transformation réalisée pour le modèle Lee-Carter poissonien est le suivant :

"D/E ~ -1 + offset(o) + factor(x) + Mult(factor(x), factor(t), inst = 1)"

Avec :

- D/E : la variable cible $m_{x,t}$
- -1 : Exclut l'intercept de la formule
- Une variable `offset(o)` pour ajuster par rapport à l'exposition
- `factor(x)` : Ce terme est la variable catégorique âge, pour laquelle un coefficient est estimé pour chaque niveau.
- La fonction `Mult` (multiplication) ici indique une interaction entre l'âge x et l'année t , où l'on modélise une interaction multiplicative entre ces deux facteurs.
- L'argument `inst = 1` est un paramètre technique utilisé par le package **gnm** pour spécifier que l'interaction doit être modélisée de manière instantanée (sur la période t).

Une fois la transformation faite, l'ajustement peut être réalisé de manière classique en optimisant le maximum de vraisemblance.

Nous allons présenter, sous chaque scénario, les résultats d'ajustement de nos modèles.

Scénario Baseline

Estimations du modèle Lee-Carter poissonien

Pour rappel, le modèle de Lee-Carter propose la modélisation suivante du taux de mortalité :

$$\ln(\mu_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

Ainsi, le modèle de Lee-Carter se base sur trois paramètres principaux à estimer : deux paramètres qui varient selon l'âge (α_x et β_x), et une composante temporelle (κ_t) qui évolue au fil des années.

Les taux historiques des femmes françaises jusqu'à 2019 ont été utilisés pour calibrer le modèle Lee-Carter poissonien dans le scénario Baseline. Les 3 paramètres du modèle ont ainsi été estimés par la méthode du maximum de vraisemblance, après la transformation en GLM décrite précédemment.

En observant les paramètres estimés sur la figure 4, nous trouvons qu'ils paraissent conformes aux résultats classiquement obtenus :

- La courbe du paramètre α_x a la forme d'une droite croissante, ce qui reflète l'augmentation du taux de mortalité avec l'âge. Nous remarquons également un léger pic au niveau des jeunes adultes (18-25 ans), ce qui pourrait souvent être attribué à une hausse des prises de risques et des comportements dangereux à cet âge⁸.
- La courbe de la composante temporelle κ_t est quant à elle de la forme d'une droite décroissante reflétant ainsi l'allongement de l'espérance de vie au fil du temps, cela indique donc une amélioration générale de la mortalité.
- Enfin, la courbe du paramètre β_x mesure la sensibilité de la mortalité à la composante temporelle κ_t . En d'autres termes, un β_x élevé (en valeur absolue) pour un âge donné x indique que la mortalité à cet âge est davantage sensible aux évolutions de la mortalité au fil du temps, souvent dues à des progrès médicaux et/ou technologiques. Nous pouvons remarquer que la courbe présente des fluctuations, avec une sensibilité plus importante pour les jeunes âges et personnes âgées, ce qui suggère alors que ces groupes sont plus touchés par les changements au cours du temps.

⁸ Source : [Remund et al. \[2021\]](#)

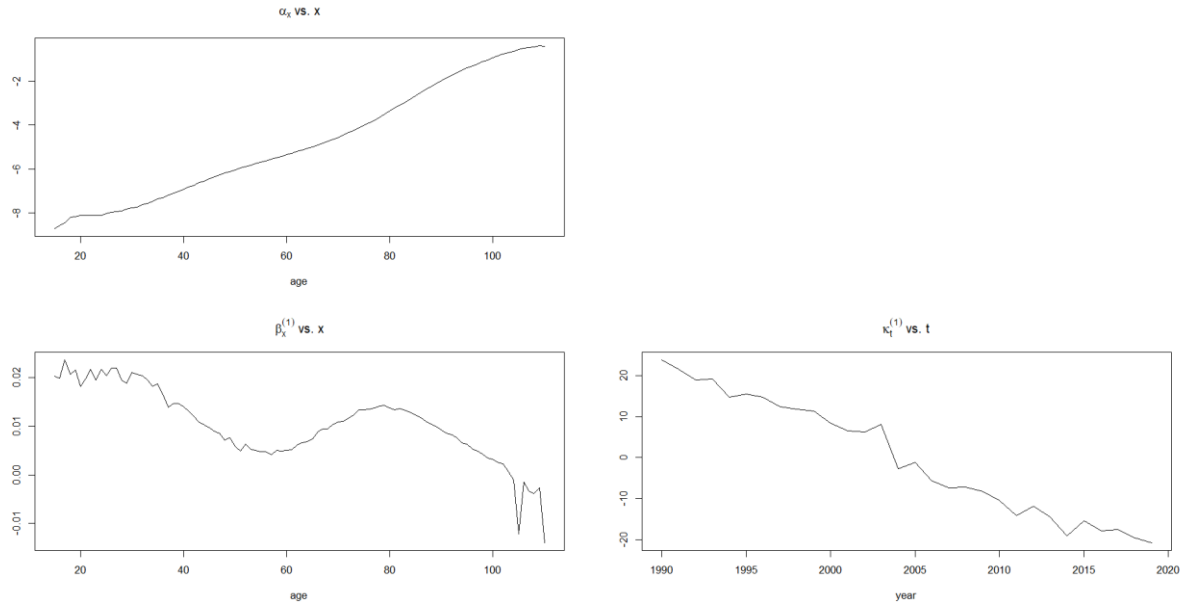


Figure 4 : Estimation des paramètres du modèle Lee-Carter dans le scénario Baseline pour les femmes en France

Estimations du modèle Renshaw-Haberman (RH)

Le modèle RH est quant à lui plus complexe, avec plus de paramètres à estimer afin de prendre en compte l'effet de cohorte :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t + \beta_x^{(0)}\gamma_{t-x} + \epsilon_{xt}$$

On reconnaît les premiers paramètres du modèle LC, auxquels s'ajoutent une autre composante âge $\beta_x^{(0)}$ et la composante cohorte γ_{t-x} .

Des simplifications du modèle ont été proposées par Renshaw et Haberman en ajustant un des termes modulant l'âge, ce qui peut améliorer la stabilité et l'interprétabilité du modèle tout en conservant sa capacité à capturer les tendances de mortalité liées à l'âge, au temps et aux cohortes. Une sous-structure particulièrement intéressante est obtenue en fixant $\beta_x^{(0)} = 1$, donnant ainsi :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t + \gamma_{t-x} + \epsilon_{xt}$$

Haberman et Renshaw ont proposé cette structure comme une version simplifiée, capable de réduire certains problèmes de stabilité que nous avons mentionnés dans la première section du mémoire. Nous adopterons donc cette version pour la calibration de notre modèle RH.

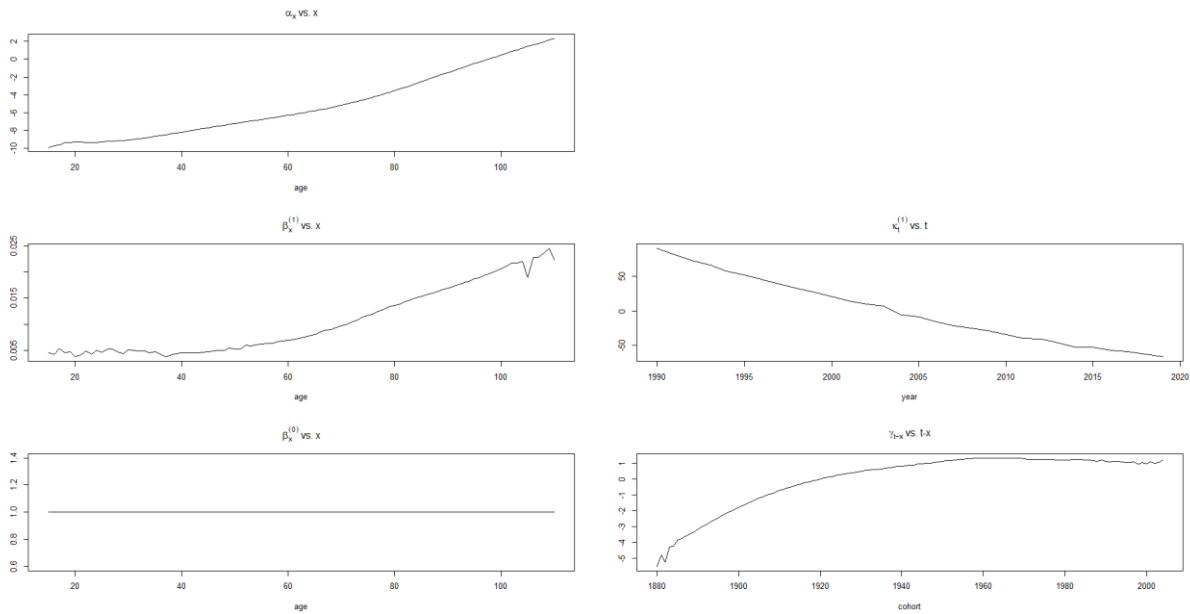


Figure 5 : Estimation des paramètres du modèle RH dans le scénario Baseline pour les femmes en France

Nous pouvons remarquer sur la figure 5 que les courbes des paramètres α_x et k_t présentent des allures similaires à celles du modèle Lee-Carter. De plus, nous remarquons que les âges élevés représentent une sensibilité β_x relativement forte aux évolutions de la mortalité au fil du temps.

Enfin, nous pouvons observer dans le dernier graphique de la figure 5, en bas à droite, le paramètre propre au modèle RH qui prend en compte l'effet de cohorte. Il est clair que cet effet est bien présent et varie en fonction de chaque génération.

Estimations du modèle APC

Le modèle Age-Period-Cohort correspond à une variation du modèle de Renshaw-Haberman où l'on a imposé les contraintes suivantes :

$$\beta_x^{(1)} = 1 \quad \beta_x^{(0)} = 1$$

Donnant ainsi :

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

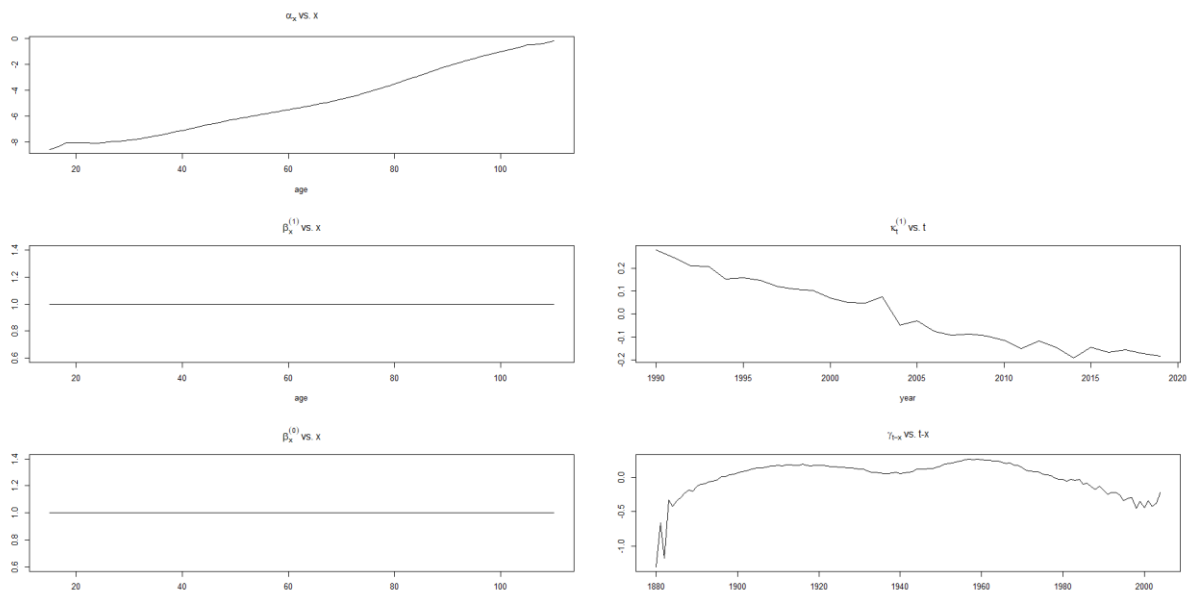


Figure 6 : Estimation des paramètres du modèle APC dans le scénario Baseline pour les femmes en France

En observant les paramètres estimés sur la figure 6, nous pouvons dire que les courbes des paramètres d'âge α_x et de période k_t conservent des allures similaires aux modèles LC et RH, reflétant ainsi les phénomènes connus d'augmentation de la mortalité avec l'âge ainsi que l'allongement de l'espérance de vie.

De plus, nous observons un effet de cohorte plus marqué pour les générations les plus anciennes et légèrement présent pour les plus récentes. Cet effet est moins prononcé pour les autres générations, où les paramètres γ_{t-x} sont proches de 0.

Scénario Covid

Dans cette partie, nous allons examiner l'effet du scénario Covid sur les paramètres du modèle. Plus précisément, nous cherchons à déterminer si l'intégration de l'année 2020, incluant le choc pandémique comme année *jump-off* dans la calibration du modèle modifie considérablement ou non les paramètres par rapport au scénario Baseline.

Cette comparaison sera réalisée pour les trois modèles pour les femmes en France. Pour les autres pays et pour les hommes, par souci de lisibilité, nous limiterons cette analyse uniquement pour le modèle sélectionné.

Estimations du modèle Lee-Carter poissonien

En superposant les paramètres du modèle sous les deux scénarios (figure 7), nous observons que les paramètres du modèle Lee-Carter sont presque identiques dans les deux cas.

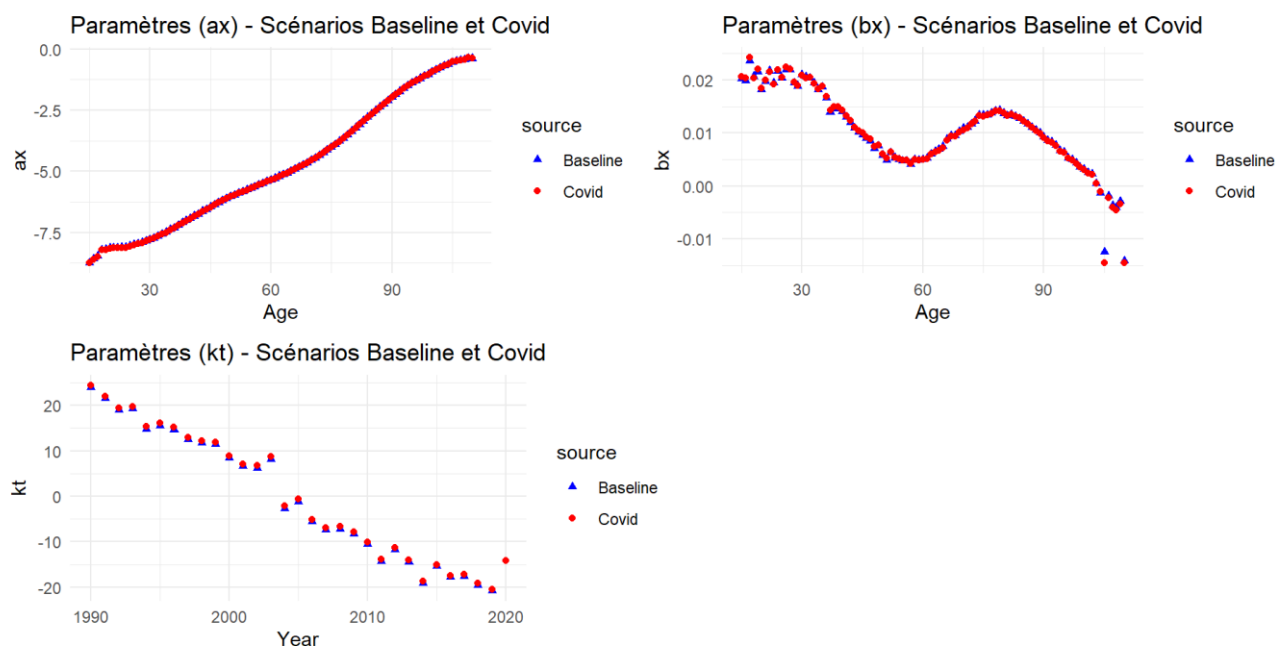


Figure 7 : Lee-Carter, Comparaison des paramètres estimés sous les deux scénarios pour les femmes en France

En revanche, l'effet du choc causé par la Covid est capturé par la composante temporelle k_{2020} . En effet, nous constatons que pour l'année 2020, la composante k_t dévie de la tendance générale décroissante observée au fil des années. Au lieu de continuer à diminuer, comme cela a été le cas dans les années précédentes, k_t présente un pic vers le haut en passant de -20,53 à -14,26. Ce saut indique une perturbation dans la tendance normale, reflétant l'augmentation soudaine des taux de mortalité due à la pandémie capturée par le modèle.

Pour illustrer l'impact de ce saut, chez les femmes françaises de 65 ans et toutes choses étant égales par ailleurs, cette augmentation de k_{2020} a conduit à une hausse dans la mortalité calibrée, passant de 0,0051 en 2019 à 0,0054 en 2020, soit une augmentation de 5,1% en un an. Pour les femmes de 15 ans, cette hausse atteint 14 %. Cela n'est pas choquant puisque la mortalité des femmes jeunes de notre population est plus sensible par rapport à l'évolution générale k_t , comme on peut le voir sur la courbe du paramètre β_x . En moyenne, l'augmentation est de 6,9 % pour l'ensemble de la population, toutes choses étant égales par ailleurs.

Cette augmentation de k_{2020} , a non seulement impacté les mortalités ajustées entre 2019 et 2020, mais elle influencera également les projections futures de la série.

Estimations du modèle Renshaw-Haberman

Nous avons réalisé la même analyse de comparaison pour le modèle Renshaw-Haberman (Figure 8).

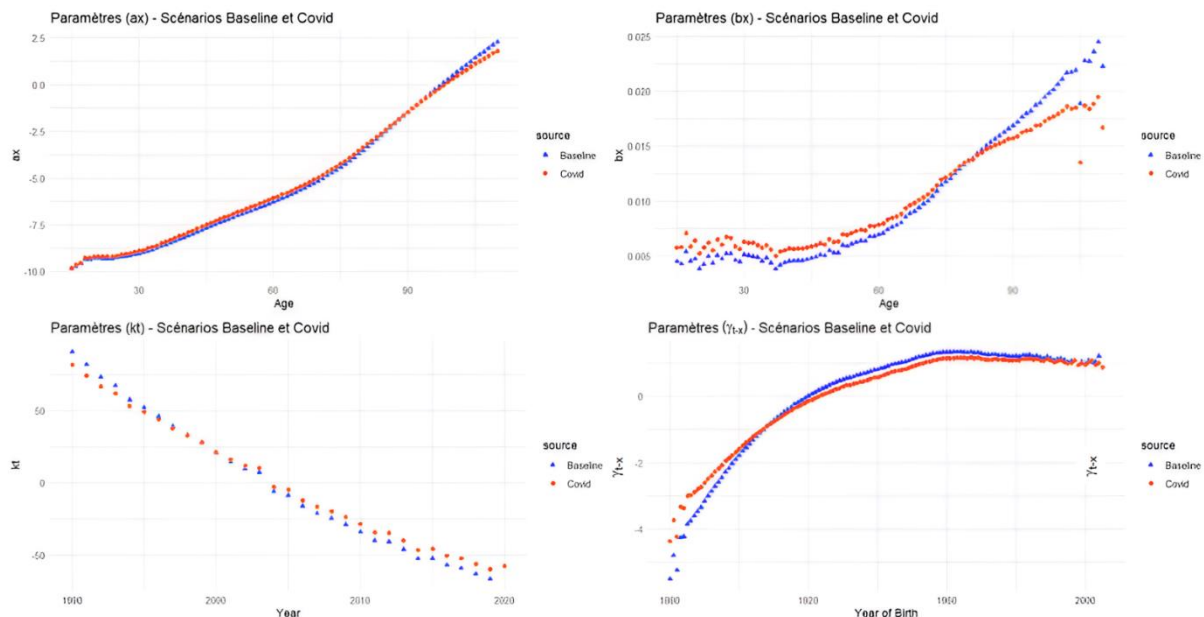


Figure 8 : Renshaw-Haberman, Comparaison des paramètres estimés sous les deux scénarios pour les femmes en France

Pour la composante α_x , on observe que les courbes pour les scénarios Baseline et Covid sont quasiment identiques. Cela signifie que l'effet de la composante âge sur la mortalité reste inchangé, même en tenant compte de la pandémie.

