

Impact du montant de rente sur la longévité d'un portefeuille de rentiers

Une approche par les modèles additifs généralisés

Christian Borel WAFO KANKEU

Wébinaire Institut des Actuaires

22 mai 2024



EURIA
Euro-Institut d'Actuariat



Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

- Situation vis à vis de la référence réglementaire

4 Résultats

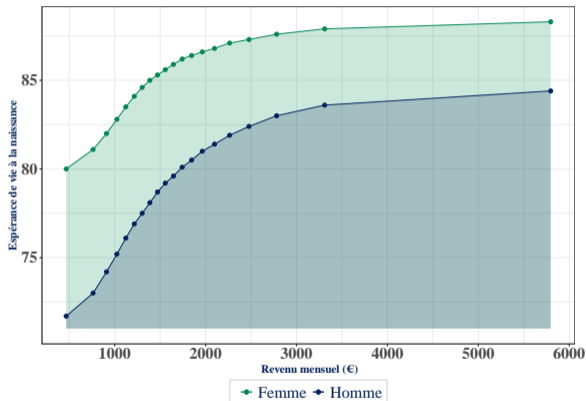
- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

6 Conclusion

Contexte

- Le **risque longévité** c'est le risque pour un assureur sur un portefeuille donné, d'observer une augmentation non anticipée de l'espérance de vie aux grands âges.
- Selon l'INSEE, la durée de vie augmente de génération en génération. L'espérance de vie à la naissance des hommes en 2022 est ainsi supérieure de plus de **9 ans** par rapport à celle de 1980.
- Les individus les plus aisés vivent en moyenne plus longtemps que les plus modestes.
- Les 5% des hommes les plus aisés, ont une espérance de vie à la naissance supérieure de **13 ans** par rapport aux 5% les plus pauvres. Cet écart est de **8 ans** pour les femmes



Problématique et objectif

- Nous disposons d'un portefeuille de rentiers de l'assureur **Swiss Life**
- L'assureur souhaiterait mesurer les différences de longévité entre individus liées au niveau de vie
- Le niveau de vie est approché par le **montant de la rente** perçue
- Le pire des scénarios ici étant de sous estimer la longévité des plus riches, et donc de payer plus longtemps que prévu les rentes les plus élevées.

💡 Objectif principal

L'objectif principal de l'étude menée ici sera ainsi de tenir compte du **niveau de vie** (au moyen du **montant de la rente** perçue) dans la construction d'une **table de mortalité d'expérience prospective** pour les assurés de ce portefeuille de rentiers.

Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

- Situation vis à vis de la référence réglementaire

4 Résultats

- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

6 Conclusion

Construction d'une table de mortalité prospective

Deux principales approches sont généralement utilisées pour la construction des tables prospectives:

- Une approche **intrinsèque** basée exclusivement sur les données dont dispose l'assureur
- Une approche par référence **externe** qui lie la mortalité du portefeuille à une référence externe. Elle se décline comme suit :

- 1 Calcul des taux bruts de décès
- 2 Lissage des taux bruts de décès
- 3 Liaison à la référence externe
- 4 Fermeture de la table aux grands âges

! Limites de ces approches

Elles ne permettent pas la prise en compte de variables explicatives autres que l'âge dans la modélisation de la mortalité

💡 Les modèles additifs généralisés

Ils se présentent comme le candidat idéal pour mener à bien l'objectif qui est le notre ici, ils permettent :

- Prise en compte des impacts de variables
- Application directe de lissage suivant les variables explicatives
- Mise en **offset** de la référence réglementaire
- Extrapolation du modèle aux grands âges

