HAL
open science

Modélisation stochastique du résultat d'une garantie dépendance

Pierre Kubiak

▶ To cite this version:

Pierre Kubiak. Modélisation stochastique du résultat d'une garantie dépendance. Économie et finance quantitative [q-fin]. 2014. dumas-02489605

HAL Id: dumas-02489605
https://dumas.ccsd.cnrs.fr/dumas-02489605v1

Submitted on 7 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire présenté devant

l'UFR de Mathématique et d'Informatique

pour l'obtention du Diplôme Universitaire d'Actuaire de Strasbourg

et l'admission à l'Institut des Actuaires

le 01/10/2014

Par : Pierre Kubiak

Titre: Modélisation stochastique du résultat d'une garantie dépendance

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

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

Membres du jury de l'Institut des Actuaires signature

Membres du jury :

M. Jean Bérard
M. Karl-Théodor Eisele
M. David Fitouchi
M. Jacques Franchi
M. Jean-Lucien Netzer
M. Alexandre You

Invités :

M. David Dubois
Mme Stéphanie Foata
M. Christian Hess
Mme Sylvain Gadenne
Mme Magali Kelle-Vigon
M. Myriam Maumy-Bertrand
Mme Jean Modry
M. Vincent Vigon

Secrétariat : Mme Stéphanie Richard

Bibliothèque : Mme Christine Disdier

Entreprise :

Nom : Hannover Re

Signature :

Directeur de mémoire en entreprise :

Nom : Célya Meunier

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

Le vieillissement de la population représente une problématique grandissante pour les pays développés. Ses risques ne se limitent pas à des problèmes de financement de la retraite. En effet, il se pose aussi la question de l'état de santé de ces personnes âgées, avec des ressources faibles et nécessitant peut être une aide extérieure, un aménagement du domicile voir un placement en établissement spécialisé.

Le risque financier lié à la survenue de la dépendance, couvert faiblement par l'Etat, peut être prévenu par la souscription de garantie dépendance.

Néanmoins, la tarification de ce type de garantie présente un risque, lié à la méconnaissance des lois de probabilité régissant l'entrée en dépendance et la durée de survie dans cet état.

C'est dans ce contexte que se situe l'objet du mémoire. Le but va être de vérifier qu'une garantie dépendance commercialisée vers la fin des années 90 ne présente pas un risque de perte certain lié à une sous évaluation des lois précitées.

Pour ce faire, nous estimerons ces lois sur les observations des sinistres des assurés ayant souscrit cette garantie.

Par la suite, grâce à ces lois, nous projetterons le résultat futur de cette garantie par avec des simulations de Monte Carlo.

Nous nous apercevrons alors que ce risque de perte est justifié et nous proposerons alors une solution pour y faire face.

Population ageing is a growing issue for developed countries. Its risks are not limited to problems of retirement funding. Indeed it also raises the question of the health of these elderly people with limited resources that can require non-medical care services at home, home adaptation or institutionalization. The financial risk associated with the onset of loss of autonomy is poorly covered by the State but can be prevented by the subscription of long term care coverage.

However the pricing for this kind of coverage is risky for the insurer due to poor knowledge of probability's laws for the onset of dependency and survival time in this state.

This memory aims to assess if a long term care covers marketed in the late 90's does not present a certain risk of loss because of an underestimation of the laws above cited.

Firstly we will estimate these laws with observations of policyholders' sinisters.

Then we will be able to project the future result of this cover with Monte Carlo Simulations.

Finally we will find that the risk of loss is justified and we will propose a solution to address it.