En revanche, en regardant la composante β_x , on déduit que la sensibilité de la mortalité à l'âge x par rapport à l'évolution générale k_t a été impactée. On observe un décalage léger vers le haut pour les âges en dessous de 80 ans, ce qui veut dire qu'elles sont plus sensibles à l'évolution générale dans le scénario Covid. Un choc traduit par k_t aura alors un peu plus d'effet sur la mortalité de ces femmes que dans le modèle Baseline, toutes choses étant égales par ailleurs. En revanche, on observe une diminution de la sensibilité pour les âges élevés (>80 ans). Toutes choses étant égales par ailleurs, ces changements dans les coefficients β_x ont entraîné :

- Une augmentation moyenne entre les deux scénarios de 16% de la mortalité en 2019 pour la tranche d'âge de 80 à 110 ans.
- Une diminution de 6,46% de la mortalité en 2019 pour la tranche d'âge de 15 à 80 ans.
- Une diminution de 0,74% de la mortalité pour l'ensemble de la population.

Concernant la composante cohorte γ_{t-x} , on peut observer que les deux courbes ont globalement la même allure mais avec quelques divergences. Plus particulièrement, on voit une augmentation des paramètres pour les générations très anciennes entre 1880 à 1910 par rapport au scénario Baseline. Cette augmentation pourrait refléter une anticipation d'une plus grande vulnérabilité à la pandémie pour ces cohortes, qui auraient atteint une tranche d'âge (110 à 140 ans en 2020) pour laquelle les

risques de mortalité liés à la Covid-19 sont particulièrement élevés. Cependant, il est important de souligner que les données correspondantes à ces générations sont très limitées, ce qui entraîne une variance élevée dans les estimations, car ces individus sont relativement moins nombreux et meurent avant d'atteindre un âge aussi avancé. Ainsi, dans ce mémoire, nous ne nous intéresserons pas à cette tranche d'âge et notre analyse s'arrêtera à 110 ans.

Enfin, concernant la composante temporelle k_t , nous ne remarquons pas de changement significatif entre les deux scénarios. Cependant, nous retrouvons un pic à l'année 2020 mais relativement moins prononcé que celui du modèle Lee-Carter.

Modèle Age-Period-Cohort

Nous avons enfin réalisé la même démarche pour le modèle APC et nous avons trouvé les résultats suivants sur la figure 9 :

- Les paramètres d'âge α_x et de cohorte γ_{t-x} n'ont pas été modifiés par le scénario Covid : les courbes ont conservé des valeurs quasiment identiques.
- Les valeurs de la composante temporelle k_t sont restées globalement inchangées entre les deux scénarios, mais nous retrouvons un saut en 2020. Celui-ci, toutes choses étant égales par ailleurs, a permis une augmentation de 7% de la mortalité entre 2019 et 2020.

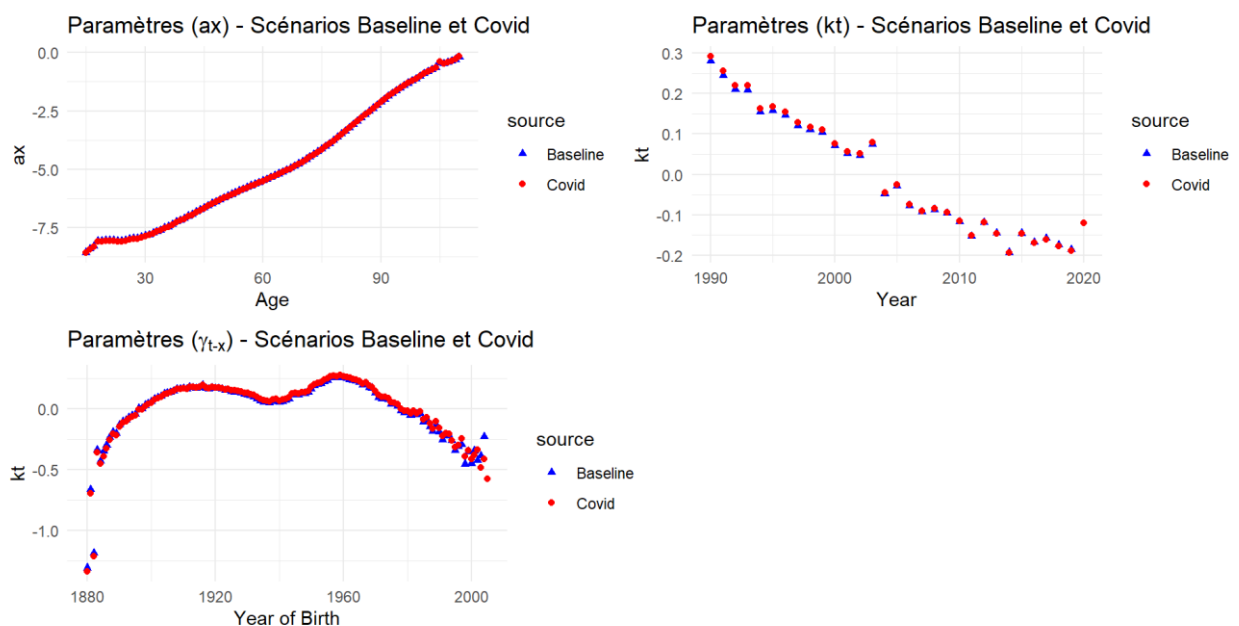


Figure 9 : Age-Period-Cohort, Comparaison des paramètres estimés sous les deux scénarios pour les femmes en France

3.1.2) Analyse des résidus

La qualité de l'ajustement d'un modèle est généralement analysée en examinant les résidus du modèle. On s'intéresse aux résidus de déviance ajustés définis par :

$$r_{xt} = \text{sign}(d_{xt} - \widehat{d}_{xt}) \sqrt{\frac{\text{dev}(x,t)}{\widehat{\phi}}}, \quad \widehat{\phi} = \frac{D(d_{xt}, \widehat{d}_{xt})}{K-v}$$

Ou :

$$\text{dev}(x,t) = 2[d_{xt} \log\left(\frac{d_{xt}}{\widehat{d}_{xt}}\right) - (d_{xt} - \widehat{d}_{xt})]$$

Avec :

$$\left\{ \begin{array}{l} d_{xt} \text{ la valeur observée à l'âge } x \text{ et au temps } t \\ \widehat{d}_{xt} \text{ la valeur prédite à l'âge } x \text{ et au temps } t \\ D(d_{xt}, \widehat{d}_{xt}) \text{ la déviance totale} \\ K \text{ le nombre d'observations} \\ v \text{ le nombre de paramètres} \end{array} \right.$$

Les résidus de déviance ajustés peuvent être considérés comme un indicateur pertinent pour évaluer la qualité des modèles. En effet, ils permettent de mesurer l'écart entre les données réelles et les données prédites, tout en prenant en compte la variabilité de ces données ainsi que la complexité du modèle reflétée par le nombre de paramètres.

Nous allons alors tracer ces résidus pour chaque modèle sous les deux scénarios afin de les examiner et de les comparer entre eux.

Il est important de noter que la famille des modèles Lee-Carter et ses variantes supposent la normalité des résidus. Ainsi, les résidus devraient idéalement être distribués aléatoirement autour de 0. Des schémas réguliers, c'est à dire des motifs non aléatoires, dans les résidus indiquent l'incapacité du modèle à décrire correctement toutes les caractéristiques des données.

Lee-Carter poissonien

En traçant les résidus du modèle Lee-Carter pour les deux scénarios, on remarque qu'ils présentent la même tendance diagonale (figure 10). Cela indique que le modèle n'est pas capable de décrire correctement toute la structure des données.

Nous pouvons déduire, dans un premier temps, que la performance du modèle n'a pas changé de manière drastique entre les deux cas, puisque les allures des résidus obtenues sont quasi similaires.

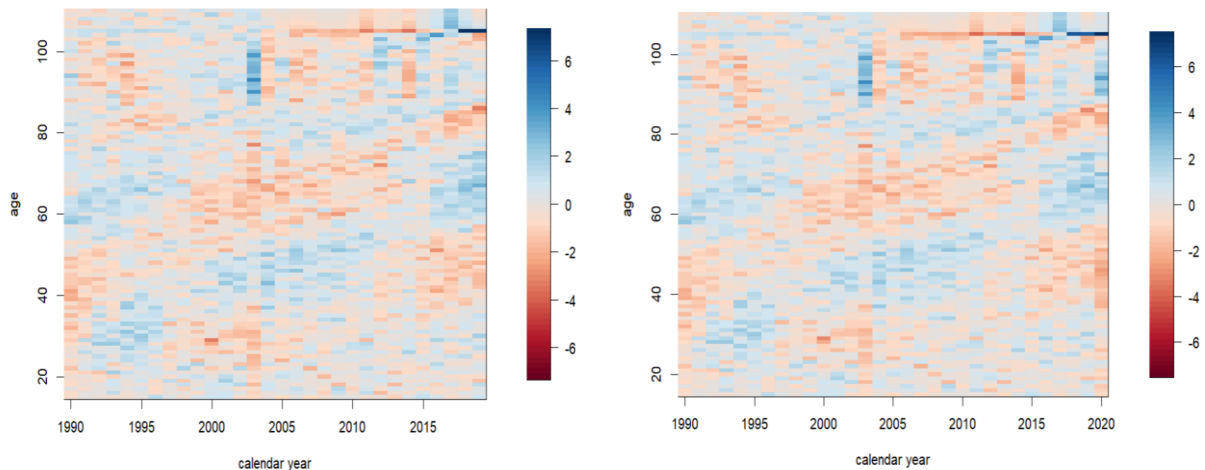


Figure 10 : Les résidus du modèle Lee-carter pour le scénario Baseline et le scénario Covid respectivement sont quasi similaires

Ensuite, nous avons tracé sur la figure 11 la distribution des résidus, séparée en fonction de l'âge, le temps et l'année de naissance afin de voir quel effet contient un « motif » ou une information non prise en compte par le modèle. Les graphiques des deux scénarios étant similaires, nous présentons un seul scénario (Baseline) pour plus de lisibilité.

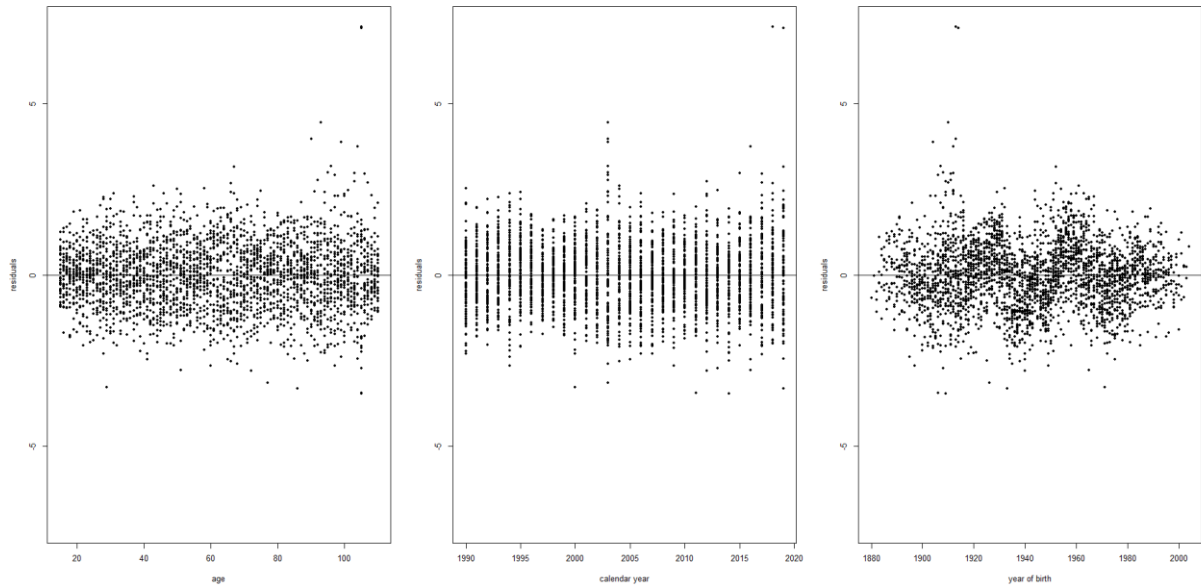


Figure 11 : Analyse des résidus pour le modèle Baseline Lee-Carter

En observant les deux premiers graphiques, les résidus semblent aléatoires par rapport à l'âge et au temps. En revanche, nous avons retrouvé une tendance dans le graphique à droite, qui illustre les résidus en fonction de l'année de naissance. Cela est probablement dû au fait que le modèle Lee-Carter ne prend pas en compte l'effet de cohorte, qui est contenu donc dans ces résidus.

Renshaw-Haberman

Le modèle Renshaw-Haberman, avec sa capacité à prendre en compte, en plus des effets de l'âge et du temps, cet effet de cohorte permet d'obtenir des résidus aléatoires, sans facteur caché ou non expliqué, et toujours très similaires entre les deux scénarios comme observé ci-dessous :

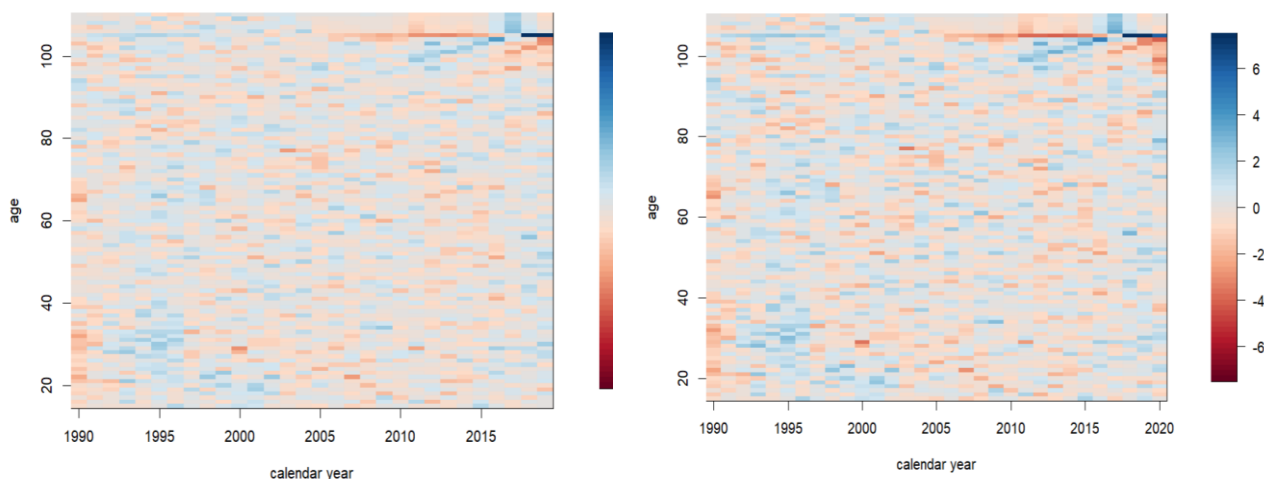


Figure 12 : Les résidus du modèle Renshaw-Haberman pour le scénario Baseline et le scénario Covid respectivement sont quasi similaires et semblent aléatoires

En examinant plus en détail les graphiques des distributions des résidus en fonction de l'âge, du temps et de l'année de naissance sur la figure 13, nous pouvons déduire que la tendance observée en fonction de l'année de naissance dans le modèle Lee-Carter n'est plus présente et que les résidus semblent aléatoires en fonction des trois effets.

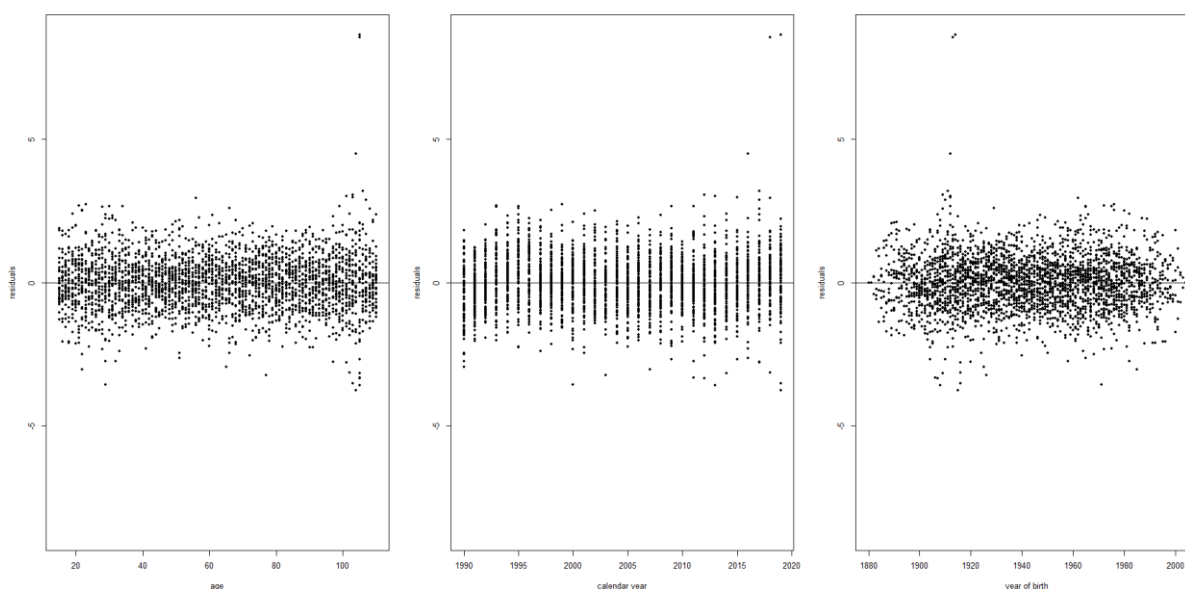


Figure 13 : Analyse des résidus du modèle Baseline Renshaw-Haberman

Le modèle RH semble donc offrir une approche plus complète pour capturer les dynamiques générationnelles des tendances de mortalité.

Age-Period-Cohort

Enfin, en traçant les résidus du modèle Age-Period-Cohort (figure 14), nous observons des regroupements de résidus présentant des valeurs similaires (de même couleur). Cela pourrait potentiellement être dû à l'incapacité du modèle à capturer la sensibilité par âge en fonction de l'évolution générale de la mortalité.

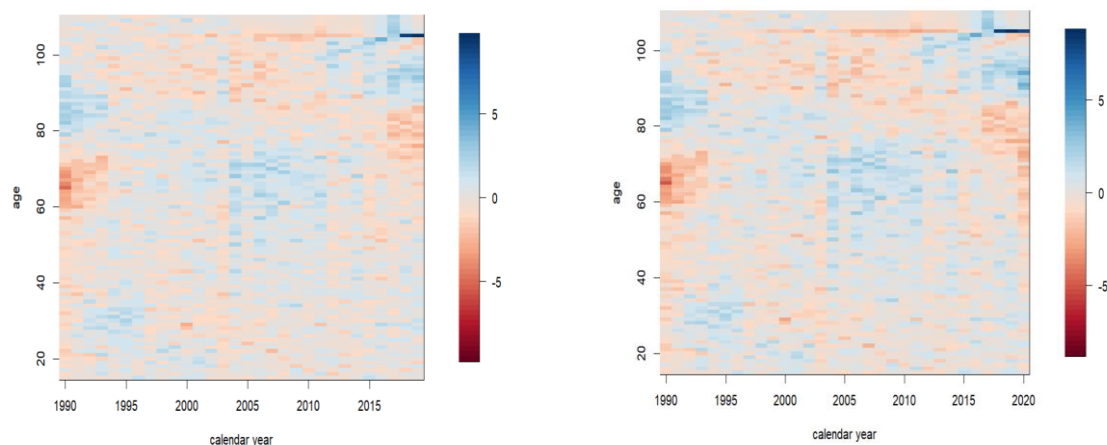


Figure 14 : Les résidus du modèle Age-Period-Cohort pour le scénario Baseline et le scénario Covid sont quasi similaires

En revanche, en observant la figure 15, nous ne retrouvons pas de motif particulier dans la distribution des résidus par facteur.