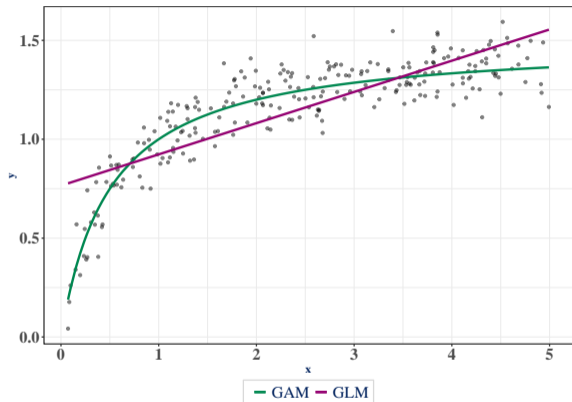
Principe

Un modèle additif généralisé en abrégé **GAM** (**G**eneralized **A**dditive **M**odel) est un modèle linéaire généralisé avec des prédicteurs linéaires qui sont des fonctions **lisses** des variables explicatives.

$$g(\mu_i) = A_i\theta + f_1(x_{1i}) + f_2(x_{2i}) + \dots$$

Avec:

- $\mu_i \equiv \mathbb{E}(Y_i)$, Y_i étant la variable d'intérêt de loi appartenant à la **famille exponentielle**
- g une **fonction de lien**
- $A_i\theta$ correspondant à la **partie paramétrique** du modèle, où A_i est une ligne de la matrice de design, et θ le vecteur de paramètre associé
- f_j une **fonction lisse** associée à la covariable x_j



Ces modèles permettent une plus grande **flexibilité** par rapport aux GLM dans la matérialisation des dépendances entre les variables quantitatives et la variable d'intérêt.

Représentation des fonctions lisses

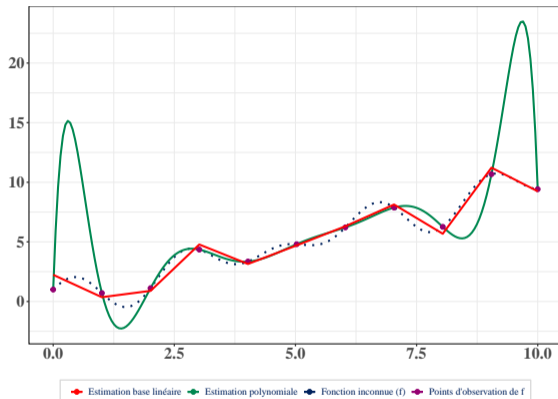
Les fonctions lisses f sont représentées dans des bases de fonctions élémentaires b_k .

$$f(x) = \sum_{k=1}^q b_k(x) \times \beta_i$$

Avec β_i des paramètres à estimer. On se rapproche ainsi d'une forme linéaire classique.

La spécification de la fonction lisse f se ramène donc au choix de la base dans laquelle elle sera représentée.

- Base de **fonctions polynomiales**
- Base de **fonctions linéaires par morceaux**
- Base de **splines** : splines cubiques, P-splines, thin plate regression spline



Dans le cas des splines, le choix de la famille de base n'est pas très déterminant.

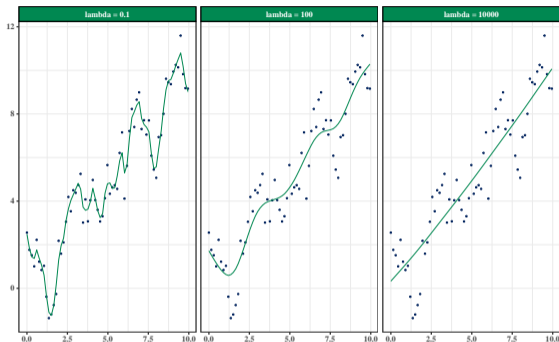
Estimation des paramètres et degré de lissage

Les paramètres du modèle sont estimés par maximisation de la fonction de **logvraisemblance pénalisée** suivante :

$$l_p(\beta) = l(\beta) - \frac{1}{2\phi} \sum_j^p \lambda_j \beta^T S_j \beta$$

Avec :

- $l(\beta)$ la log-vraisemblance associée à la distribution $EF(\mu, \phi)$ et aux observations y_1, \dots, y_n de la variable réponse Y
- λ_j les paramètres de lissages associés aux fonctions lisses.