Mots clés : Kaplan Meier, Dépendance, Whittaker Henderson, Monte Carlo, table d'expérience, splines, modèle relationnel

Remerciements

J'adresse mes remerciements aux personnes qui m'ont aidé dans la réalisation de ce mémoire.
En premier lieu, je tiens à remercier Célya Meunier, maître de stage en entreprise, pour son suivi et ses conseils tout au long de ce stage. Je la remercie chaleureusement ainsi que Claude Vercasson de m'avoir accueilli pour effectuer ce stage et de la confiance qu'ils m'ont accordée et m'accorde pour l'avenir.
Mes remerciements se dirigent aussi vers tous mes collègues de Hannover Re. Leur bonne humeur a rendu le travail particulièrement agréable. Je remercie en particulier Jonathan Legros et Cristophe Guilloux, pour leur aide et réflexion sur ce travail.
Je remercie vivement ma tutrice universitaire, Angélique Dillinger, pour son suivi, ses conseils et ses nombreuses relectures, m'ayant permis sans cesse d'améliorer mon travail. Son expertise m'a été d'une grande aide. Chacun de nos échanges a été un plaisir, que j'espère partagé.
Pour finir, un grand merci à mes relecteurs pour leur aide dans la correction de ce document.

Section I

Introduction

Les pays industrialisés sont témoins du vieillissement de plus en plus important de leurs populations. D'après l'INSEE, en 2060 une personne sur trois aura plus de 60 ans contre environ 1 sur 5 aujourd'hui. Plusieurs facteurs peuvent expliquer ce phénomène, notamment un taux de fécondité trop faible, la progression constante de l'hygiène de vie et les progrès de la médecine.

Ce vieillissement s'accompagne toutefois d'effets socio-économiques délétères. D'une part, il pose le problème du financement des retraites. D'autre part, ce vieillissement s'accompagne également de problèmes de santé liés à l'âge et le problème de la perte d'autonomie des individus âgés se pose alors. Par ailleurs, la désolidarisation des liens familiaux observables dans nos sociétés actuelles entraîne un isolement de ces individus en mauvaise santé. Ils deviennent alors dépendants d'une aide extérieure ou doivent être placés en maison médicalisée.

Cette dépendance a un coût élevé, par exemple, selon la FFSA, le placement en maison médicalisée coûte environ 2800 euros par mois.

Ce phénomène, déjà important, va prendre de l'ampleur. En effet, 1.13 million de personnes seraient dépendantes à ce jour en France et ce chiffre est amené à doubler d'ici 2040 selon le Ministère de la Santé.

L'Etat peine à protéger la population contre ce risque coûteux. C'est pourquoi les assurances ont commencé à proposer des couvertures pour répondre à la demande de protection financière contre ce risque vers le milieu des années 80. Toutefois, afin d'assurer un risque et demander une cotisation juste, l'assureur a besoin de connaître le coût probable que la couverture peut entraîner pour lui. L'observation de ce risque étant récente, peu de données étaient disponibles, entraînant ainsi une incertitude concernant la tarification. La crainte actuelle sur ces portefeuilles est la possible sous estimation de ce coût, ce qui pourrait entraîner des pertes pour les assureurs.

Conscient de ce possible risque de sous tarification, l'assureur se protège lui même, en souscrivant des couvertures de réassurance. Cela atténue ainsi son risque et le fait bénéficier de l'expérience du réassureur quant à la tarification de telles garanties.

Dans ce contexte d'incertitude quant à la rentabilité de tels produits, l'objectif de ce mémoire sera d'étudier du point de vue du réassureur une garantie dépendance proposée depuis la fin des années 90. Plus précisément, nous tacherons d'évaluer si la commercialisation de ce produit ne présente pas un risque de pertes pour l'assureur et le réassureur lié à une sous estimation du coût de la dépendance. Pour ce faire, nous projetterons le compte de résultat du réassureur grâce à des simulations de la vie des assurés du portefeuille. Ces projections devront être le plus conforme possible à la réalité.

Ainsi, dans une première partie et après avoir introduit le cadre général de la dépendance en France, nous présenterons la théorie utile à l'estimation de lois de probabilités. Ces lois de probabilités nous permettront de simuler la durée de vie

des assurés autonomes et dépendants par la suite.