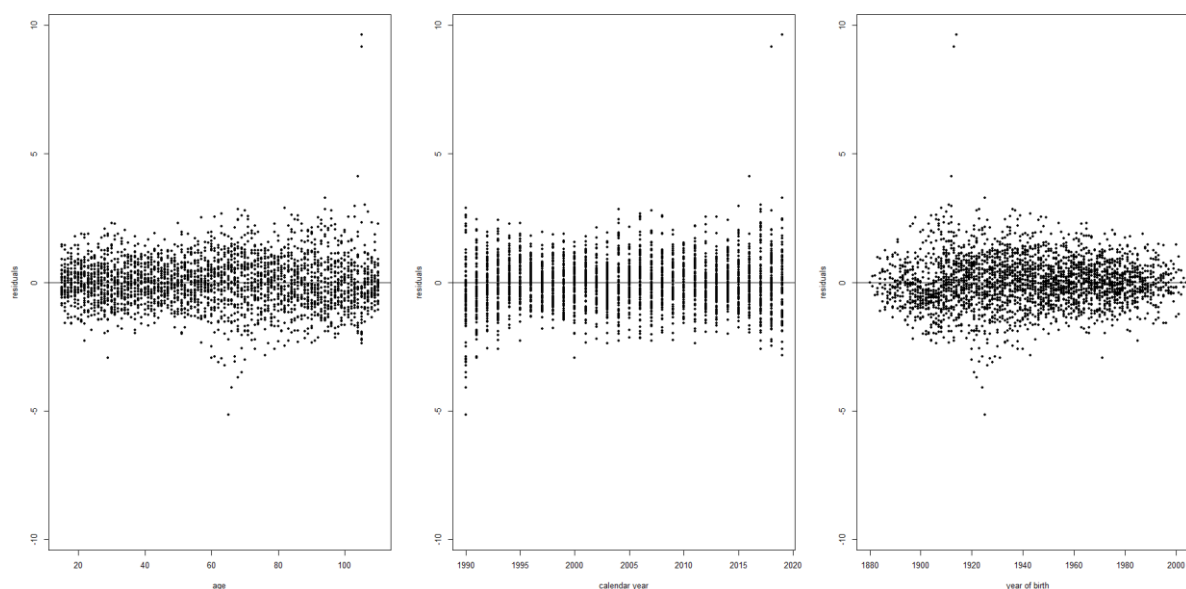


Figure 15 : Analyse des résidus du modèle Baseline Age-Period-Cohort

3.1.3) Sélection des modèles

Dans cette partie, l'objectif est d'abord de choisir le meilleur modèle, c'est à dire le plus adapté pour notre population, dans le scénario Baseline en utilisant des critères de sélection. L'idée ensuite est de réaliser ce même choix du modèle optimal mais cette fois-ci sous le scénario Covid. Cela permet d'évaluer si le modèle Baseline initialement choisi reste le mieux adapté après le choc de la Covid. En d'autres termes : **un modèle de longévité pertinent pour une population l'est-il toujours après le choc de la pandémie ?**

Scénario Baseline

Nous avons commencé par choisir le modèle de durée le plus adapté pour les femmes en France dans le scénario Baseline. Pour cela, nous avons comparé les modèles LC poissonien, RH et APC calibrés sur la population concernée pour la période 1990-2019.

Afin d'évaluer la performance des modèles, nous avons utilisé les critères d'information d'Akaike (AIC) et Bayésien (BIC). Ces critères permettent de comparer les modèles en prenant compte deux aspects :

- 1) D'une part leur capacité à s'ajuster aux données,
- 2) D'une autre part leur complexité, en pénalisant les modèles trop complexes en fonction du nombre de paramètres estimés.

Cela permet alors d'avoir un indicateur favorisant les modèles qui équilibrent la précision de l'ajustement et la simplicité du modèle.

Le critère AIC

Le critère d'information Akaike – *Akaike information criterion* (AIC) – est utile pour les modèles estimés par la méthode du maximum de vraisemblance. Il s'écrit comme ci-dessous :

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

Avec k le nombre de paramètres à estimer et la fonction L le maximum de la fonction de vraisemblance du modèle.

L'objectif est de minimiser l'AIC, puisque le critère pénalise la complexité du modèle via le terme $2k$, qui augmente avec le nombre de paramètres, tout en favorisant une bonne qualité d'ajustement avec le terme $-2 \times \log (L)$. En effet, plus le modèle s'ajuste bien aux données, plus le maximum de vraisemblance est élevé, ce qui conduit à la réduction du terme $-2 \times \log (L)$.

Le critère BIC

Le critère d'information bayésien – *Bayesian information criterion* (BIC) – est également utilisé pour les modèles estimés par la méthode du maximum de vraisemblance. Il s'écrit comme ci-dessous :

$$BIC = k \times \ln(N) - 2 \times \log (L)$$

Avec la fonction L le maximum de la fonction de vraisemblance du modèle, N le nombre d'observations, et k le nombre de paramètres à estimer.

De la même manière que le critère AIC, nous cherchons le modèle avec le BIC le moins élevé. Cependant, nous pouvons remarquer que le critère BIC est plus pénalisant que le critère AIC car la pénalité dépend non seulement du nombre de paramètres k , mais également de la taille de l'échantillon N .

Il n'existe pas de critère universellement meilleur, le choix doit dépendre de l'objectif de l'analyse. Par exemple, l'AIC est préférable lorsque l'on recherche un modèle offrant un bon compromis entre précision et complexité, tout en privilégiant la prédiction sur de nouvelles données. En revanche, si l'objectif principal est la parcimonie, le critère BIC serait préférable puisqu'il possède une pénalité plus stricte. Dans une optique de robustesse, il est courant de calculer les deux critères même s'ils donnent souvent des résultats similaires.

Nous avons obtenu les résultats suivants :

Modèles	Nombre de paramètres estimés	AIC	BIC
Lee-Carter	220	31 922	33 234
Renshaw-Haberman	344	29 315	31 367
APC	248	31 853	33 332

Même en ayant le plus grand nombre de paramètres estimés, nous remarquons que le modèle Renshaw-Haberman est meilleur pour cette population au sens des critères AIC et BIC, nous décidons alors de le garder pour la suite.

Scénario Covid

Nous avons effectué la même comparaison avec les modèles prenant en compte le choc de la COVID afin de voir si le modèle Renshaw-Haberman était toujours le plus pertinent pour notre population. Nous obtenons les résultats suivants :

Modèles	Nombre de paramètres estimés	AIC	BIC
Lee-Carter	221	33 470	34 796
Renshaw-Haberman	346	30 455	32 531
APC	250	33 409	34 909

Nous remarquons que le modèle Renshaw-Haberman reste préférable aux modèles Lee-Carter et APC, même après l'intégration du choc.

3.1.4) *Choix de la méthode de projection de la composante temporelle*

Une fois le modèle choisi, nous allons nous concentrer sur la composante temporelle k_t . Comme vu précédemment, les paramètres α_x et β_x de notre modèle ne dépendent que de l'âge et ne varient pas en fonction du temps. En revanche, la composante temporelle k_t ainsi que la composante cohorte γ_{t-x} doivent être projetées dans le temps afin de réaliser des nouvelles prédictions de mortalité. La composante k_t est cruciale dans notre modèle car elle capture les effets du temps sur les taux de mortalité. En d'autres termes, k_t représente les tendances temporelles des variations de la mortalité au fil des années. Comme nous avons vu dans la littérature, la projection de la composante k_t repose souvent sur des modèles de séries temporelles, nous allons donc en comparer plusieurs et en choisir une.

Les deux méthodes étudiées sont la méthode de marche aléatoire multivariée avec dérive MAD (méthode 1) et la méthode ARIMA (méthode 2).

La méthode 1 est utile pour modéliser des séries temporelles lorsque nous supposons qu'il existe une tendance générale et systématique de mortalité. La dérive permet d'intégrer une composante représentant des améliorations ou des détériorations prévisibles dans les taux de mortalité tout en intégrant une composante prenant en compte les fluctuations aléatoires.

La méthode 2, quant à elle, est plus flexible et permet de capter des tendances plus complexes de mortalité. Elle est particulièrement utile lorsque les séries temporelles présentent des motifs non linéaires, des cycles saisonniers ou des fluctuations irrégulières.

Afin de choisir le meilleur processus ARIMA pour notre série, nous avons utilisé les critères AIC et BIC avec la fonction `auto.arima` sur R et nous avons trouvé que le modèle ARIMA optimal au sens des deux critères est ARIMA(0,2,2).

Il est important de noter que la méthode de projection de la composante temporelle k_t , que ce soit ARIMA ou une marche aléatoire avec dérive, peut avoir un impact sur la trajectoire future de la série. On constate, dans la figure 16, que la série k_t pour le scénario Baseline diminue plus rapidement d'ici 2050 lorsque la méthode de la marche aléatoire avec dérive optimale est appliquée, comparativement à la méthode optimale ARIMA(0,2,2). Cette observation se maintient également pour le scénario Covid. Cela impactera par conséquent les mortalités projetées. Il est donc crucial de sélectionner la méthode de projection la plus adéquate.

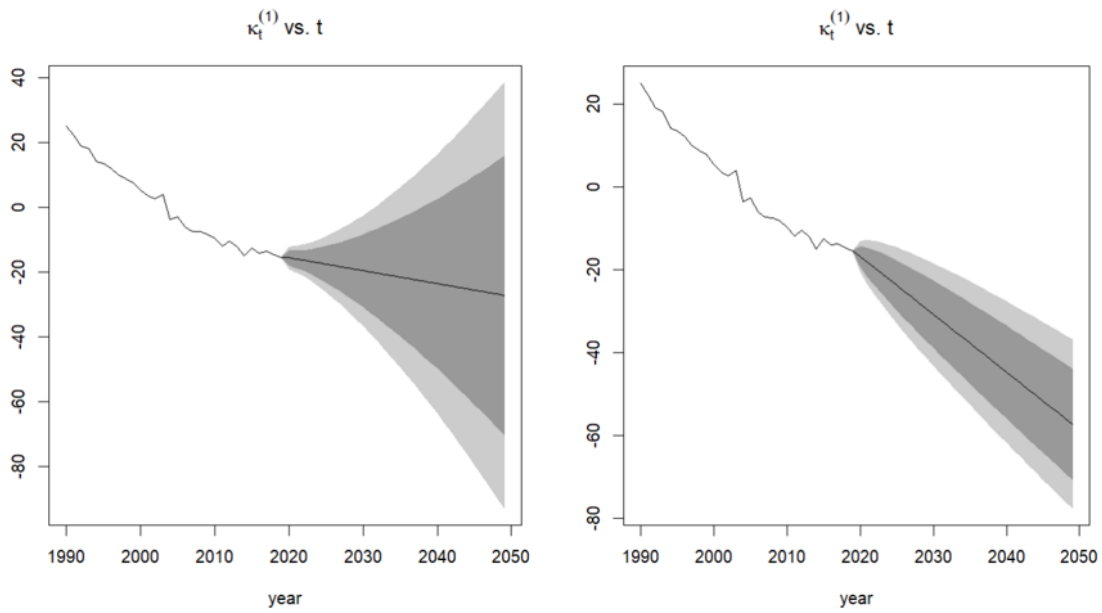


Figure 16 : Comparaison des trajectoires de la composante k_t projetées avec la méthode ARIMA(0,2,2) et la marche aléatoire avec dérive respectivement pour le scénario Baseline

Afin de pouvoir comparer les méthodes de projection de k_t entre elles, nous avons décidé de comparer leurs erreurs quadratiques moyennes (RMSE). Le RMSE, ou *Root Mean Square Error* est une mesure utilisée pour évaluer la précision d'un modèle de prédiction ou de régression. Elle calcule l'écart moyen quadratique entre les valeurs observées et les valeurs prédites par le modèle. Il s'agit de la racine carrée de la moyenne des carrés des erreurs (différences entre les valeurs observées et prédites).

On a :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (k_i - \hat{k}_i)^2}$$

Avec n le nombre de données prédites, k_i la valeur observée et \hat{k}_i la valeur prédite.

Pour cela, nous avons divisé notre série en deux ensembles :



avec T l'année *jump-off* et $1990 < t < T$

L'ensemble train détermine la partie à utiliser dans la modélisation, il s'agit de la série dans l'intervalle de temps $[1990 ; t]$. Une fois le modèle entraîné, nous allons prédire la série jusqu'à l'année T . Ensuite nous allons calculer le RMSE en comparant les projections effectuées avec l'ensemble test qui contient les données réelles de la série.

Après avoir calculé les RMSE des deux modèles, nous avons alors décidé de retenir le modèle ARIMA pour la projection des k_t pour la population féminine Française.

3.2) Calibration et résultats de modélisation pour chaque pays

Nous avons suivi la même démarche de calibration et de sélection des modèles pour les hommes en France ainsi que dans les 4 autres pays : l'Espagne, le Japon, le Canada et le Royaume-Uni. Nous allons présenter dans cette partie les résultats obtenus par pays et par sexe.

3.2.1) Sélection des modèles par pays

Nous avons commencé par réaliser la sélection du modèle comme nous l'avons fait pour les femmes en France. Comme nous l'avons mentionné précédemment et comme l'avait également remarqué Currie, nous avons rencontré des problèmes de convergence lors de l'utilisation du package **gnm** pour ajuster le modèle Renshaw-Haberman. Ces difficultés rendaient impossible l'ajustement du modèle et ne permettaient pas d'obtenir des estimations fiables. Dans ces cas, nous avons donc décidé de choisir un autre modèle, en sélectionnant celui ayant le meilleur AIC/BIC parmi les modèles Lee-Carter (LC) et APC.

Ci-dessous les résultats obtenus :

Pour les femmes :

Critères d'information pour les modèles Baseline (1990-2019) :

Pays	Espagne		Japon		Canada		Royaume-Uni	
Critère	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
LC	30 014	31 326	37 327	38 639	27 433	28 746	32 943	34 256
RH	-	-	31 078	33 130	-	-	28 619	30 672
APC	28 319	29 799	51 563	53 042	26 655	28 134	29 380	30 860

Critères d'information pour les modèles Covid (1990-2020) :

Pays	Espagne		Japon		Canada		Royaume-Uni	
Critère	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
LC	31 408	32 733	38 634	39 960	28 528	29 853	34 210	35 536
RH	-	-	32 299	34 375	-	-	29 725	31 800
APC	29 398	30 897	53 357	54 957	27 685	29 184	30 618	32 118

Nous remarquons que pour chaque pays, le modèle optimal selon les critères AIC et BIC reste identique dans les deux scénarios. Cela suggère que le choix du modèle pour nos populations n'a pas été influencé par l'intégration de la pandémie en année *jump-off*. Ainsi, nous avons retenu le modèle Age-

Period-Cohort pour les femmes en Espagne et le Canada, et le modèle Renshaw-Haberman pour les femmes au Japon et au Royaume-Uni.

Pour les hommes :

Critères d'information pour les modèles Baseline (1990-2019) :

Pays	Espagne		Japon		Canada		Royaume-Uni		France	
Critère	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
LC	33 430	34 742	43 943	45 255	29 645	30 956	33 087	34 399	35 991	37 303
RH	-	-	-	-	26 836	28 886	-	-	29 717	31 769
APC	30 911	32 390	46 128	47 607	28 143	29 621	30 942	32 414	31 527	33 006

Critères d'information pour les modèles Covid (1990-2020) :

Pays	Espagne		Japon		Canada		Royaume-Uni		France	
Critère	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
LC	35 344	36 669	45 944	47 269	31 569	32 893	34 485	35 810	37 781	39 106
RH	-	-	-	-	27 827	29 901	-	-	30 733	32 808
APC	32 160	33 659	47 478	48 977	29 703	31 201	32 449	33 941	32 927	34 426

Comme pour les femmes, les modèles optimaux pour chaque pays, selon les critères AIC et BIC, ont été identiques chez les hommes dans les deux scénarios. Nous avons décidé de retenir le modèle Age-Period-Cohort pour l'Espagne et le Royaume-Uni, le modèle Renshaw-Haberman pour le Canada et la France et le modèle Lee-Carter poissonien pour le Japon.

Comme nous pouvons le constater, les modèles retenus pour les femmes ne sont pas toujours les mêmes que ceux sélectionnés pour les hommes dans chaque pays. Cela justifie la pertinence de la segmentation réalisée par sexe, en raison des différences dans leurs caractéristiques face à la mortalité.

Méthode de projection choisie pour la composante temporelle k_t

Comme pour les femmes en France, nous avons comparé la méthode de marche aléatoire avec drift (MAD) et la méthode ARIMA pour projeter la composante temporelle k_t , en se basant sur le RMSE. Le modèle ARIMA optimal a été sélectionné en fonction du critère AIC.

Les méthodes optimales retenues sont présentées ci-dessous :

Femmes

Pays	Espagne	Japon	Canada	Royaume-Uni
Baseline	ARIMA(0,1,1)	MAD(-1,33)	MAD(-1,08)	ARIMA(1,1,0)
Covid	ARIMA(2,2,1)	MAD(-1,4)	ARIMA(1,1,0)	ARIMA(0,2,2)

Hommes

Pays	Espagne	Japon	Canada	Royaume-Uni	France
Baseline	ARIMA(0,1,1)	ARIMA(1,1,0)	MAD(-1,38)	ARIMA(1,1,0)	ARIMA(0,2,2)
Covid	ARIMA(1,1,0)	ARIMA(1,1,0)	ARIMA(0,2,2)	ARIMA(1,2,1)	ARIMA(0,2,2)

Nous pouvons remarquer que pour les hommes au Japon et en France, les modèles optimaux pour projeter la composante temporelle sont restés identiques entre les deux scénarios.

Pour les autres cas, nous observons soit un changement de méthode, passant d'une marche aléatoire avec drift à un modèle ARIMA plus complexe, soit une modification des paramètres des modèles entre les deux scénarios.

Par exemple, pour les femmes en Espagne, dans le scénario Baseline, le modèle optimal retenu est ARIMA(0,1,1). Ce dernier inclut donc un terme de différenciation ($d=1$) pour capturer la non-stationnarité, un terme de moyenne mobile MA ($q=1$) pour prendre en compte le terme d'erreur de la période précédente mais il ne contient pas de terme autorégressif AR ($p=0$). En revanche, pour cette même population, sous le scénario Covid, le modèle optimal devient plus complexe : ARIMA(2,2,1), ajoutant deux termes autorégressifs ainsi qu'une différenciation supplémentaire.

Pour les hommes au Canada, la composante temporelle passe d'être modélisée par une simple marche aléatoire avec drift à un modèle ARIMA(0,2,2), permettant de mieux capturer les dynamiques spécifiques de la nouvelle série temporelle.

Comme nous l'avons vu dans la figure 16, ces changements de méthodes et de paramètres pourraient impacter les trajectoires des projections de nos séries k_t . Cela impacterait, par conséquent, nos taux de mortalité projetés, entraînant ainsi des différences dans les estimations de vie résiduelles.