Les paramètres λ_j **contrôlent la régularité** des fonctions lisses. Leurs valeurs sont choisies de façon à minimiser un critère statistique, ici le **GCV**:

$$GCV = \frac{n \sum_{i=1}^n (y_i - \hat{f}_i)^2}{[n - \text{tr}(A)]^2}$$

Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

● Situation vis à vis de la référence réglementaire

4 Résultats

- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

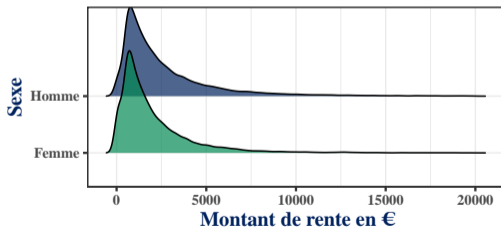
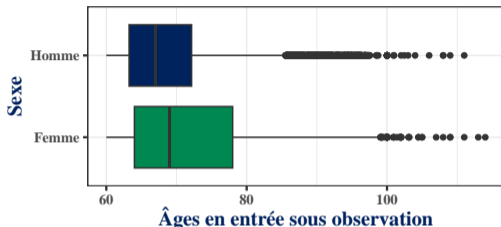
6 Conclusion

Présentation des données

Les données

Les données à disposition proviennent d'un portefeuille de rentes viagères en cours de restitution de **Swiss Life** :

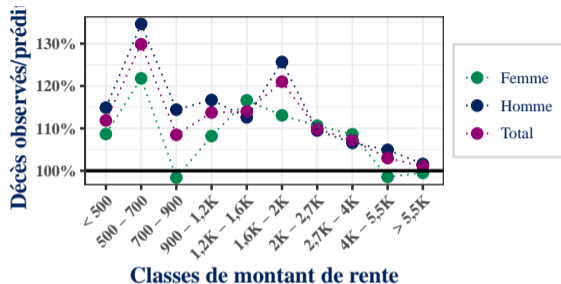
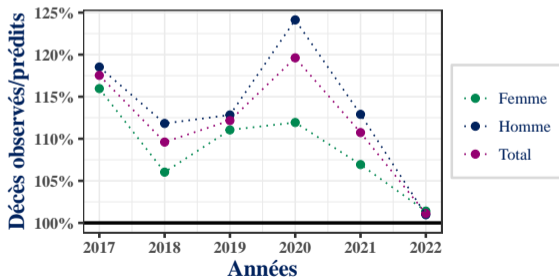
- **6 ans** d'observations, de **Janvier 2017** à **Décembre 2022**
- **56 220** assurés, **67%** d'hommes et **33%** de femmes
- **6 716** décès dont **63%** d'hommes et **37%** de femmes
- Âge moyen d'entrée en observation de **71** ans pour les femmes et de **68,5** ans pour les hommes
- Âge moyen au décès de **87** ans chez les femmes et de **81** ans pour les hommes
- **97%** des montants de rente inférieurs à **10 000 €**



Situation vis à vis de la référence réglementaire

Comparaison aux tables TGH-TGF 05

- Surmortalité de **10%** par rapport aux tables réglementaires
- Une mortalité plus importante en **2020**



Écarts par classe de montant de rente

- Tendence à la baisse de la mortalité avec le montant de rente
- Fluctuations importantes d'une classe à l'autre

Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

- Situation vis à vis de la référence réglementaire

4 Résultats

- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

6 Conclusion

Spécification du modèle

- Départ de la **log-vraisemblance** d'un modèle de durée avec données **tronquées à gauche** et **censurées à droite** : $l(\theta) = \sum_{i=1}^n \left(\delta_i \log(h_{\theta}(t_i)) - \int_0^{t_i} h_{\theta}(t) dt \right)$, avec une **pénalisation des paramètres** dont dépend la fonction de hasard.
- Agrégation des données, sous hypothèse de **force de mortalité constante**
- Obtention d'un modèle parfaitement équivalent à un **GAM Poisson**
- En variable cible le nombre de décès, comme variables explicatives on a l'âge, le montant de rente et le sexe
- L'exposition centrale au risque et la force de mortalité des tables réglementaires sont mises en **offset**.