Les lois estimées présentant souvent des irrégularités, nous introduirons alors des méthodes permettant d'une part d'interpoler ces données lorsque l'estimation n'a pas été possible durant certaines périodes et d'autre part, de les lisser. Néanmoins, une problématique se pose généralement lors de l'estimation de ces différentes probabilités : le manque d'observations de personnes très âgées (plus de 90 ans). Nous introduirons alors des méthodes permettant d'estimer la mortalité pour ces âges ou de l'indexer sur la mortalité d'une population utilisée comme référence. Nous présenterons également des méthodes permettant de valider si ces lois correspondent à la réalité. Nous proposerons alors différents tests afin de valider les ajustements effectués.

Dans une deuxième partie, avant de mettre en pratique les techniques présentées précédemment, nous vérifierons que les données à étudier sont de bonnes qualités et nous effectuerons des statistiques descriptives sur celles-ci. Puis nous passerons à l'estimation de la loi de mortalité des assurés autonomes, leur loi d'entrée en dépendance et la loi de mortalité des dépendants.

Enfin, dans une troisième partie, nous présenterons brièvement le fonctionnement de la réassurance et la construction du compte de résultat permettant de visualiser chaque année si l'entreprise s'appauvrit ou s'enrichit. Nous introduirons le fonctionnement des simulations utilisées, notamment la méthode de Monte Carlo. Nous présenterons ensuite le résultat des projections obtenues par la méthode précitée. Nous essayerons ainsi de prédire pour chaque année une estimation du résultat du réassureur, en simulant divers éléments, et sur un horizon suffisamment long pour visualiser d'éventuelles pertes.

Table des matières

I Introduction **2**

II Aspects théoriques **13**

1 Le cadre **14**

1.1 Perspectives démographiques 14

1.2 Évaluation de l'état de dépendance 20

1.2.1 Définition 20

1.2.2 Indicateur de Katz 20

1.2.3 Grille AGGIR (Autonomie Gérontologie Groupe Iso-Ressource) 21

1.2.4 La grille Colvez 22

1.2.5 L'indicateur EHPA 22

1.2.6 Test Blessed 23

1.2.7 Importance du choix de l'indicateur 23

1.3 Enjeux économiques 24

1.3.1 Prise en charge par l'Etat 24

1.3.2 Réformes politiques 24

1.3.3 Évolution et coûts de la dépendance en France 26

1.3.4 Problème de développement 26

2 Tarification et provisionnement de la dépendance **28**

2.1 Notions préliminaires 28

2.2 Evaluation de la prime 30

2.2.1 Hypothèses de tarification 30

2.3 Calcul des provisions 36

2.3.1 Provision mathématique de rente 37

2.3.2 Provision pour Risques Croissants 37

2.4 Valeur de réduction 39

3 Importance de l'estimation des lois d'expérience **40**

3.1 Analyse de survie 40

3.1.1 Notions préliminaires 41

3.1.2 Modèles non paramétriques 44

3.1.3 Hypothèse pour les sous intervalles 45
3.1.4 Censure et troncature 47
3.1.5 Modèles paramétriques 52
3.2 Lissage et interpolation des données brutes 54
3.2.1 Interpolation 54
3.2.2 Lissage non paramétrique 56
3.2.3 Lissage paramétrique 58
3.3 Fermeture des tables 63
3.3.1 Méthode de Coale et Kisker 64
3.3.2 Méthode de Denuit et Goderniaux 65
3.4 Abattement par expérience 66
3.4.1 Approche SMR 66
3.4.2 Méthode semi paramétrique avec utilisation d'un modèle relationnel de type Brass 67
3.5 Validation des estimations 69
3.5.1 Analyse de la qualité de l'ajustement 69
3.5.2 Tests d'hypothèses : différenciation de courbe de survie 71

III Applications : Détermination des lois du portefeuille 74

4 Étude des données 75
4.1 Statistiques générales 75
4.1.1 Vérification de la qualité des données 75
4.1.2 Étude de la population autonome du portefeuille 76
4.2 Statistiques descriptives 83
4.2.1 Statistiques sur les autonomes 83
4.2.2 Statistiques sur l'incidence en dépendance 84
4.2.3 Statistiques sur les dépendants 86