3.2.2) Résultats de calibration des modèles choisis par pays

Comme ce que nous avons fait pour les femmes en France, nous avons tracé les paramètres du modèle retenu par pays et par sexe sous les deux scénarios. L'idée est d'abord de visualiser l'allure des paramètres reflétant les caractéristiques spécifiques de chaque population étudiée. Ensuite, l'objectif serait d'analyser, pour chaque sexe et chaque pays, comment les paramètres de leur modèle sous le scénario Baseline ont été influencés par le scénario Covid. Nous cherchons alors à analyser l'effet de la pandémie en année *jump-off* sur la calibration des modèles.

1) L'Espagne

En observant les paramètres α_x , k_t et γ_{t-x} du modèle APC sous les deux scénarios, nous constatons qu'ils n'ont pas été significativement affectés par le scénario Covid pour la population espagnole, aussi bien pour les femmes (figure 17) que pour les hommes (figure 18).

Femmes

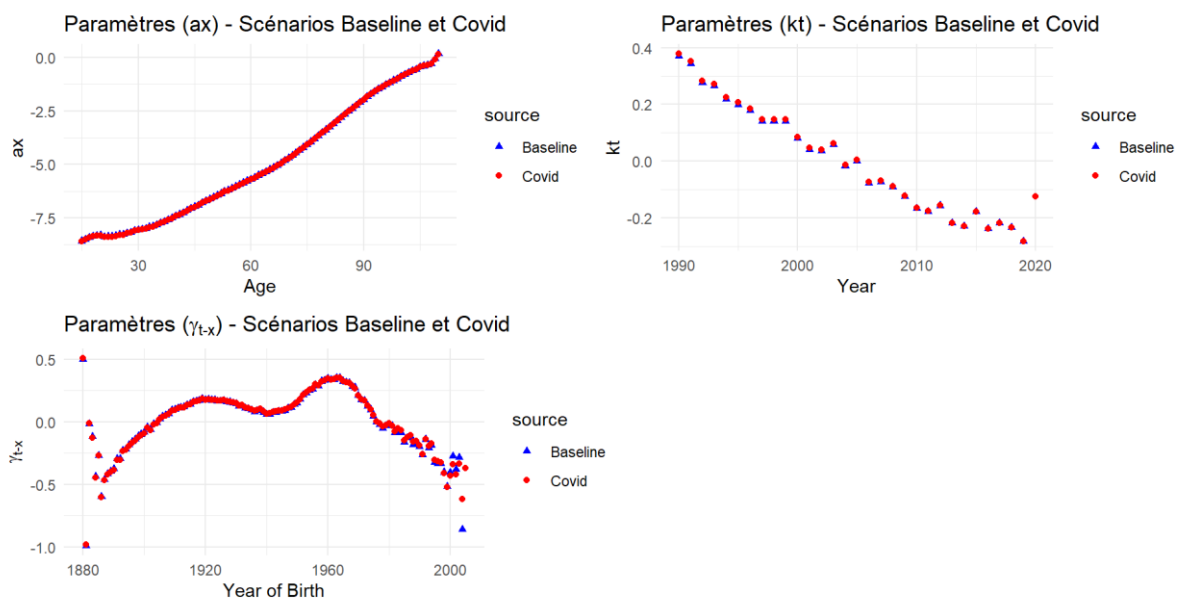


Figure 17 : Effet du scénario covid sur les paramètres du modèle APC pour les femmes en Espagne

Cependant, nous pouvons voir un saut de la composante temporelle k_t pour l'année 2020 présent chez les deux sexes. En effet, toutes choses étant égales par ailleurs, ce pic de k_t entraîne une augmentation moyenne de 17% de la mortalité entre 2019 et 2020 chez les femmes espagnoles, tous âges confondus, et de 15,7% chez les hommes.

Hommes

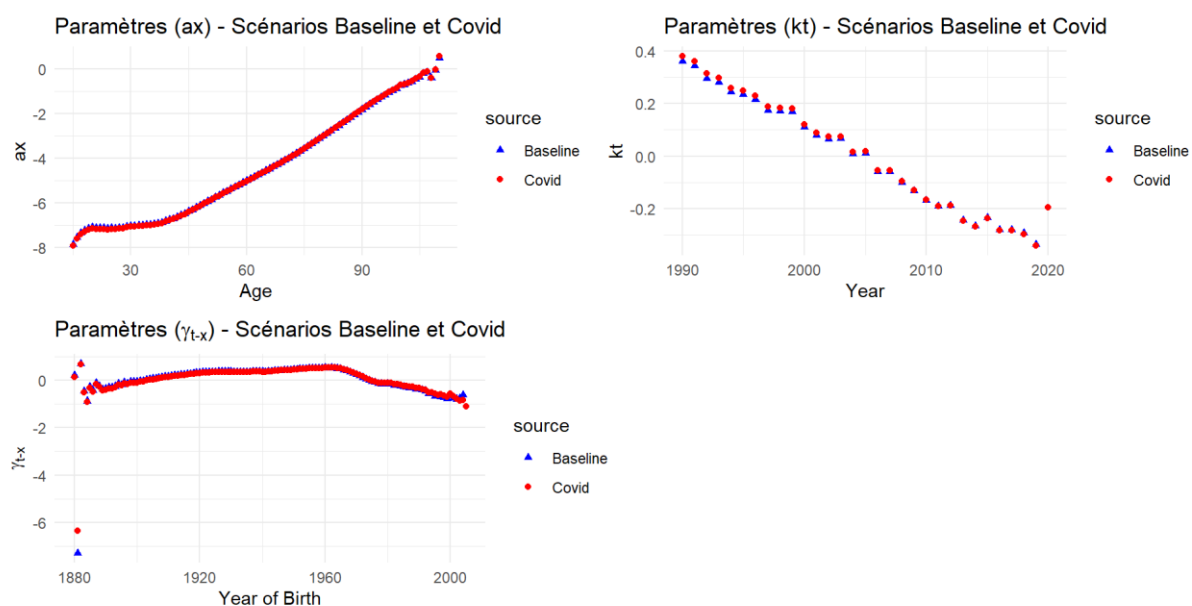


Figure 18 : Effet du scénario covid sur les paramètres du modèle APC pour les hommes en Espagne

2) Le Japon

A contrario, nous observons plus de changements pour le modèle RH chez les femmes au Japon (figure 19). Par exemple, en regardant la composante β_x , nous déduisons que la sensibilité de la mortalité à l'âge x par rapport à l'évolution générale k_t a été impactée pour les âges élevés. En effet, nous observons un décalage vers le haut, ce qui veut dire que les personnes âgées sont devenues plus sensibles à l'évolution générale dans le scénario Covid.

Concernant cette évolution générale de mortalité, nous remarquons que la courbe de la composante temporelle k_t n'a pas présenté de saut, comme cela a été observé chez les femmes en France et en Espagne, mais a plutôt vu sa pente diminuer. Cela suggère que sous le scénario Covid, l'amélioration de la mortalité au cours des années est plus « lente » que celle observée sous le scénario Baseline.

Femmes

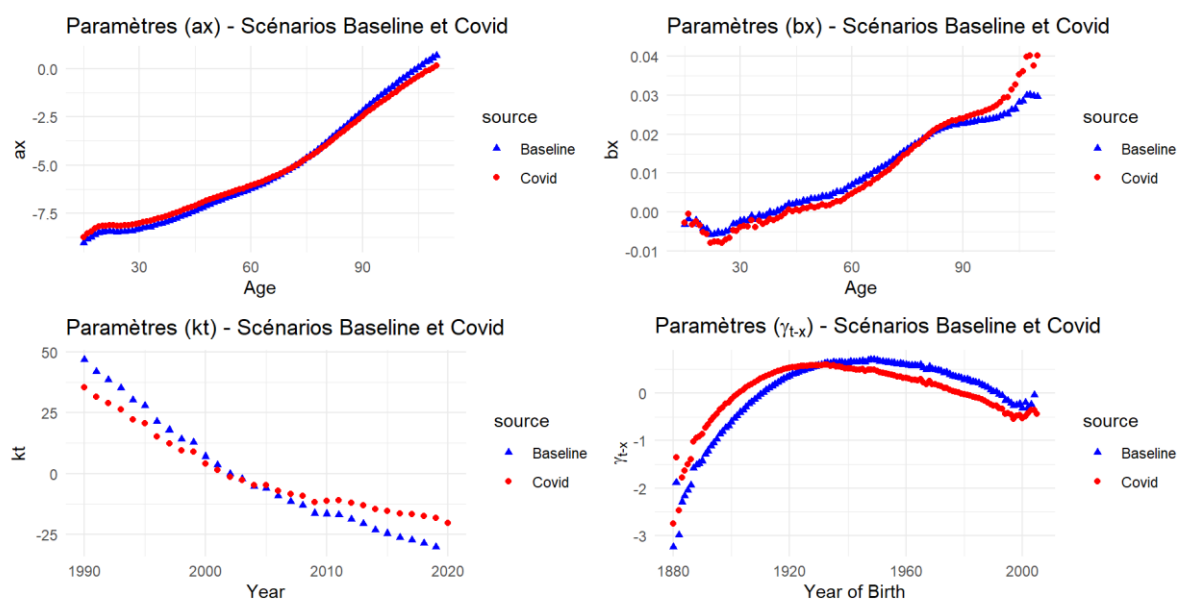


Figure 19 : Effet du scénario covid sur les paramètres du modèle RH pour les femmes au Japon

Pour les hommes au Japon, nous pouvons observer, dans la figure 20, que les paramètres du modèle LC n'ont pas été significativement affectés.

De plus, il est particulièrement intéressant de noter que, contrairement aux autres pays, le paramètre k_t n'a pas été impacté en 2020 sous le scénario Covid et a continué de suivre une tendance décroissante.

Hommes

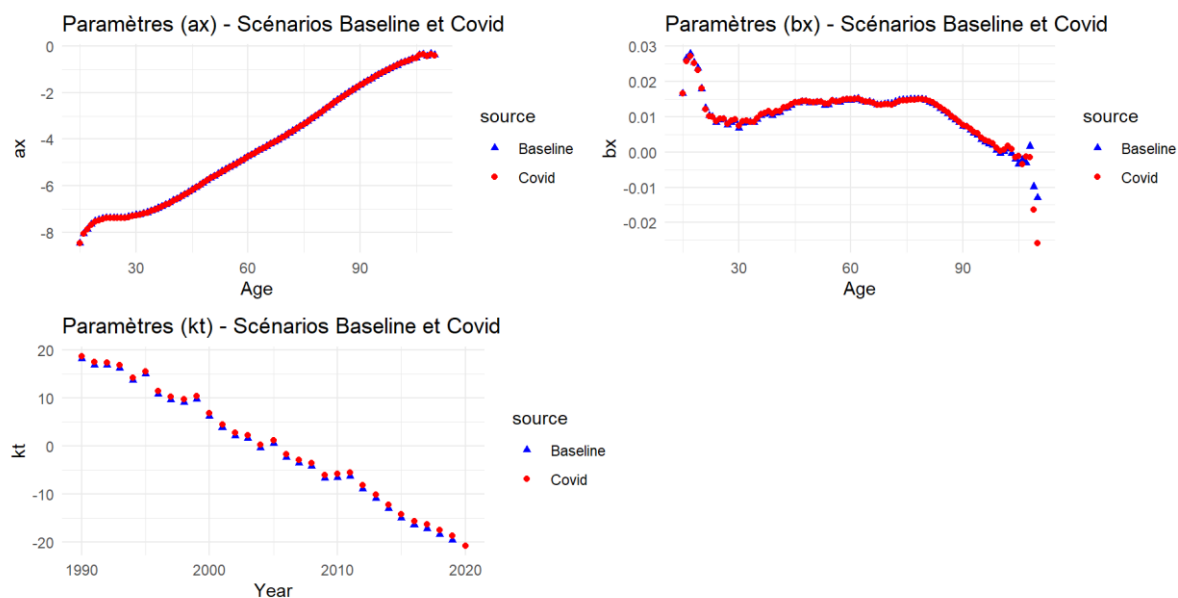


Figure 20 : Effet du scénario covid sur les paramètres du modèle LC pour les hommes au Japon

3) Canada

Concernant le Canada, nous pouvons observer dans la figure 21 que les paramètres du modèle APC chez les femmes n'ont pas été beaucoup affectés, à l'exception du saut observé pour la composante temporelle k_t qui, toutes choses étant égales par ailleurs, conduit à une augmentation moyenne de 5% de la mortalité, tous âges confondus, entre l'année 2019 et 2020.

Femmes

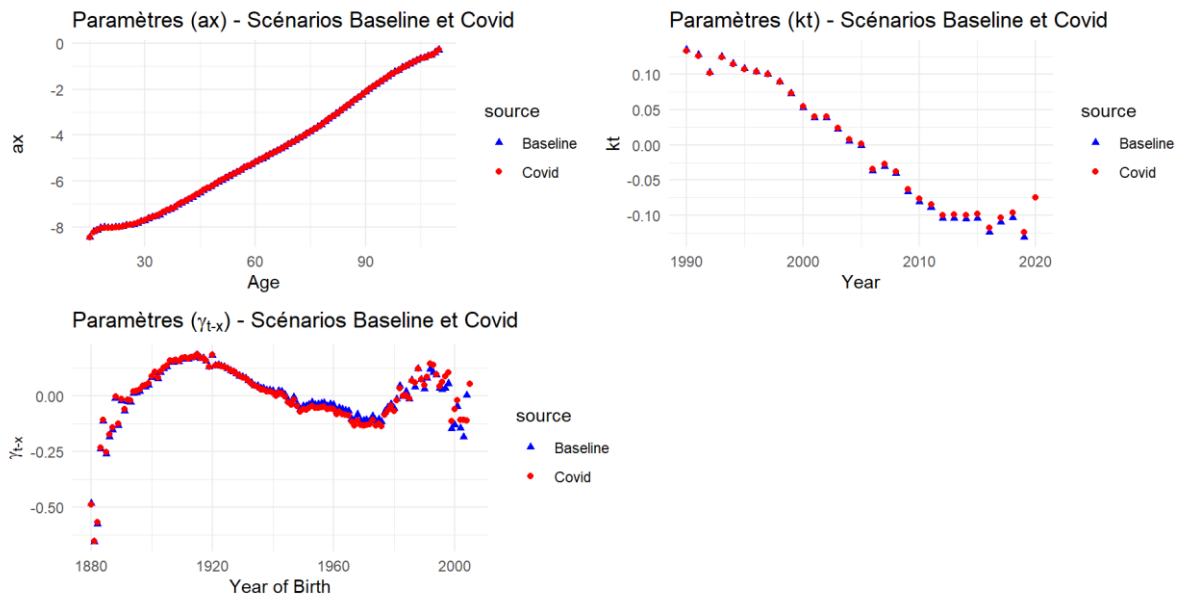


Figure 21 : Effet du scénario covid sur les paramètres du modèle APC pour les femmes au Canada

Pour les hommes, nous pouvons observer plusieurs changements dans les paramètres du modèles RH sur la figure 22. En effet, comme pour le modèle RH chez les femmes aux Japon, nous constatons un changement de pente de la courbe de la composante k_t . Cependant, cette fois-ci, la pente ne diminue pas mais augmente par rapport au scénario Baseline. De plus, non seulement la pente a été modifiée, mais un saut en 2020 est également présent sur la courbe Covid.

Hommes

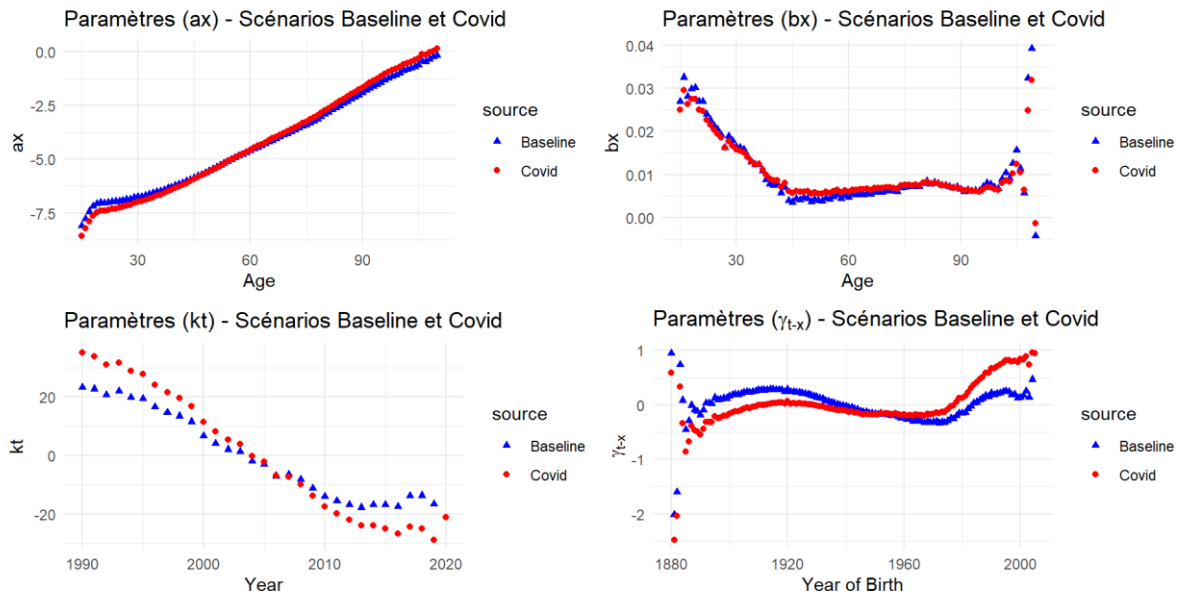


Figure 22 : Effet du scénario covid sur les paramètres du modèle RH pour les hommes au Canada

4) Le Royaume-Uni

Au Royaume-Uni, chez les femmes, nous remarquons que les paramètres α_x et γ_{t-x} n'ont pas été modifiés et ne sont donc pas sensibles au choc de la pandémie en année *jump-off* (figure 23). Concernant le paramètre k_t , nous pouvons premièrement constater que la courbe n'a pas une forme décroissante comme pour les autres pays et que ses valeurs ne s'éloignent pas fortement de 0. Cela pourrait suggérer que, dans ce cas, l'évolution de la mortalité au cours du temps est considérée comme relativement stable pour les femmes au Royaume-Uni. Cependant, nous observons un saut d'environ 10 unités en 2020 pour le scénario Covid.