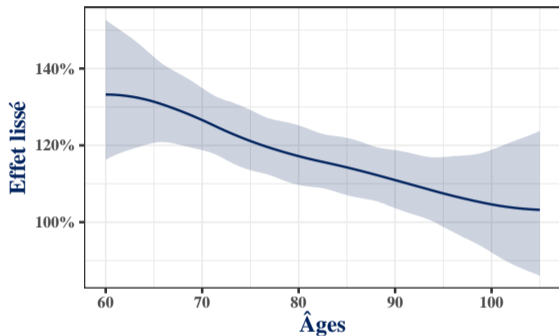
$$D \sim \mathcal{P}(h_{\theta} \times ec), \text{ et } \log(\mathbb{E}(D)) = 1 \times \log(ec) + 1 \times \log(\mu_{reg}) + \beta_0 + \beta_1 \times s + S(x) + S(m_2)$$

Soit :

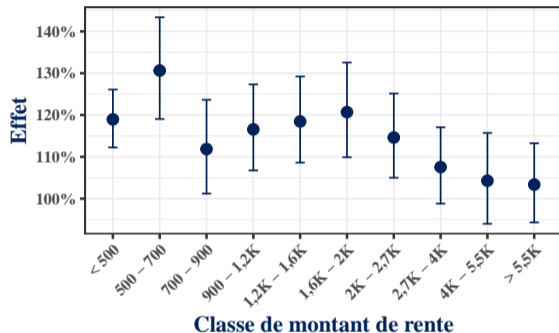
$$\frac{\mathbb{E}(D)}{ec} = \mu_{reg} \times \exp(\beta_0 + \beta_1 \times s + S(x) + S(m_2))$$

- Les coefficients passés à l'exponentiel sont des **facteurs multiplicatifs** à appliquer à la force de mortalité des **tables réglementaires** pour aboutir à la **force de mortalité du portefeuille**

Effets des variables explicatives

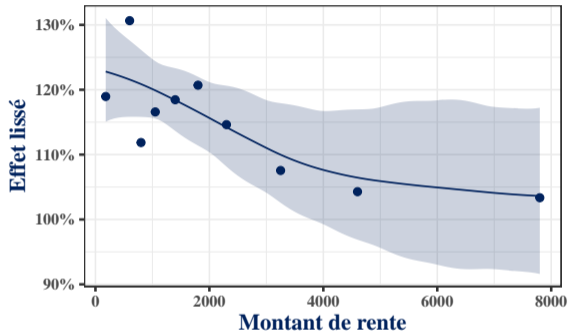


- Décroissance des écarts par rapport à la référence réglementaire avec l'âge
- +32% de correctif à appliquer à la référence réglementaire pour aboutir à la force de mortalité du portefeuille à 60 ans

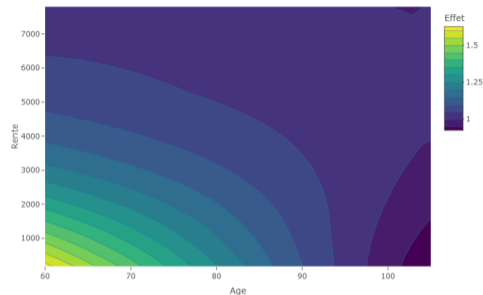


- Une tendance à la baisse des écarts avec le montant de rente
- Des fluctuations erratiques d'une classe à l'autre, bande de confiances larges
- Nécessité de **lisser**

Effets des variables explicatives



- Capture plus nette de la tendance à la baisse de la mortalité avec le montant de rente
- Élimination des fluctuations d'une classe de rente à l'autre



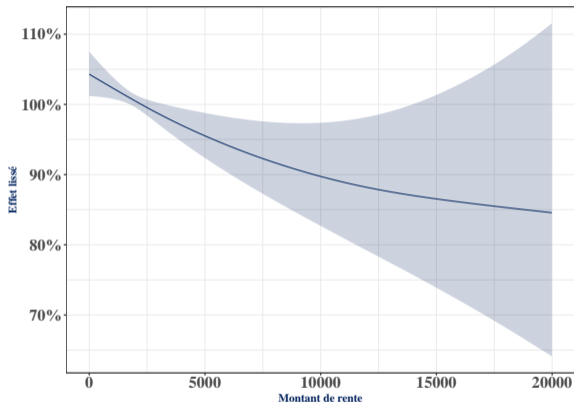
- Effet d'interaction lissé entre âge et montant de rente
- Les différences liées au montant de rente se dissipent à mesure que l'âge augmente
- Aux grands âges la mortalité est exclusivement déterminée par des facteurs biologiques

Modèle de Cox

Le modèle de Cox est le modèle de durée de référence lorsqu'il est question d'évaluer l'impact d'une variable explicative sur la mortalité.