Femmes

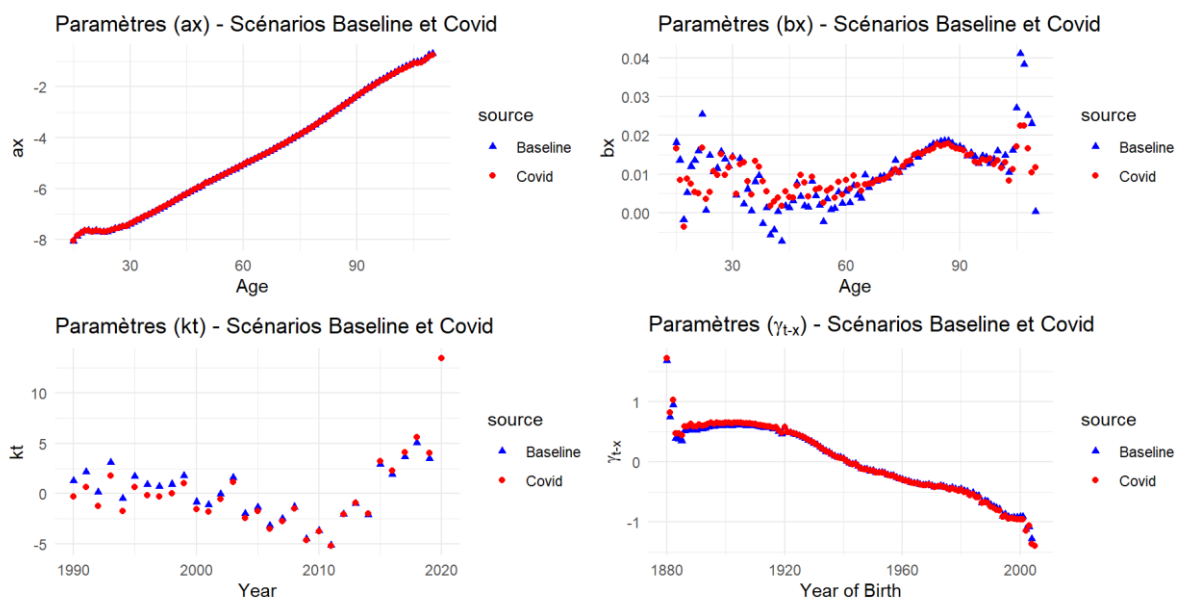


Figure 23 : Effet du scénario covid sur les paramètres du modèle RH pour les femmes au Royaume-Uni

Hommes

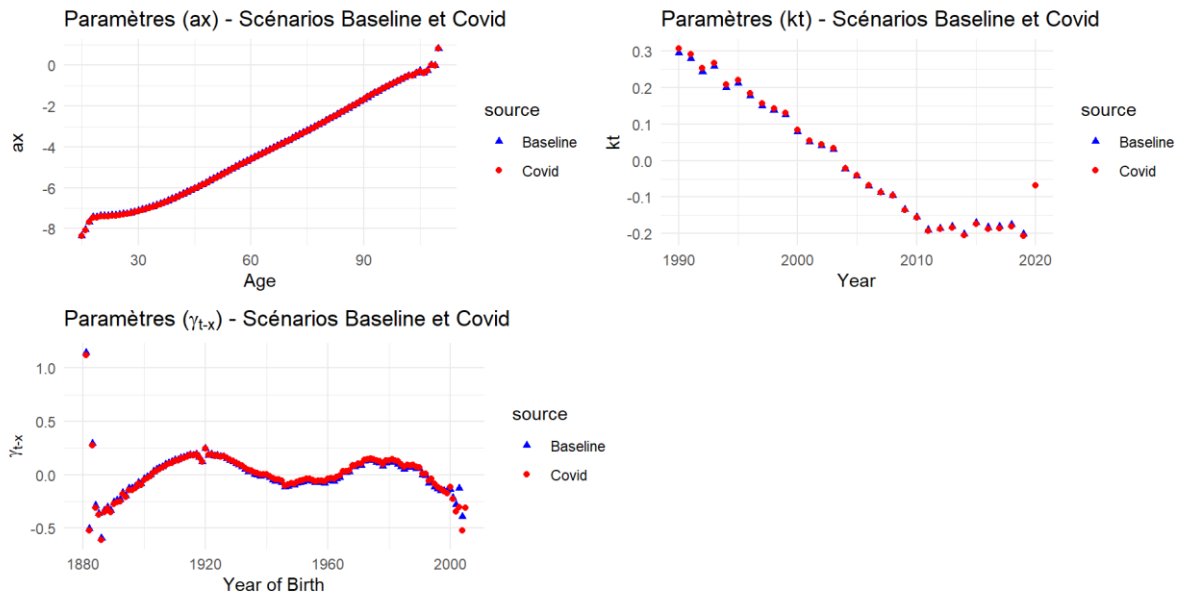


Figure 24 : Effet du scénario covid sur les paramètres du modèle APC pour les hommes au Royaume-Uni

Chez les hommes, les paramètres α_x et β_x du modèle APC n'ont pas été sensibles au choc de la pandémie (figure 24). En revanche, nous retrouvons le choc de la composante temporelle k_t en 2020. Ce choc a comme conséquence, toutes choses étant égales par ailleurs, une augmentation de 15% en moyenne de la mortalité entre 2019 et 2020, tous âges confondus.

5) France – Hommes

Enfin, pour les hommes en France (figure 25), les paramètres du modèle Renshaw-Haberman ont été plutôt sensibles au choc de la pandémie. En effet, nous remarquons que la courbe des paramètres k_t a vu sa pente diminuer, comme c'était le cas pour les femmes au Japon. Nous constatons également un saut en 2020 sous le scénario Covid, reflétant ainsi le choc de la pandémie.

En ce qui concerne les paramètres β_x , nous pouvons remarquer une faible augmentation pour les jeunes âges (jusqu'à 40 ans). Cela suggère que ces derniers sont plus sensibles aux évolutions de la mortalité sous le scénario Covid par rapport au scénario Baseline. En d'autres termes, un progrès technologique améliorant la mortalité générale aurait un impact plus marqué sur les jeunes âges dans le scénario Covid que dans le scénario Baseline.

Pour le paramètre de cohorte γ_{t-x} , nous observons que l'effet de cohorte est plus prononcé dans le scénario Baseline pour les générations les plus âgées que sous le scénario Covid. En revanche, pour les générations plus récentes, l'effet de cohorte est plus marqué sous le scénario Covid.

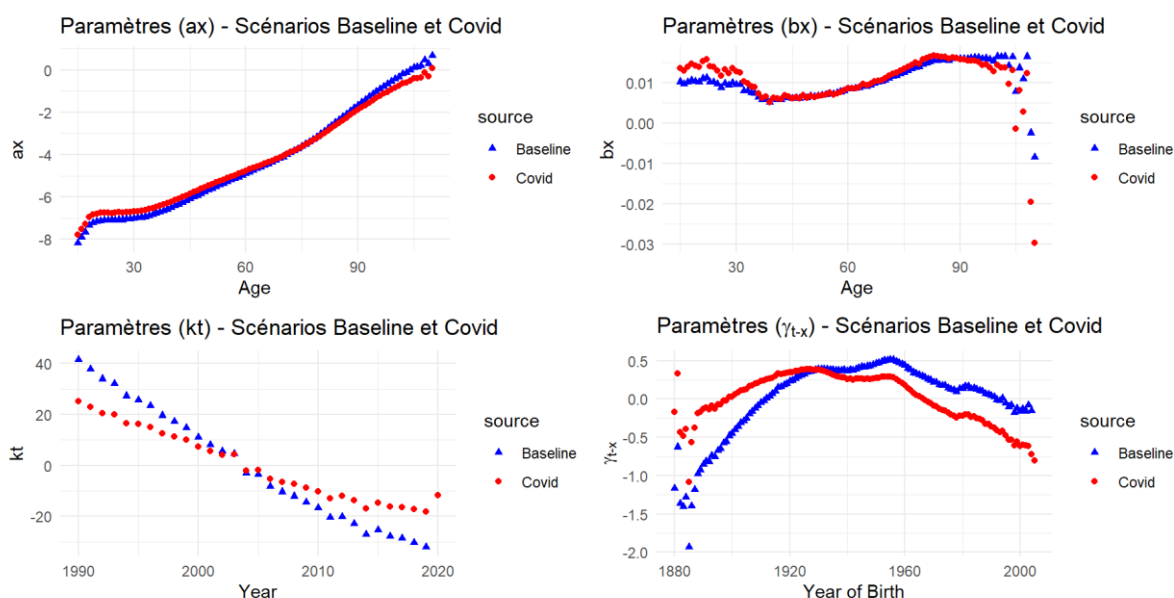


Figure 25 : Effet du scénario covid sur les paramètres du modèle RH pour les hommes en France

Nous pouvons remarquer que, en général, les paramètres des modèles APC et LC n'ont pas changé de manière significative à la suite du choc de la Covid à la dernière année de calibration. En effet, à part le saut de la composante temporelle k_t en 2020, tous les paramètres sont restés quasiment identiques entre les deux scénarios. En revanche, les paramètres du modèle RH ont été en général plus sensibles à ce choc de la Covid.

Maintenant que nous avons sélectionné le modèle optimal pour chaque population ainsi que la méthode de projection de la composante temporelle k_t , il est désormais possible de réaliser les projections des taux de mortalité.

Partie III : Projections et applications actuarielles

Dans la section précédente nous avons vu l'effet du choc pandémique de l'année *jump-off* sur la calibration et les paramètres de nos modèles. Nous allons maintenant, dans cette partie, étudier cet effet sur les projections issues de ces derniers afin de pouvoir quantifier l'impact actuariel. Notre étude se déroulera en trois étapes :

- Dans un premier temps, nous allons comparer les taux de mortalité réels de l'année 2020 avec ceux projetés par les modèles Baseline. L'objectif est de voir si les taux réels se situent dans la marge d'erreur prévue par le modèle non choqué.
- Ensuite, nous allons calculer les espérances de vie résiduelles à partir des taux de mortalité projetés pour l'année 2021 sous les scénarios Baseline et Covid. Dans un premier temps, nous comparerons ces estimations entre elles afin d'évaluer l'impact qu'a le choix du scénario sur les estimations. Puis, nous allons les comparer aux données réelles observées afin de déterminer lequel se rapproche le plus de la réalité.
- Enfin, nous allons réaliser une application de cette étude sur un portefeuille de rentiers viagers afin de quantifier cet effet du choc en année *jump-off* sur les engagements de l'assureur.

1) Comparaison des taux de mortalité du modèle Baseline avec les taux observés en 2020

Nous allons dans cette section comparer les taux de mortalité observés à l'année 2020 avec les taux prédits par le modèle Baseline non choqué. Nous cherchons à savoir si les données réelles du choc auraient pu être contenues dans la marge d'erreur du modèle de base.

Afin d'obtenir des projections plus robustes, nous avons réalisé 10 000 simulations. Chaque trajectoire est générée à travers les termes d'erreur du modèle ARIMA ou MAD tirés aléatoirement selon une loi normale, que ce soit pour la projection de k_t ou aussi pour la composante cohorte y_{t-x} dans les modèles RH ou APC.

A partir de ces simulations, en calculant les différents quantiles, nous avons pu obtenir des intervalles de confiance à 95 % et 80 %. L'intervalle à 95% (respectivement 80%) contient 95 % (respectivement 80%) des taux estimés, ce qui signifie que, si le modèle est toujours adéquat, il existe 95 % (respectivement 80%) de chances que le taux de mortalité réel soit compris à l'intérieur de cet intervalle. Ces intervalles définissent ainsi nos marges d'erreur, permettant d'évaluer la précision du modèle.

De plus, nous avons décidé de réaliser un test statistique de Student pour des échantillons appariés afin de voir si la moyenne observée des taux de mortalité prédits par le modèle Baseline est significativement différente de celle des taux réels.

Pour rappel, le test de Student pour échantillons appariés est une méthode statistique utilisée pour comparer les moyennes de deux échantillons liés. Ce test permet de déterminer si la différence moyenne entre les deux échantillons est significativement différente de zéro, c'est-à-dire si les moyennes observées des deux échantillons sont statistiquement différentes :

- **Hypothèse Nulle (H_0)** : Il n'existe pas de différence significative entre les moyennes de nos deux échantillons
- **Hypothèse Alternative (H_1)** : Il existe une différence significative entre les moyennes de nos deux échantillons

La statistique **t** est calculée à partir de la moyenne des différences entre chaque paire (taux réel - taux prédit), leur écart type ainsi que le nombre de paires. Une fois la statistique **t** calculée, il est possible de déterminer sa p-value associée.

Cependant, plusieurs conditions de validité du test doivent être vérifiées :

- 1) Les données doivent être appariées : Cette condition est vérifiée puisque nous comparons les taux pour la même population ayant le même âge.
- 2) Les différences entre les paires doivent suivre une loi normale : Cette condition n'est pas nécessaire pour les échantillons suffisamment grands (>30) grâce au Théorème Central Limite. Ce qui est notre cas.
- 3) Indépendance des paires, c'est-à-dire que chaque paire d'observations est indépendante des autres paires. Nous partons du principe que cette condition est remplie, pour mener à bien le test même si l'hypothèse n'est pas entièrement vérifiée.

Nous avons obtenu les résultats suivants :

	France	Espagne	Japon	Canada	Royaume-Uni
Femmes	0,02	0,00	0,16	0,03	0,00
Hommes	0,07	0,24	0,26	0,60	0,73

Tableau 1 : P-values du test de Student réalisé sur les taux de mortalité prédits par le modèle Baseline et les taux observés pour l'année 2020.

Nous pouvons remarquer dans le tableau 1 que chez les femmes en France, Espagne, Canada et Royaume-Uni, les p-values sont inférieures à 0,05, ce qui nous permet de rejeter l'hypothèse nulle selon laquelle la moyenne des valeurs prédites pour tous les âges n'est pas significativement différente de celle des valeurs observées. En revanche, chez les femmes au Japon ainsi que chez les hommes dans les cinq pays étudiés, l'hypothèse nulle est conservée, indiquant que les deux moyennes ne présentent pas de différence significative.

Toutefois, en observant, dans le tableau 2, les proportions des taux observés se situant dans les intervalles de confiance estimés par le modèle, nous constatons que les résultats ne sont pas satisfaisants. En effet, à l'exception du Japon, une faible proportion des taux observés se trouve à l'intérieur de nos IC dans les autres pays chez les deux sexes.

	France	Espagne	Japon	Canada	Royaume-Uni
Femmes	32% (15%)	15% (9%)	47% (17%)	14% (7%)	10% (5%)
Hommes	16% (10%)	1% (1%)	55% (44%)	10% (7%)	4% (3%)

Tableau 2 : Proportion des taux de mortalité observés en année 2020 se situant dans l'IC 95% (et 80%) du modèle Baseline

Comparaison pour la population âgée de 65 ans

Nous avons ensuite tracé, pour chaque pays, les écarts entre les taux de mortalité observés et prédits pour les années 2020 et 2021 chez les femmes (figure 26) et les hommes (figure 27) âgés de 65 ans, correspondant à l'âge légal moyen de départ à la retraite dans la majorité des pays.

Nous pouvons voir que dans la majorité des cas, les taux de mortalité observés ont été supérieurs à ceux prédits. Cela était attendu puisque le modèle a été calibré sur la période avant le choc et a ainsi prédit les taux dans un scénario "normal" sans évènement exceptionnel. En revanche, au Japon, chez les deux sexes, nous n'observons presque pas d'écart entre les valeurs prédites et celles observées pour l'année 2020, mais un écart a été observé chez les hommes en 2021

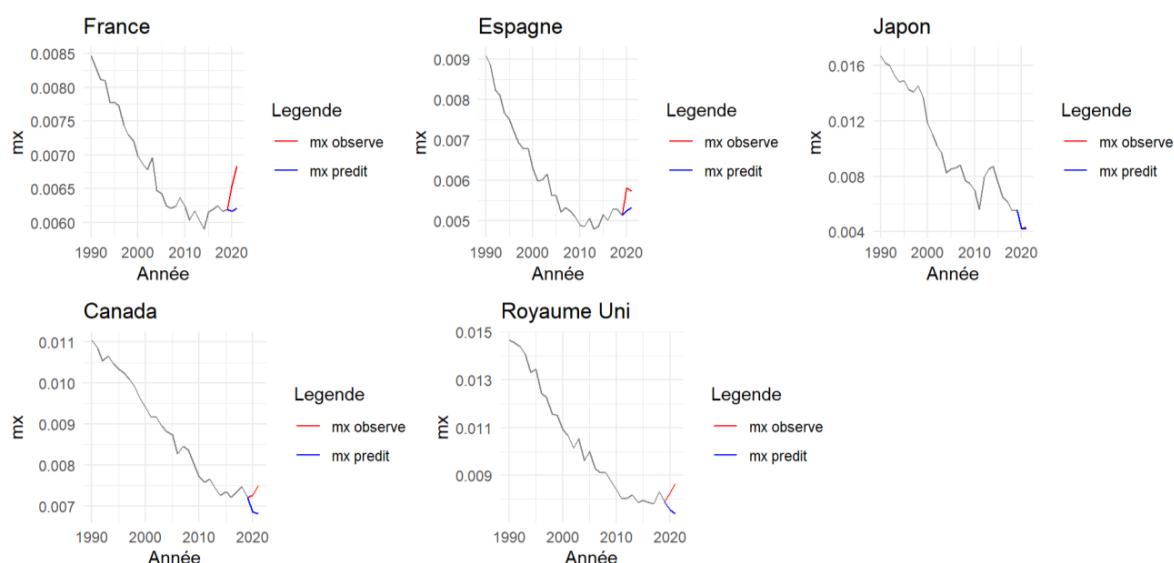


Figure 26 : Taux de mortalité observés VS prédits pour les femmes âgées de 65 ans

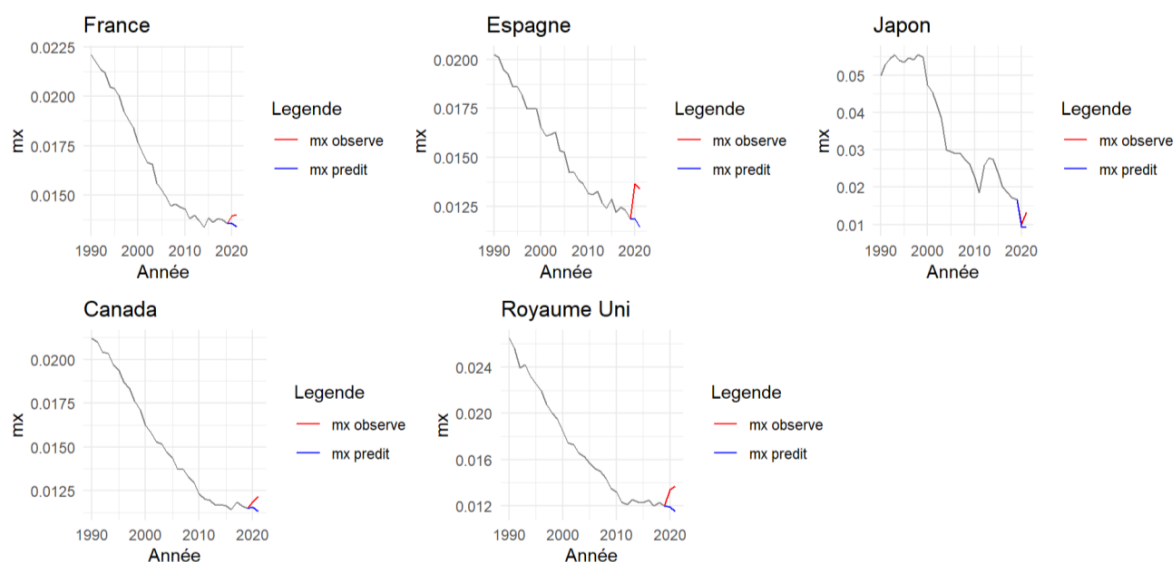


Figure 27 : Taux de mortalité observés VS prédits pour les hommes âgés de 65 ans

2) Comparaison des espérances de vie projetées en 2021

Dans cette partie, nous allons construire les tables de mortalité à partir des taux de mortalité prédits et observés en 2021.

Une fois les espérances de vie résiduelles calculées pour chaque âge, nous analyserons l'effet du choc survenu à l'année *jump-off* sur ces estimations en comparant celles issues du modèle Baseline avec celles issues du modèle Covid.

Ensuite, en comparant les écarts entre les espérances de vie estimées sous les deux scénarios et celles observées, nous allons chercher à déterminer quel modèle, avec ou sans prise en compte du choc pandémique en année *jump-off*, a prédit le mieux la perte d'espérance de vie observée en 2021.

Pour calculer les espérances de vie, nous avons procédé ainsi :

- 1) Nous avons d'abord calculé les probabilités de survie p_x à partir des taux de mortalité m_x
- 2) Ensuite, nous avons calculé les espérances de vie résiduelles E_x selon la formule suivante :

$$e_x = \frac{l_{x+1} + l_{x+2} + \dots + l_{110}}{l_x}$$

2.1) Comparaison entre les estimations des deux modèles

Dans cette partie, nous nous intéressons à la comparaison entre les estimations issues des deux modèles, celui du scénario Baseline et celui du scénario Covid. Nous cherchons à évaluer l'impact du choix du modèle sur nos estimations. Plus précisément, nous cherchons à mesurer combien d'années d'espérance de vie notre population gagne ou perd en moyenne en utilisant le modèle Covid plutôt que le modèle Baseline.

Après avoir estimé les espérances de vie résiduelles pour l'année 2021 à partir de nos deux modèles, nous avons calculé les écarts entre ces estimations.

Ecart Covid/Baseline	France	Espagne	Canada	Japon	Royaume-Uni
Femmes	-3,5 mois	-6,1 mois	-2,7 mois	1,2 mois	-7,9 mois
Hommes	-4,9 mois	-5,9 mois	-5,9 mois	0,1 mois	-8,4 mois

Tableau 3 : Ecart (en mois) entre les estimations d'espérance de vie résiduelles issues du modèle Covid et celles issues du modèle Baseline, tous âges confondus

Les résultats présentés dans le tableau 3 montrent qu'en France, au Canada et au Royaume-Uni, l'écart entre les estimations des deux modèles a été, en moyenne, plus important (en valeur absolue) chez les hommes que chez les femmes. Par exemple, au Canada, l'utilisation du modèle Covid a conduit à une

diminution moyenne de 5,9 mois de l'espérance de vie des hommes par rapport aux estimations du modèle Baseline.

De plus, nous remarquons que l'écart le plus élevé a été observé au Royaume-Uni chez les deux sexes, et le moins élevé au Japon. Nous pouvons alors a priori s'attendre que l'impact du choix du modèle sur les engagements serait en général plus important au Royaume-Uni et un effet moins important au Japon.

Enfin, au Japon, l'écart est positif pour les hommes et pour les femmes, indiquant que, contrairement aux autres pays, le modèle Covid a conduit à une légère augmentation de l'espérance de vie plutôt qu'à une diminution.

Comparaison pour la population âgée de 65 ans

Maintenant, nous allons regarder plus en détail les écarts entre les estimations des deux modèles en 2021, en nous concentrant uniquement sur la population âgée de 65 ans.

Sur les tableaux 4 et 5, nous remarquons que les observations ci-dessus sur les écarts tous âges confondus restent identiques pour les estimations de la population âgée de 65 ans.

En effet, en France, Canada et Royaume-Uni, les écarts entre les estimations des modèles restent plus importants chez les hommes que chez les femmes. Par exemple, en France, les hommes âgés de 65 ans ont perdu 0,51 an d'espérance de vie, en utilisant le modèle Covid par rapport au modèle Baseline. Cet écart est de 0,37 an chez les femmes.

De plus, les écarts entre les deux modèles au Royaume-Uni restent supérieurs chez les deux sexes que les autres pays étudiés, avec une perte de 0,84 an chez les hommes et 0,78 an chez les femmes.

Enfin, nous remarquons que le Japon présente toujours les écarts les plus faibles parmi les pays mais aussi les seuls écarts positifs.

	E_{65,B} (95% IC)	E_{65,C} (95% IC)	E_{65,C} - E_{65,B}
France	23,05 (22,52 ; 23,61)	22,68 (22,22 ; 23,14)	-0,37
Espagne	23,11 (22,67 ; 23,55)	22,52 (21,99 ; 23,04)	-0,59
Canada	22,14 (21,87 ; 22,41)	21,95 (21,72 ; 22,17)	-0,19
Japon	24,41 (23,94 ; 24,88)	24,52 (24,17 ; 24,87)	0,11
Royaume Uni	21,24 (20,76 ; 21,73)	20,46 (20,03 ; 20,89)	-0,78

Tableau 4 : Comparaison des espérances de vie résiduelles à 65 ans chez les femmes en 2021 : Espérances de vie estimées par le modèle Baseline (E_{65,B}) et Covid (E_{65,C}), et écarts associés

	E_{65,B} (95% IC)	E_{65,C} (95% IC)	E_{65,C} - E_{65,B}
France	19,19 (18,85 ; 19,53)	18,68 (18,32 ; 19,05)	-0,51
Espagne	19,21 (18,86 ; 19,58)	18,66 (18,12 ; 19,21)	-0,55
Canada	19,51 (19,25 ; 19,76)	19,06 (18,84 ; 19,28)	-0,45
Japon	19,62 (19,36 ; 19,89)	19,63 (19,39 ; 19,87)	0,01
Royaume Uni	18,79 (18,43 ; 19,15)	17,95 (17,50 ; 18,39)	-0,84

Tableau 5 : Comparaison des espérances de vie résiduelles à 65 ans chez les hommes en 2021 : Espérances de vie estimées par le modèle Baseline (E_{65,B}) et Covid (E_{65,C}), et écarts associés

2.2) Comparaison entre les estimations et les valeurs observées

Dans cette partie, nous allons procéder cette fois-ci à la comparaison des estimations issues de nos deux scénarios aux données réelles en 2021. Pour cela, nous analysons les écarts entre les valeurs observées et celles prédites par chaque modèle, afin d'évaluer leur précision et déterminer lequel des deux scénarios se rapproche le plus de la réalité.

Nous observons, sur les graphiques 28 et 29, que dans la majorité des cas, l'écart entre les valeurs observées et celles estimées a été négatif puisque les courbes se situent sous l'axe des abscisses. Cela signifie que les espérances de vie observées en 2021 ont été inférieures à celles prédites par nos modèles, quel que soit le scénario choisi.

Toutefois, le modèle Covid semble mieux refléter la réalité que le modèle Baseline. En effet, dans la majorité des cas, les écarts entre les valeurs prédites par le modèle Baseline et les valeurs réelles sont généralement plus importants que ceux obtenus avec le modèle Covid.

Cependant, ces observations ne sont pas présentes au Japon. En effet, nous remarquons que chez les hommes les deux modèles présentent des écarts similaires vis-à-vis de la réalité, tandis que chez les femmes, le modèle Baseline semble même offrir des estimations plus proches de la réalité que le modèle Covid.

Enfin, nous remarquons que les écarts sont plus marqués aux âges jeunes et diminuent progressivement avec l'âge.

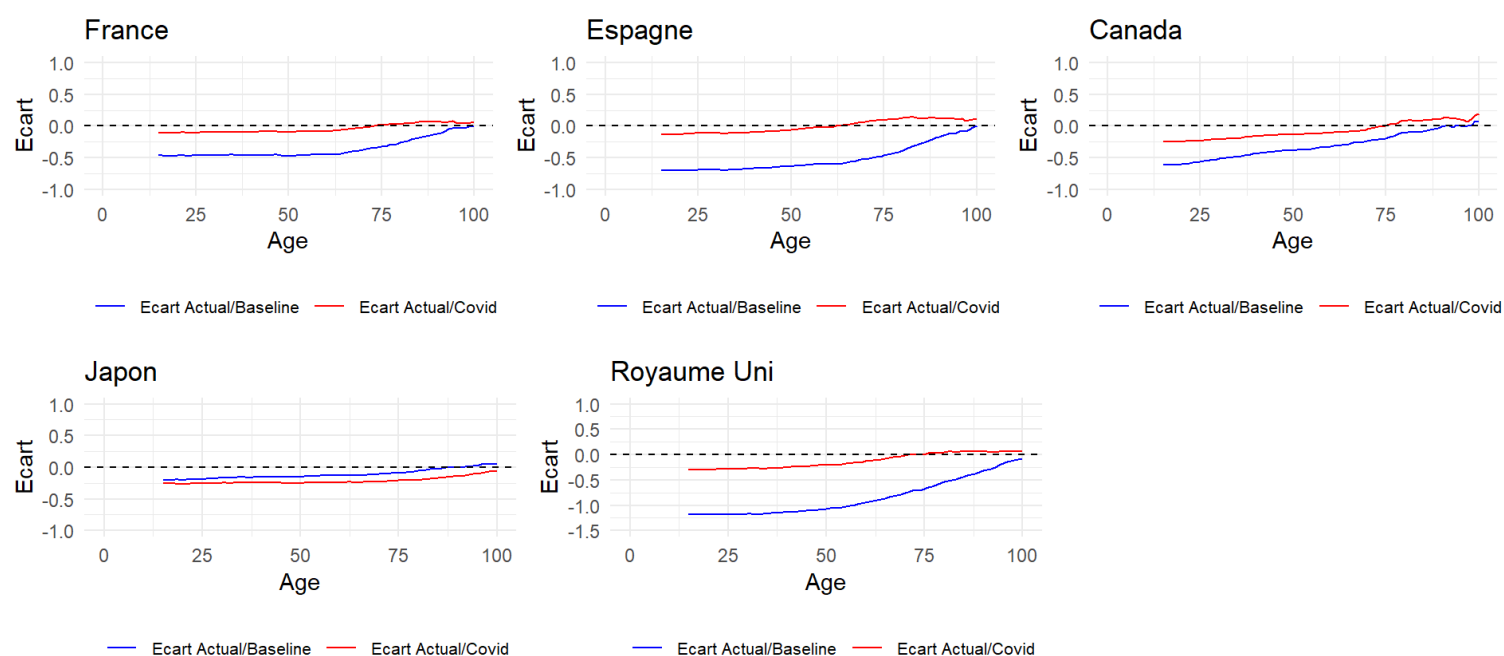


Figure 28 : Ecart par âge entre les espérances de vie estimées et celles observées en 2021 chez les femmes

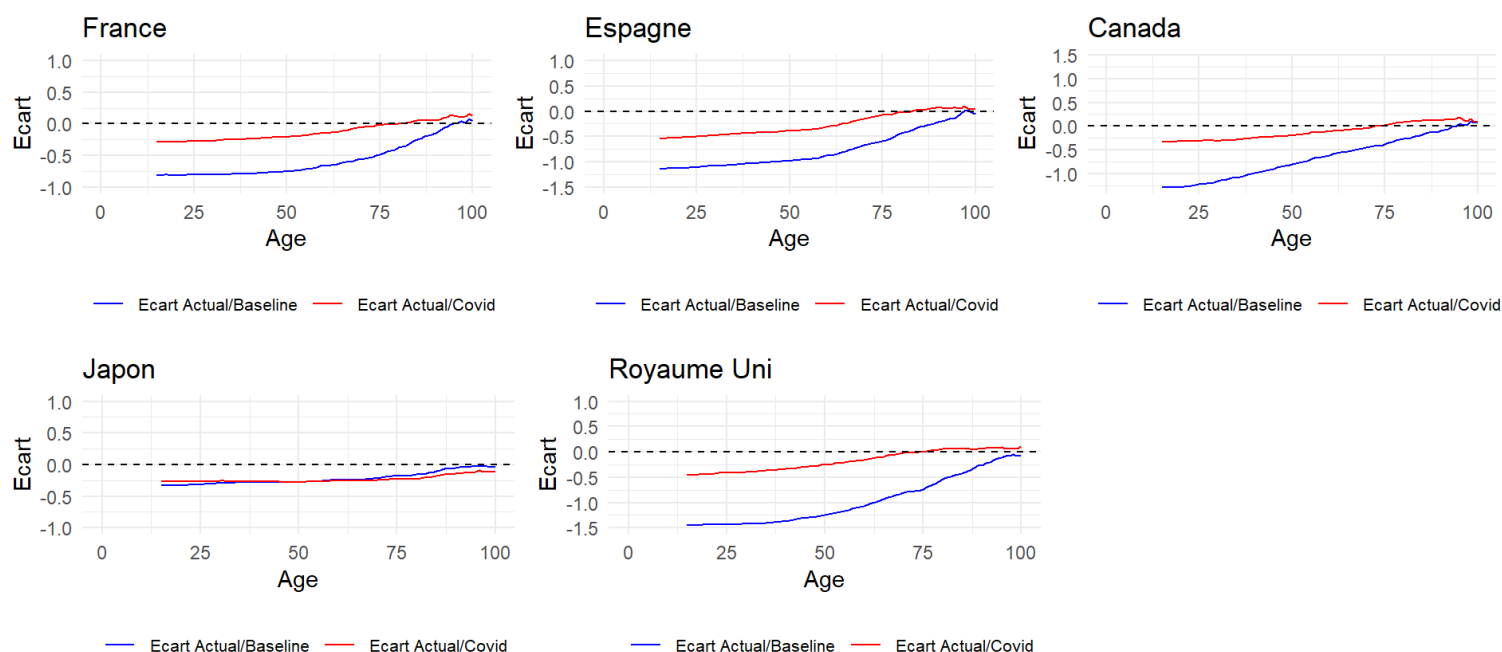


Figure 29 : Ecart par âge entre les espérances de vie estimées et celles observées en 2021 chez les hommes

Afin d'obtenir une vision plus globale de la performance de nos modèles, nous avons calculé l'écart moyen en espérance de vie entre les valeurs prédites et les valeurs observées, tous âges confondus.

- Comme observé sur les tableaux 6 et 7, le modèle Covid s'est révélé, en moyenne, plus proche de la réalité chez les deux sexes. Par exemple, au Royaume-Uni, l'écart moyen entre les estimations du modèle Baseline et les valeurs effectivement observées s'élève à 9,4 mois chez les femmes et 10,7 mois chez les hommes. Grâce au modèle Covid, cet écart a été réduit à 1,8 mois chez les femmes et 2.4 chez les hommes.
- De plus, nous remarquons que les écarts sont plus élevés chez les hommes dans tous les pays.
- Nous pouvons aussi noter que le modèle Covid chez les femmes en Espagne a montré une très bonne précision, avec un écart moyen de seulement 0,05 mois.
- Enfin, au Japon, chez les deux sexes, le modèle Baseline présente un écart moyen inférieur à celui du modèle Covid. Cela a été attendu au vu des graphiques 28 et 29.

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	4 mois	5,6 mois	3,6 mois	1,4 mois	9,4 mois
Covid	0,9 mois	0,05 mois	1,6 mois	2,5 mois	1,8 mois

Tableau 6 : Ecart moyen entre les espérances de vie prédites et celles observées chez les femmes pour l'année 2021 (en mois)

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	6,4 mois	8,1 mois	7,5 mois	2,5 mois	10,7 mois
Covid	2 mois	2,7 mois	2,2 mois	2,7 mois	2,4 mois

Tableau 7 : Ecart moyen entre les espérances de vie prédites et celles observées chez les hommes pour l'année 2021 (en mois)

De plus, nous avons calculé les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance estimés sous les deux scénarios. Nous remarquons dans les tableaux 8 et 9 que, à l'exception du Japon, chez les deux sexes, les modèles Covid ont été significativement plus précis, contenant une plus grande proportion des valeurs réelles dans leurs marges d'erreur.

Par exemple, en Espagne, chez les femmes, 97% des valeurs réelles se situent à l'intérieur de l'IC₉₅ du modèle Covid contre 24% pour le modèle Baseline. Chez les hommes, 93% des valeurs réelles sont contenues dans l'IC₉₅ contre seulement 5% pour le modèle Baseline.

A contrario, au Japon, les modèles Baseline se sont révélés plus précis, avec 98% des données réelles dans l'IC₉₅ chez les femmes et 94% chez les hommes.

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	13% (8%)	5% (3%)	16% (9%)	94% (17%)	7% (2%)
Covid	82% (72%)	93% (57%)	76% (32%)	53% (0%)	94% (76%)

Tableau 8 : Les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance 95% (et 80%) chez les femmes pour l'année 2021

	France	Espagne	Canada	Japon	Royaume Uni
Baseline	99% (33%)	24% (12%)	38% (23%)	98% (97%)	14% (8%)
Covid	96% (95%)	97% (94%)	78% (51%)	96% (39%)	99% (98%)

Tableau 9 : Les proportions des valeurs réelles qui se situent à l'intérieur des intervalles de confiance 95% (et 80%) chez les hommes pour l'année 2021

Comparaison pour la population âgée de 65 ans

Nous nous intéressons maintenant à la population âgée de 65 ans.

Nous remarquons sur les tableaux 10 et 11, qu'à l'exception du Japon, le modèle Covid s'est révélé plus précis que le modèle Baseline. En effet, les écarts entre les estimations issues du scénario Covid et les données observées sont, en valeur absolue, inférieurs à ceux du modèle Baseline. De plus, ces résultats semblent satisfaisants, avec des écarts ne dépassant pas 0,1 an chez les femmes et 0,25 an chez les hommes sous le scénario Covid.

En revanche, au Japon, le modèle Baseline a mieux reflété la réalité que le modèle Covid.

En examinant les intervalles de confiance, cette tendance se confirme. En effet, chez les femmes, seuls la France et le Japon présentent des données réelles incluses dans l'intervalle de confiance à 95 % du modèle Baseline, alors que toutes les données observées se situent dans celui du modèle Covid. Chez les hommes, à l'exception du Japon, aucune donnée réelle ne se trouve à l'intérieur de l'intervalle de confiance du modèle Baseline, alors qu'elles y sont toutes sous le scénario Covid.

Enfin, nous pouvons remarquer que, dans la grande majorité des cas, les modèles sur les populations féminines ont été plus performants que les modèles sur les populations des hommes, quel que soit le scénario choisi.

	E_{65,O}	E_{65,B} (95% IC)	E_{65,O} - E_{65,B}	E_{65,C} (95% IC)	E_{65,O} - E_{65,C}
France	22,63	23,05 (22,52 ; 23,61)	-0,42	22,68 (22,22 ; 23,14)	-0,05
Espagne	22,54	23,11 (22,67 ; 23,55)	-0,57	22,52 (21,99 ; 23,04)	0,02
Canada	21,86	22,14 (21,87 ; 22,41)	-0,28	21,95 (21,72 ; 22,17)	-0,09
Japon	24,28	24,41 (23,94 ; 24,88)	-0,13	24,52 (24,17 ; 24,87)	-0,24
Royaume Uni	20,38	21,24 (20,76 ; 21,73)	-0,86	20,46 (20,03 ; 20,89)	-0,08

Tableau 10 : Comparaison des espérances de vie résiduelles à 65 ans chez les femmes en 2021 : Espérances de vie observées (E_{65,O}), Espérances de vie estimées par le modèle Baseline (E_{65,B}) et Covid (E_{65,C}), et écarts associés

	E_{65,O}	E_{65,B} (95% IC)	E_{65,O} - E_{65,B}	E_{65,C} (95% IC)	E_{65,O} - E_{65,C}
France	18,57	19,19 (18,85 ; 19,53)	-0,62	18,68 (18,32 ; 19,05)	-0,11
Espagne	18,42	19,21 (18,86 ; 19,58)	-0,79	18,66 (18,12 ; 19,21)	-0,24
Canada	18,98	19,51 (19,25 ; 19,76)	-0,53	19,06 (18,84 ; 19,28)	-0,08
Japon	19,37	19,62 (19,36 ; 19,89)	-0,25	19,63 (19,39 ; 19,87)	-0,26
Royaume Uni	17,86	18,79 (18,43 ; 19,15)	-0,93	17,95 (17,50 ; 18,39)	-0,09

Tableau 11 : Comparaison des espérances de vie résiduelles à 65 ans chez les hommes en 2021 : Espérances de vie observées (E_{65,O}), Espérances de vie estimées par le modèle Baseline (E_{65,B}) et Covid (E_{65,C}), et écarts associés

3) Applications sur un portefeuille de rentiers

Nous allons dans cette partie quantifier l'impact du choc pandémique à l'année *jump-off* sur un portefeuille de rentiers viagers. Nous allons d'abord présenter notre portefeuille, ensuite nous allons calculer les engagements de l'assureur à l'année 2021 sous les deux scénarios dans chaque pays.

3.1) Présentation de notre portefeuille

Notre portefeuille est composé de 76 103 rentiers et comprend les variables suivantes :

- **Sexe** : Variable qualitative indiquant le sexe de chaque rentier (M ou F)
- **DDN** : Variable présentant la date de naissance de chaque rentier. A partir de la date de naissance, nous avons calculé l'âge de chaque rentier à l'année 2021.
- **Rentes** : Variable quantitative représentant le montant contractuel des rentes viagères que le rentier percevra à partir de son départ à la retraite.