- Il s'agit d'un modèle de semi-paramétrique avec une fonction de hasard qui est de la forme :
$$h_X(x|z; \theta) = \exp(z^t \theta) h_0(x)$$
- Il repose sur l'hypothèse de **hasard proportionnel** qui postule la constance dans le temps des effets des variables explicatives
- Il est possible d'inclure des effets lisses de variables quantitatives. La fonction de hasard devient alors :
$$h_X(t|(x, s, m_2)) = \exp(S_1(x) + \beta \times s + S_2(m_2)) h_0(t)$$

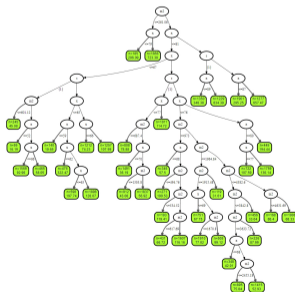
i La fonction de hasard est une probabilité instantanée de décès par unité de temps



- Tendence à la baisse avec le montant de rente du facteur multiplicatif à appliquer au hasard de base

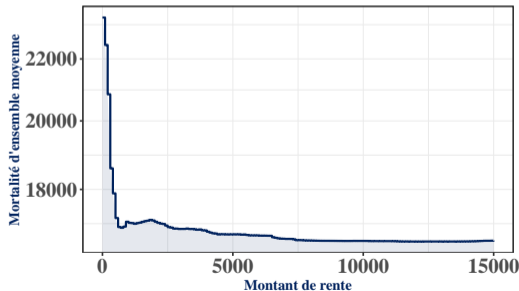
Forêts aléatoires

- 1 **Segmentation de la population** en sous-groupes ayant des comportements **homogènes en terme de survie**.
- 2 Calcul d'une fonction de hasard cumulée dans chaque sous groupe constitué (Estimateur de **Nelson Aalen**)
- 3 La fonction de hasard cumulée globale est obtenue en faisant une **moyenne sur tous les arbres de la forêt**



L'effet des variables explicatives est représenté grâce à la **mortalité d'ensemble**, il s'interprète comme un nombre de décès théorique si tous les individus avaient les mêmes caractéristiques :

$$M_i = \mathbb{E}_i \left(\sum_{j=1}^n H(T_j | X_i) \right)$$



Comparaison des modèles

Ayant des modèles calibrés sur des données de formats différents, il faut un critère global permettant de les comparer :

$$V = \sum_{i=1}^n \left(d_i \log(p_i) + (1 - d_i) \log(1 - p_i) \right)$$

Avec :

- d_i l'indicatrice de décès sur la période pour l'individu i
- p_i la probabilité de décès sur la période d'observation pour l'individu i , n le nombre d'individus

Modèles	R2	R2 ajusté	AIC	Vraisemblance (V)	edf
Poisson données individuelles	16,2 %	16,19 %	42987	-21 482	12,24
Cox	15,97 %	15,97 %	43097	-21 540	9,61
Poisson données agrégées	15,77 %	15,76 %	43201	-21 592	9,73
Tables avec coeff d'abattement	15,58 %	15,58 %	43279	-21 639	1
Méthode de Brass	15,34 %	15,34 %	43404	-21 701	2
Forêts aléatoires	14,67 %	14,55 %	44482	-21 873	368,752
Intercept seul	0 %	0 %	51267	-25 633	1

Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

- Situation vis à vis de la référence réglementaire

4 Résultats

- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

6 Conclusion

Provisions avec différents modèles

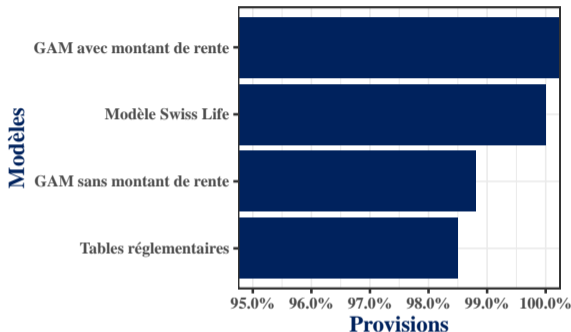
Expression de la PM de rente

$$PM = Rente \times a_x$$

$$\begin{aligned} a_x &= \sum_{k=1}^{N-x} v^k \times_x P_k \\ &= \sum_{k=1}^{N-x} v^k \times \mathbb{P}(T > k+x | T > x) \\ &= \sum_{k=1}^{N-x} v^k \times \frac{\mathbb{P}(T > k+x)}{\mathbb{P}(T > x)} \end{aligned}$$

Avec :

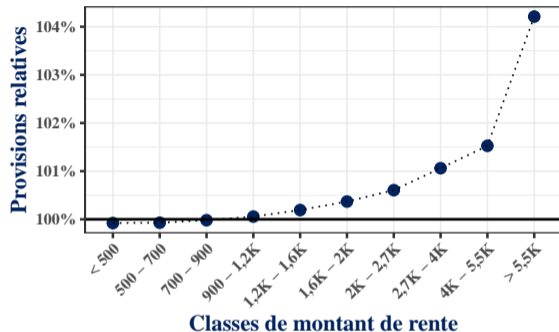
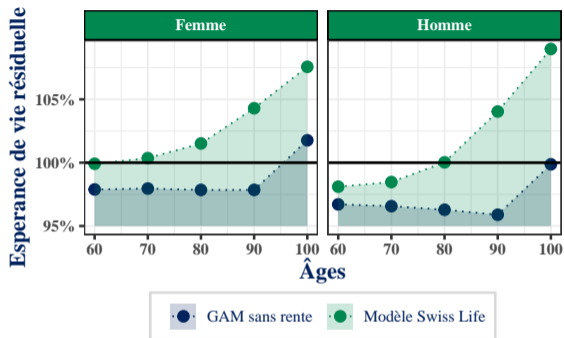
- $v = \frac{1}{1+i}$: le facteur d'actualisation, et i le taux technique en vigueur
- x : l'âge du rentier à la date de calcul



Comparaison des provisions

- Des provisions Swiss Life assez prudentes
- Modèle avec rente plus prudent

Provisions : Origine des écarts



💡 Différences entre les modèles

- Modèle Swiss Life prudent aux grands âges
- Des espérances de vie supérieures aux GAM

💡 Effet montant de rente

- Des écarts de provisions plus importants vers les rentes élevées
- In fine des provisions plus prudentes

Table des matières

1 Introduction

- Contexte
- Problématique et objectif

2 Méthodologie

- Construction d'une table de mortalité prospective
- Les modèles additifs généralisés

3 Analyses descriptives

- Présentation des données

- Situation vis à vis de la référence réglementaire

4 Résultats

- Spécification du modèle
- Effets des variables explicatives
- Approches alternatives
- Comparaison des modèles

5 Calcul de provisions

6 Conclusion

Conclusion

Bilan

Il a été question de mesurer l'impact du niveau de vie (au moyen du montant de rente) sur la longévité d'un portefeuille de rentiers :

- Il en ressort une longévité plus importante à mesure que le montant de rente augmente
- La prise en compte de cet effet dans la construction d'une table de mortalité prospective aboutit à des provisions plus prudentes que celles qui étaient évaluées jusqu'ici par Swiss Life

Axes d'amélioration

- L'utilisation du montant de rente comme proxy du niveau de vie peut induire des biais. Le code postal aurait pu être utilisée en complément d'information sur le niveau de richesse des individus
- Une indicatrice aurait pu être utilisée pour essayer de tenir compte de la période Covid dans l'étude.

Fin de la présentation

Merci pour votre attention !



EURIA
Euro-Institut d'Actuariat