Comme l'illustre la figure 30, notre portefeuille est composé de 62,6 % d'hommes et 37,4 % de femmes. Nous observons une répartition similaire pour les montants de rentes contractuels, avec 64,2 % détenus par les hommes et le reste par les femmes.

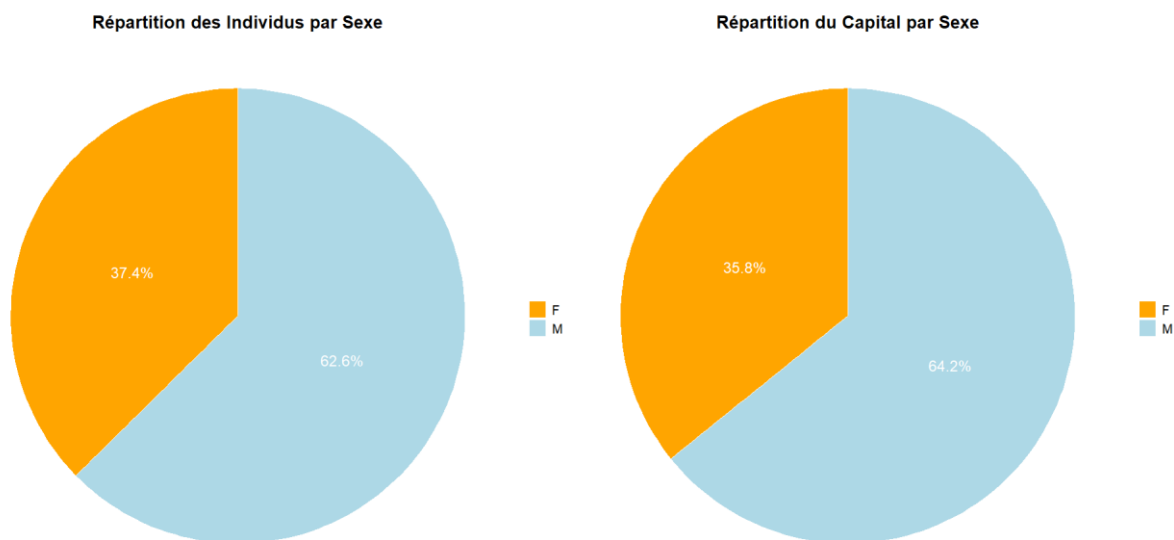


Figure 30 : Répartition des individus et du montant de rentes contractuel par sexe

En analysant plus en détail les montants contractuels des rentes dans la figure 31, nous constatons que la grande majorité des individus, hommes ou femmes, perçoivent ou percevront une rente inférieure à 20 000.

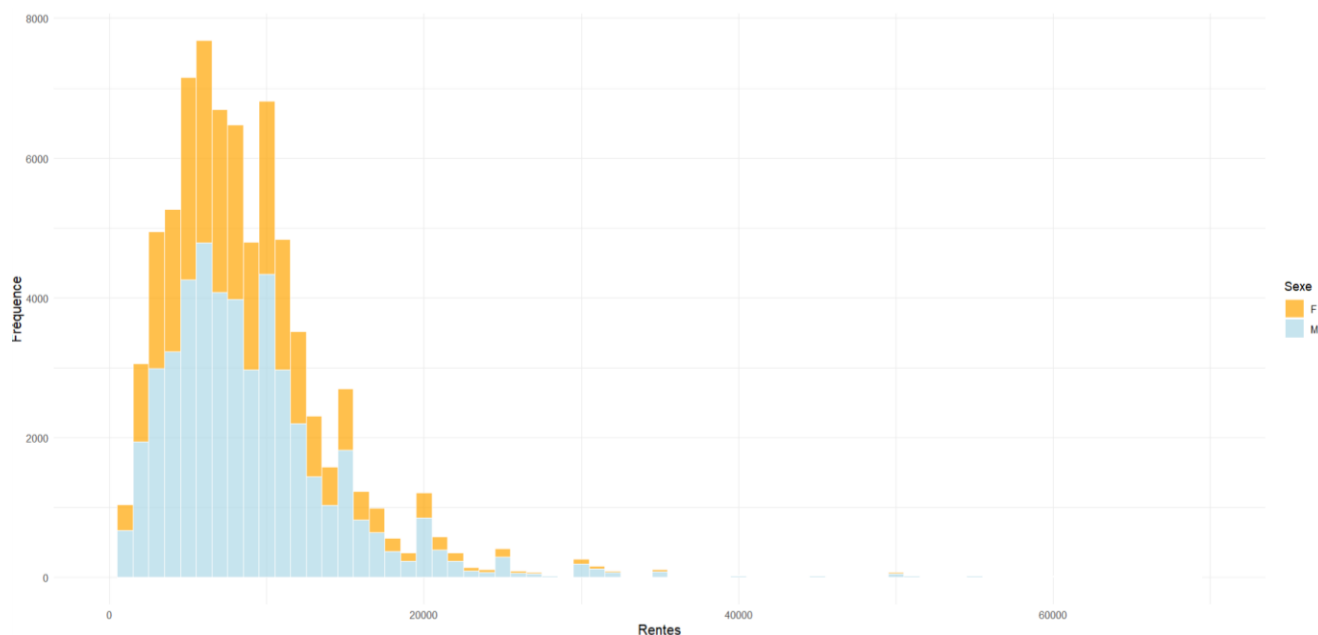


Figure 31 : Distribution des montants des rentes contractuels, par sexe

Enfin, dans la figure 32, nous pouvons voir que la tranche d'âge la plus représentée se situe entre 50 et 70 ans. De plus, l'âge minimal observé dans notre portefeuille est de 26 ans, tandis que l'âge maximal atteint 87 ans.

Il est également intéressant de noter que notre portefeuille compte 9 936 retraités, âgés de 65 ans ou plus, qui perçoivent déjà leurs rentes, contre 66 167 non-retraités.

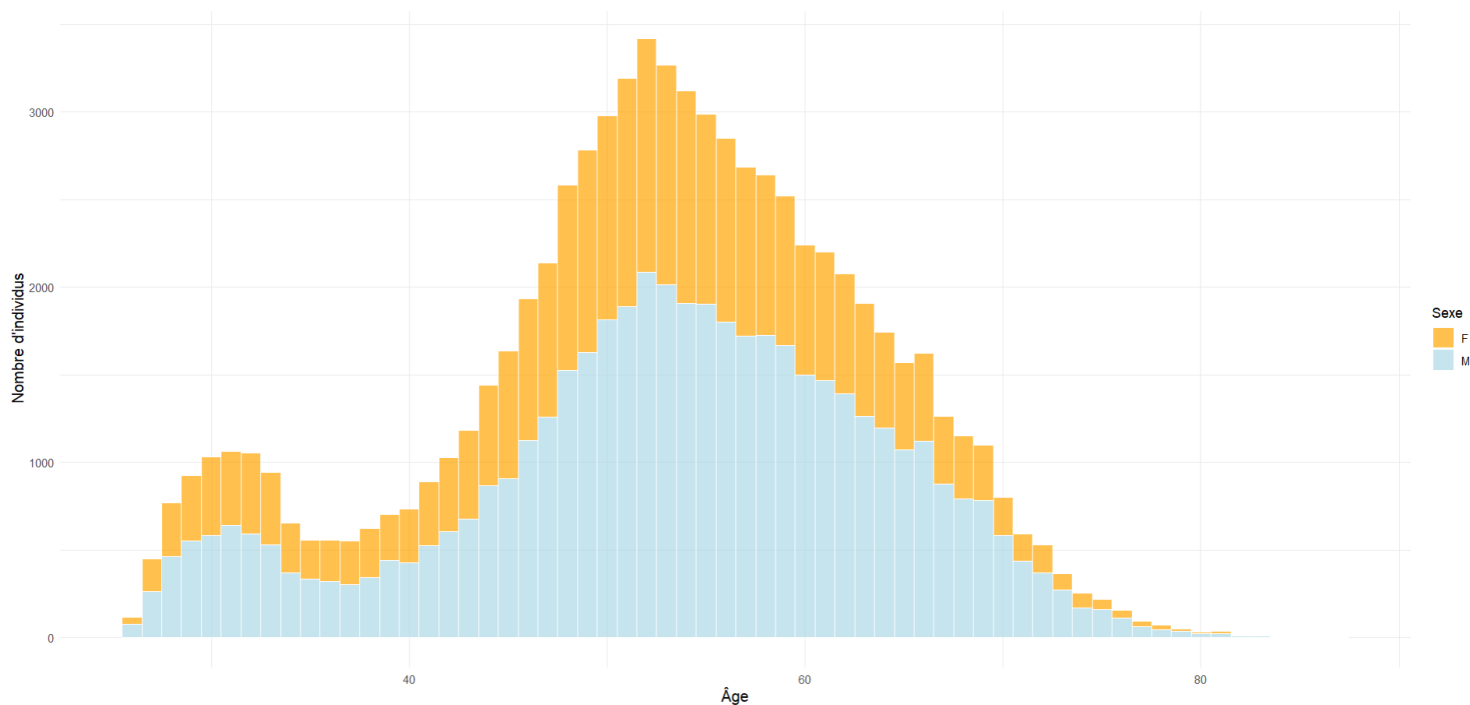


Figure 32 Distribution des âges, par sexe

3.2) Calcul des engagements

Nous avons supposé plusieurs hypothèses dans notre étude :

- 1) L'âge de départ à la retraite est de 65 ans dans les 5 pays.
- 2) Le taux d'actualisation est supposé identique entre les pays et constant dans le temps. Il est égal à 3%.
- 3) Nous utilisons l'Euro comme devise pour tous les pays.

Ces hypothèses ont pour but de mieux isoler l'effet du COVID des autres facteurs externes susceptibles d'influencer notre portefeuille, mais aussi de faciliter les comparaisons entre les pays.

Chaque rentier verse soit une prime unique, soit cotise pendant une période donnée. Ensuite, à son départ à la retraite, il commence à percevoir ses rentes annuelles jusqu'au décès. Dans ce mémoire, nous nous concentrons sur les engagements de l'assureur en 2021 vis-à-vis des rentiers, c'est-à-dire le capital constitutif total nécessaire pour couvrir l'ensemble des rentes du portefeuille.

Pour le calcul de ce dernier, nous avons procédé ainsi :

- 1) Nous avons tout d'abord réalisé des projections en utilisant nos modèles pour obtenir des taux de mortalité m_x jusqu'à l'année 2105, qui correspond à l'année où l'ensemble du portefeuille aurait déjà atteint l'âge 110 ans. Cela correspond à une matrice de dimensions 96 lignes (pour les âges de 15 à 110 ans) et 85 colonnes (représentant les années de 2021 à 2105).
- 2) Ensuite, nous avons calculé les probabilités de survie p_x associées à chaque âge et chaque année projetée.
- 3) Après, nous avons calculé le facteur d'annuités ainsi :
 - Pour les rentes viagères immédiates :

Considérons une rente viagère de 1 euro payable immédiatement à la fin de chaque année tant que le rentier est vivant. La prime unique pour un individu âgé de x est égale à :

$$a_x = \sum_{k=1}^{110} v^k {}_k p_x$$

Avec v le terme d'actualisation suivant la formule $\frac{1}{1+3\%}$.

- Pour les rentes viagères différées :

Pour les individus âgés de moins de 65 ans, la rente n'est pas payable immédiatement mais uniquement à partir du départ à la retraite. Dans ce cas, nous avons utilisé la formule suivante pour une tête âgée de x ans :

$${}_n | a_x = \sum_{k=1}^{110} v^k {}_k p_x - \sum_{k=1}^n v^k {}_k p_x$$

Avec n le nombre d'années restantes jusqu'à la retraite.

- 4) Enfin, après avoir calculé le facteur d'annuités des rentes unitaires (a_x) pour chaque individu, nous avons multiplié cette valeur par le montant contractuel de rentes afin d'obtenir le montant du capital constitutif par rentier.

3.3) Résultats

Nous avons obtenu les engagements totaux suivants pour chaque pays sous chaque scénario :

Pays	Baseline	Covid
France	€ 7 455 987 634	€ 5 723 125 132
Espagne	€ 7 547 653 812	€ 6 774 795 646
Japon	€ 7 535 431 371	€ 7 607 633 309
Canada	€ 7 040 888 704	€ 5 986 263 893
Royaume Uni	€ 7 162 269 933	€ 5 520 577 873

Nous remarquons que sous le scénario Baseline, l'Espagne possède les engagements les plus élevés parmi les 5 pays. Cela suggère que les modèles prévoient une meilleure survie dans ce pays par rapport aux autres. En effet, plus la population vit longtemps, plus l'assureur voit ses engagements augmenter. En revanche, le Canada présente les engagements les moins élevés parmi les pays.

Sous le scénario Covid, c'est le Japon qui affiche les engagements les plus élevés, tandis que le Royaume-Uni enregistre les engagements les plus faibles.

En analysant les écarts relatifs entre les engagements totaux sur la figure 33, nous constatons que l'Espagne et la France ont été les plus impactées par les deux scénarios. En effet, le portefeuille de ces deux pays présente un écart de -30% entre les engagements sous le scénario Baseline et ceux sous le scénario Covid. Autrement dit, un assureur dans ces deux pays adoptant le modèle Covid verrait ses engagements diminuer de 30 % par rapport à ceux qu'il aurait obtenus en considérant le scénario Baseline.

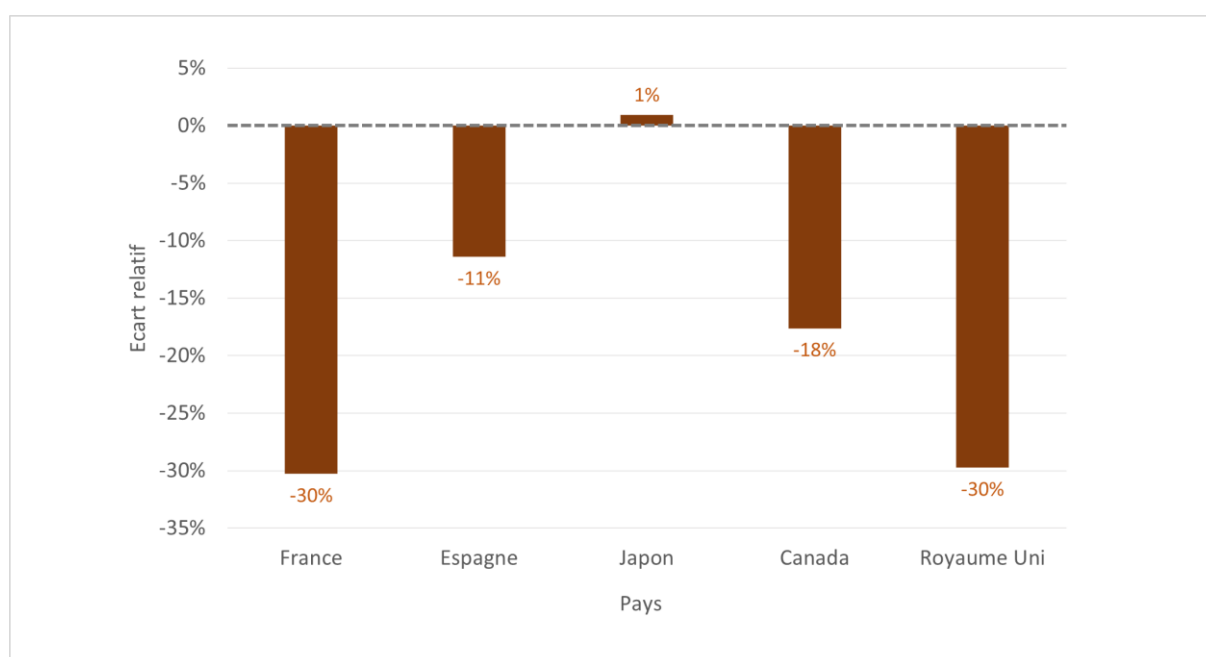


Figure 33 : Ecart relatif entre les engagements totaux sous le scénario Baseline et ceux sous le scénario Covid, par pays

En revanche, le Japon se distingue encore par un écart positif, bien que très faible, avec seulement 1% de variation entre les deux scénarios.

En observant plus en détails ces écarts mais par sexe sur la figure 34, nous pouvons voir que les variations ont été très similaires entre les deux sexes dans chaque pays.

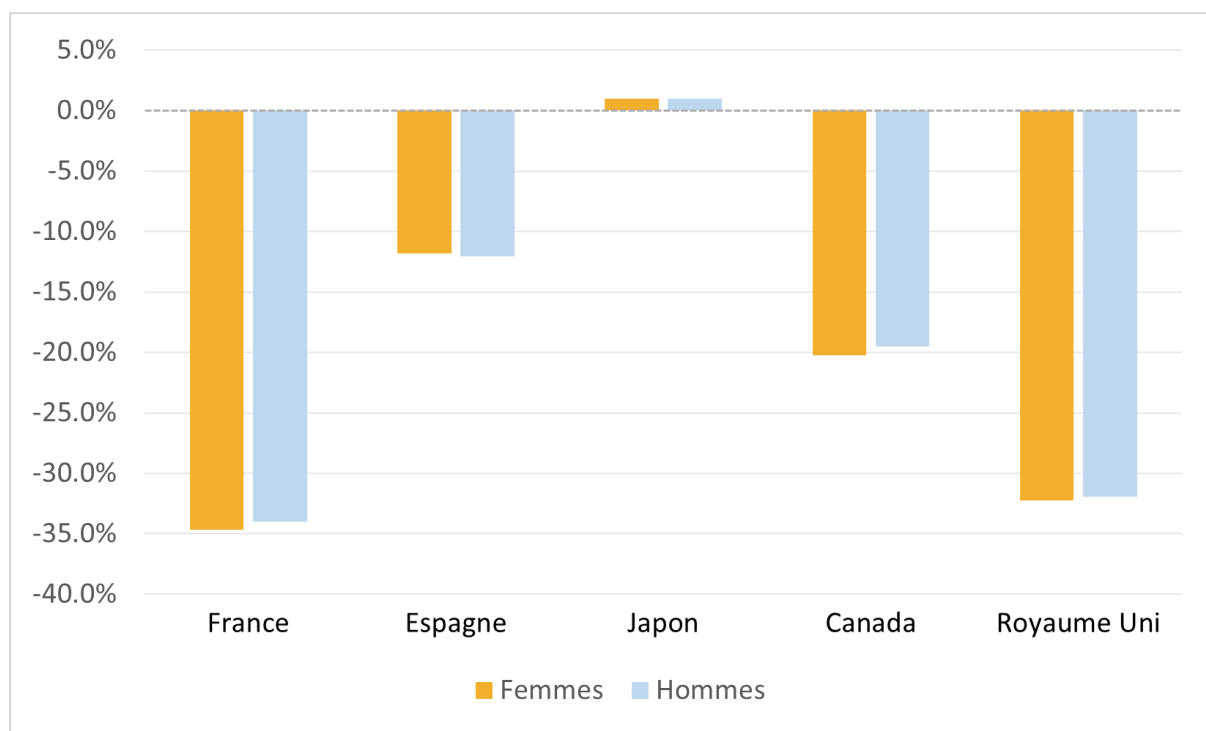


Figure 34 : Ecart relatifs par sexe et par pays

Cependant, ces écarts observés semblent disproportionnés et se baser sur ces résultats pourrait être peu prudent. En effet, le modèle Covid semble, comme nous l'avons vu, pertinent pour l'année 2021, qui a été directement impactée par le choc pandémique. Il pourrait alors être intéressant et prudent de prendre en compte ce modèle pour les engagements vie de court terme, par exemple des contrats de prévoyance collective qui verse un capital en cas de décès sur une période d'un an.

Toutefois, pour les engagements ou les produits de long terme, l'application directe de ce modèle pourrait sous-estimer les engagements des assureurs. En effet, en prenant le choc de la pandémie comme année *jump-off*, ce modèle tend à modifier durablement les tendances de mortalité, comme si l'impact de la crise devait se prolonger au long terme.

C'est pour cette raison que nous proposons une approche plus « réaliste » qui mélange à la fois les projections issues du modèle Baseline et celles issues du modèle Covid. Concrètement, en intégrant les projections COVID pour les premières années de projection, nous allons prendre en compte l'impact immédiat de la pandémie sur la mortalité. Ensuite, en revenant progressivement aux résultats du modèle Baseline, nous supposons un retour progressif à une tendance plus "normale".

Cela permettrait de mieux capter l'impact temporaire de la crise sanitaire tout en conservant une trajectoire de mortalité plus cohérente à long terme.

3.4) Approche de combinaison Baseline/Covid

Nous allons dans cette partie appliquer notre approche sur le portefeuille afin d'obtenir des résultats plus prudents. Nous allons dans un premier temps choisir une méthode de lissage afin d'amortir l'effet du choc sur plusieurs années suivant l'année *jump-off*. Ensuite, nous allons présenter les résultats obtenus.

Méthode de lissage de l'effet de la pandémie

Dans le cadre de notre approche mixte, plusieurs méthodes de lissage peuvent être proposées afin d'assurer une transition progressive entre le modèle COVID et le modèle Baseline. Le but étant d'atténuer le choc pandémique sur une période suivant l'année *jump-off* tout en conservant une tendance plus réaliste au long terme. Ayant calculé les projections long terme issues des deux modèles, nous cherchons un poids r qui nous permettra de réaliser le mélange entre les deux comme suit :

$$m_{x,t} = r \times m_{x,t,covid} + (1 - r) \times m_{x,t,baseline}$$

Nous allons ici présenter deux méthodes :

1) Méthode Linéaire

Cette méthode consiste à réduire progressivement l'impact du modèle COVID en appliquant une diminution fixe du poids de ses projections chaque année. Concrètement, nous diminuons le poids des projections du modèle COVID d'un pourcentage constant $x\%$ sur une période définie, avant de basculer entièrement vers le modèle Baseline.

Le poids du modèle Covid à l'année t , noté r_t , suit la règle suivante :

$$r_t = r_{t-1} - x\%$$

Avec $r_{2021} = 100\%$

Par exemple, si nous choisissons de réaliser l'amortissement sur 5 ans avec une réduction de 20% chaque année, nous obtenons :

Année 1 : 100 % modèle Covid

Année 2 : 80 % modèle Covid, **20 %** modèle Baseline

Année 3 : 60 % modèle Covid, **40 %** modèle Baseline

Année 4 : 40 % modèle Covid, **60 %** modèle Baseline

Année 5 : 20 % modèle Covid, **80 %** modèle Baseline

Cette méthode a pour avantage d'être simple et rapide à appliquer. Cependant, une réduction uniforme du choc pourrait ne pas être entièrement représentative de la réalité. En effet, les pandémies suivent souvent une courbe en "cloche" : une montée rapide des cas et des décès suivie d'un

ralentissement, souvent dû à l'immunité acquise et aux mesures sanitaires. La diminution rapide au début est souvent liée à une combinaison de réponses médicales et comportementales.

Une pandémie pourrait aussi se manifester par plusieurs vagues successives, qui pourraient être dues aux mutations du virus, des changements dans le comportement des populations, ou même dans les mesures sanitaires.

Enfin, le retour aux niveaux habituels dépend de la capacité du pathogène à devenir endémique ou saisonnier, comme c'était le cas pour la grippe espagnole.

2) Méthode Exponentielle

Cette méthode repose sur une décroissance exponentielle de l'influence du modèle COVID : l'effet du choc diminue rapidement au début et plus lentement ensuite.

Le poids du modèle Covid à une année t est définie par :

$$r_t = \exp(-\theta \times (t - t_0))$$

Avec :

- $t_0 = 2021$
- θ un paramètre déterminant la vitesse de décroissance. Plus θ est grand, plus la décroissance est rapide.

Cette méthode nous semble plus réaliste que la méthode linéaire, puisqu'elle modélise de manière plus précise la dynamique d'un choc. Cependant, elle ne prend pas en compte les vagues « multiples » d'une pandémie puisqu'elle suppose une seule phase de décroissance.

Pour notre cas, nous avons choisi d'appliquer cette méthode en définissant un amortissement sur 5 ans. Autrement dit, nous considérons que l'impact du modèle COVID est pris en compte entièrement en 2021, puis son effet diminue progressivement chaque année jusqu'en 2025. Nous avons également choisi un paramètre de vitesse de décroissance θ égal à 0,5, qui permet de moduler la rapidité de cette diminution.

Nous obtenons les poids suivants pour chaque année :

Année 1 : 100 % modèle Covid

Année 2 : 61 % modèle Covid, **39 %** modèle Baseline

Année 3 : 37 % modèle Covid, **63 %** modèle Baseline

Année 4 : 22 % modèle Covid, **78 %** modèle Baseline

Année 5 : 14 % modèle Covid, **86 %** modèle Baseline

Nous pouvons observer sur la figure 35 que les écarts sur notre portefeuille sont significativement moins élevés que si on utilisait uniquement le modèle Covid. Parmi les pays étudiés, le Royaume-Uni présente l'écart le plus élevé, avec une réduction de 0,32% des engagements entre le scénario Baseline et le modèle combiné. Au contraire, le Japon présente l'écart le moins élevé avec une diminution de seulement 0.01%.

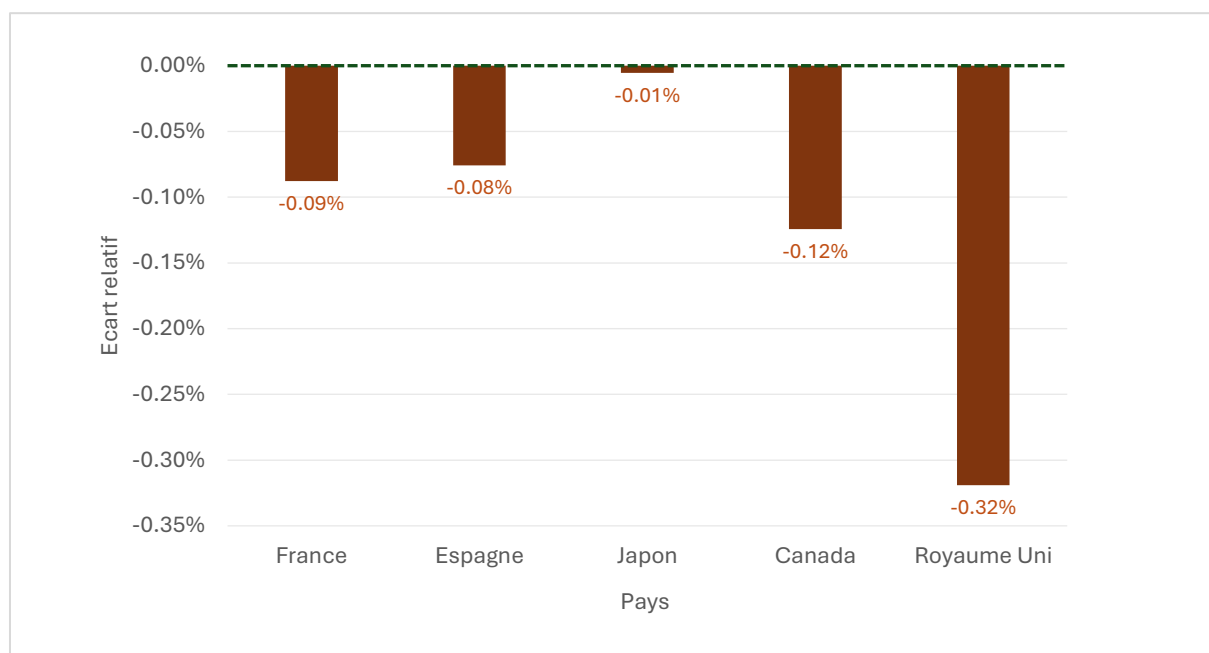


Figure 35 : Ecart relatif entre les engagements totaux sous le scénario Baseline et ceux sous le scénario mixte, par pays

Ces résultats nous semblent plus réalistes et prudents pour les engagements sur long terme. Cependant, une limite de ce modèle est le fait que le choix du paramètre θ et du nombre d'années n est crucial et pourrait influencer les résultats. Une estimation incorrecte de ces paramètres pourrait conduire à une sous-estimation ou une surestimation de la prise en compte et de la réduction du choc.

Afin d'évaluer cette sensibilité, nous avons réalisé une étude sur l'influence de ces paramètres sur l'écart relatif global de notre portefeuille. Nous avons appliqué notre approche en testant plusieurs valeurs pour les paramètres θ et n et nous avons calculé l'écart global en sommant les écarts relatifs dans les 5 pays.

n / θ	0	0,1	0,2	0,4	0,5	0,7	0,9	1	2	10
1	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%	-0,12%
3	-0,61%	-0,54%	-0,48%	-0,39%	-0,36%	-0,30%	-0,26%	-0,24%	-0,16%	-0,12%
5	-1,37%	-1,08%	-0,86%	-0,69%	-0,49%	-0,37%	-0,29%	-0,27%	-0,16%	-0,12%
7	-2,47%	-1,71%	-1,23%	-0,70%	-0,56%	-0,39%	-0,30%	-0,27%	-0,16%	-0,12%
10	-4,87%	-2,79%	-1,71%	-0,80%	-0,61%	-0,40%	-0,30%	-0,27%	-0,16%	-0,12%
85	-88,09%	-11,15%	-2,96%	-0,88%	-0,63%	-0,40%	-0,30%	-0,27%	-0,16%	-0,12%

Nous observons dans la case jaune l'écart observé de l'approche de base que nous avons étudiée, entre les engagements totaux sous le scénario Baseline et ceux sous le scénario Covid. En effet, avec un paramètre de vitesse égal à 0 et un nombre d'années maximal, cela revient à prendre en compte intégralement les projections du modèle Covid sur toute la période.

Dans la case verte, nous retrouvons l'écart relatif global obtenu dans la section précédente, en supposant que l'impact du choc s'atténue sur une période de 5 ans après le choc initial, avec un facteur de vitesse de diminution fixé à 0,5.

Ensuite, nous remarquons que l'écart diminue avec l'augmentation de la vitesse de réduction θ et augmente avec le nombre d'années d'amortissement n . Cela est attendu, car plus l'effet est pris en compte sur une période prolongée, plus l'écart avec le modèle qui ne prend pas cet effet en compte sera important. En revanche, lorsque la vitesse de réduction augmente pour un amortissement sur la même période, l'écart devient plus faible, car l'impact du modèle COVID disparaît plus rapidement, réduisant ainsi la différence avec le modèle Baseline.

De plus, nous observons que l'écart devient insensible au paramètre θ lorsque nous supposons que la transition s'effectue sur 1 an. Cela est dû au fait que dans la méthode exponentielle, nous avons le terme $(t - t_0)$ qui s'annule dans ce cas et permet d'attribuer un poids égal à 100% au modèle Covid à l'année 2021, peu importe le paramètre θ choisi.

Nous observons également que l'écart devient de plus en plus insensible au paramètre n lorsque la vitesse de diminution augmente. En effet, choisir un paramètre θ très élevé reviendrait à prendre en compte le modèle Covid que pour l'année 2021.

Partie IV : Limites de l'étude

Dans cette partie nous allons présenter quelques limites et pistes d'améliorations de notre étude.

Fermeture des tables

Une amélioration possible de notre approche aurait été d'intégrer une fermeture de la table. En effet, comme nous l'avons vu, l'estimation des paramètres de nos modèles nécessite une matrice rectangulaire contenant les taux de mortalité des âges 15 à 110, pour les années de calibration choisies. Dans notre étude, nous avons supposé que les données étaient complètes et fiables pour tous les âges dans les 5 pays. Cependant, en réalité, pour les âges avancés, les données sont parfois rares et plus volatiles en raison du faible effectif de ces populations. C'est pour cela que plusieurs méthodes d'extrapolation de ces taux existent afin d'apporter plus de robustesse et de stabilité aux estimations des paramètres aux âges avancés.

Un exemple d'une méthode courante de fermeture des tables est la méthode de Coale et Kisker. Cette approche permet d'extrapoler les taux pour les âges élevés (110 dans notre cas) en utilisant la formule suivante :

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

avec z_x le taux moyen de croissance de μ_x entre 65 et x ans.

L'idée est qu'on calcule les coefficients z_x jusqu'à un certain âge, puis on les extrapole afin de pouvoir recomposer les taux μ_x

A partir d'une étude réalisée sur 7 pays, Coale et Kisker ont montré que le taux moyen de croissance arrive à un pic vers 80 ans (ou 85 selon les pays) ensuite décroît linéairement. Par conséquent, pour tout âge > 80 , il est possible de supposer que :

$$z_x = z_{80} + s(x - 80)$$

$$\text{avec } s = -\frac{1}{465} \left(\ln \left(\frac{\mu_{79}}{\mu_{110}} + 31z_{80} \right) \right) \text{ obtenu en calculant } \mu_{110}$$

$$\text{et } z_{80} = \frac{1}{15} \left(\ln \left(\frac{\mu_{80}}{\mu_{65}} \right) \right)$$

Par ailleurs, Coale et Kisker ont fixé arbitrairement $\mu_{110} = 0.8$ chez les femmes et $\mu_{110} = 1$ chez les hommes. A partir de cela, il serait alors possible de calculer les taux au-delà de 80 ans.

Indisponibilité des données réelles au-delà de 2021

- Dans notre étude, nous avons pu comparer les projections issues de nos modèles avec les données réelles de 2021, ce qui nous a permis d'identifier le modèle offrant les estimations les plus proches de la réalité pour cette année. Toutefois, une analyse sur une année reste limitée car elle ne permet pas d'évaluer l'évolution de la pertinence des modèles dans le temps. Il aurait été donc intéressant de pouvoir réaliser cette comparaison sur plusieurs années suivant le choc initial. En effet, cela aurait permis de voir si un modèle resterait plus représentatif de la réalité que l'autre et pendant combien de temps. De plus, une telle analyse aurait permis de mieux comprendre la dynamique du retour à la « normale », facilitant ainsi le choix des paramètres de notre approche mixte.

Cependant, les données réelles accessibles sur *Human Mortality Database* n'étaient pas toutes disponibles au-delà de l'année 2021 pour plusieurs des pays étudiés.

- Une autre analyse intéressante dans le cas de disponibilité des données au-delà de l'année 2021 aurait été d'étudier l'impact de la pandémie sur les estimations des espérances de vie résiduelles, mais cette fois-ci sans que le choc soit en année *jump-off*. En effet, dans notre mémoire nous avons principalement étudié l'effet du choc sur les estimations lorsque celui-ci se produit en dernière année de calibration, et nous avons vu comment cela influence les projections issues du modèle. Cependant, avec les données réelles disponibles plusieurs années après le choc initial de 2020, il serait pertinent de mener la même étude en considérant ce choc initial survenu quelques années avant l'année *jump-off*.

Conclusion

L'analyse menée dans ce mémoire met en lumière l'impact significatif du choc pandémique de la Covid-19 en année *jump-off* sur les estimations des espérances de vie résiduelles, avec des conséquences notables sur la valorisation des engagements des assureurs. La prise en compte de ce choc dans les modèles de mortalité peut varier considérablement selon les pays et le sexe, soulignant l'importance d'une approche adaptée à chaque contexte démographique.

La prise en compte du choc pandémique en année *jump-off* s'est révélée pertinente pour modéliser les mortalités à court terme, en particulier pour les années immédiatement touchées par la pandémie. Cette approche est particulièrement utile pour les engagements à court terme, où la précision des projections de mortalité sur un horizon proche est cruciale, notamment pour des produits comme les contrats de prévoyance collective sur une durée de 1 ou 2 ans.

Cependant, pour les engagements à long terme, l'intégration du choc en année *jump-off* peut conduire à une sous-estimation exagérée des engagements. En effet, inclure ce choc en dernière année de calibration modifie les tendances de mortalité à long terme, ce qui ne reflète pas nécessairement la réalité future. Les assureurs doivent donc faire preuve de prudence et envisager des approches plus réalistes.

Une approche mixte, combinant les projections des modèles Baseline et Covid peut être envisagée et peut offrir une solution plus équilibrée. Par exemple, une atténuation exponentielle, caractérisée par une diminution rapide de l'influence du choc pandémique semble garantir une estimation plus réaliste et plus prudente des engagements futurs. En effet, elle permet de capturer l'impact immédiat de la pandémie tout en revenant à une tendance normale à long terme, jugée plus adéquate.

A noter qu'une telle approche reste sensible à la méthode et aux paramètres de mixture choisis comme cela a pu être montré par des tests de sensibilité. En effet, le nombre d'années pendant lesquelles les projections issues du scénario Covid sont prises en compte, ainsi que la vitesse d'amortissement du choc doivent être choisis : la livraison de nouvelles données de mortalité sur les années suivantes pourrait permettre de corroborer ou écarter l'approche et de faciliter le choix des paramètres.

Bibliographie

- **Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., & Dowd, J. B.** (2021). *Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis*
- **Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I.** (2009). *A quantitative comparison of stochastic mortality models using data from England and Wales and the United States*. *North American Actuarial Journal*, 13(1), 1-35.
- **Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M.** (2011). *Mortality density forecasts: An analysis of six stochastic mortality models*. *Insurance: Mathematics and Economics*, 48(3), 355-367.
- **Currie, I. D.** (2016). *On Fitting Generalized Linear and non-linear Models of Mortality*. *Scandinavian Actuarial Journal*, (4), 356–383.
- **Huang, G., Guo, F., Zimmermann, K. F., Liu, L., Taksa, L., Cheng, Z., Tani, M., & Franklin, M.** (2021). *The effect of the COVID-19 pandemic on life expectancy in 27 countries..*
- **Islam, N., Jdanov, D. A., Shkolnikov, V. M., Khunti, K., Kawachi, I., White, M., et al.** (2021). *Effects of COVID-19 pandemic on life expectancy and premature mortality in 2020: Time series analysis in 37 countries*. Disponible sur *BMJ* 2021.
- **José Manuel Aburto, Jonas Schöley, Ilya Kashnitsky, Luyin Zhang, Charles Rahal, Trifon I Missov, Melinda C Mills, Jennifer B Dowd, Ridhi Kashyap,** Quantifying impacts of the COVID-19 pandemic through life-expectancy losses: a population-level study of 29 countries, *International Journal of Epidemiology*, Volume 51, Issue 1, February 2022, Pages 63–74, <https://doi.org/10.1093/ije/dyab207>
- **Lee, R. D., & Carter, L. R.** (1992). *Modeling and Forecasting U.S. Mortality*. *Journal of the American Statistical Association*, 87(419), 659–671.
- **Ouhammou, A.** *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*. Disponible sur *Mémoire IA*.
- **Remund, A., Camarda, C., & Riffe, T.** (2021). *La surmortalité des jeunes adultes est-elle naturelle ? Population & Sociétés*, 590(6), 1-4. <https://doi.org/10.3917/popsoc.590.0001>.

- **Schnürch, S., Kleinow, T., Korn, R., & Wagner, A.** *The impact of mortality shocks on modelling and insurance valuation as exemplified by COVID-19.* Disponible sur IFOA : *Annals of Actuarial Science*.
- **Turcksin, C.** (2022). *The impact of Covid-19 on the mortality of the Belgian population.* Faculté des sciences, Université catholique de Louvain. Disponible sur Mémoire Institut de Statistique, Biostatistique et Actuariat.
- **Turner, H., & Firth, D.** (2015). *Generalized Nonlinear Models in R: An Overview of the gnm Package.* R package version 1.0-8, Disponible sur <http://CRAN.R-project.org/package=gnm>.