5 Estimation des probabilités 90
5.1 Estimation des probabilités de décès pour les autonomes 90
5.1.1 Taux brut et lissage 91
5.1.2 Méthode SMR 92
5.1.3 Modèle relationnel de Brass 94
5.1.4 Comparaison des deux méthodes 98
5.2 Estimation de la probabilité de devenir dépendant 103
5.2.1 Distinction Homme Femme 104
5.2.2 Estimation des taux d'incidence bruts 105
5.2.3 Lissage avec la méthode de Whittaker Henderson 105
5.2.4 Validation de l'ajustement 106
5.2.5 Autre idée de modélisation 107

5.2.6 Passage de dépendance partielle à dépendance totale 121
5.2.7 Entrée en dépendance partielle ou dépendance totale 122
5.2.8 Entrée en dépendance pour cause de maladie, démence ou accident 122
5.3 Probabilités de survie en dépendance 123
5.3.1 Justification de la segmentation par âge 124
5.3.2 Justification de la concaténation des deux bases 127
5.3.3 Dépendance totale 127
5.3.4 Dépendance partielle 135
5.3.5 Application des arbres de régression 140

IV Applications : Simulation du compte de résultat 142

6 Aspects comptables 144
6.1 Préliminaires : fonctionnement de la réassurance 144
6.1.1 Principes généraux et histoire 144
6.1.2 Fonctionnement d'une compagnie de réassurance 145
6.1.3 Utilités de la réassurance 145
6.2 Caractéristiques d'un traité de réassurance 146
6.2.1 La réassurance proportionnelle 146
6.3 La réassurance non proportionnelle 147
6.3.1 Traités obligatoires et traités facultatifs 148
6.4 Description du compte de résultat du réassureur 149
6.5 Description des éléments simulés 150

7 Utilisation des simulations 152
7.1 Théorie des simulations de Monte Carlo 152
7.2 Application : simulation du compte de résultat 153
7.2.1 Intérêt de la simulation de Monte Carlo 153
7.2.2 Réalisation 153
7.3 Validation des résultats 157
7.4 Limites du simulateur intégré dans VBA 158

8 Résultats des simulations 159
8.0.1 Simulation des différents éléments 159
8.0.2 Illustration de la convergence 162
8.0.3 Présentation des différents éléments simulés avec l'estimation centrale 165
8.0.4 Simulation des différents hypothèses concernant les autonomes 169
8.0.5 Simulation des différents hypothèses concernant l'incidence . . 170
8.0.6 Simulation des différents hypothèses concernant les dépendants 170

8.0.7 Simulation de différentes hypothèses de taux de rendement . . 171
8.0.8 Solution préventive de la perte future 172
8.0.9 Calcul de la VaR et TVaR 173
8.0.10 Conclusion des simulations 175

9 Conclusion **176**

10 Table des abbréviations **178**

11 Bibliographie **179**

12 Annexes **182**

Liste des illustrations

1.1 Pyramide des âges de la population française en 2013 15
1.2 Pourcentage de femmes dans la population en 2013 par âge 16
1.3 Évolution de l'espérance de vie à travers le temps 17
1.4 Évolution du taux de fécondité conjoncturel à travers le temps 18
1.5 Évolution du taux de fécondité à travers le temps 18
1.6 Évolution du nombre de centenaires 19

2.1 Illustration de l'application des différentes probabilités dans l'année pour un individu d'âge x 32
2.2 Segmentation de la population 35
2.3 Tarification d'une garantie Dépendance partielle et totale 36
2.4 Engagement probable de l'assureur et engagement probable de l'assuré 38

3.1 Illustration de l'observation complète, de la troncature et de la censure avec x l'âge d'un assuré 47
3.2 Problème de la régression sur des probabilités 67

4.1 Pyramide des âges à la souscription 77
4.2 Boîte à moustache de l'âge à la souscription (le losange étant la moyenne) 78
4.3 Évolution de l'âge à la souscription au cours des générations de polices 79
4.4 Durée de vie de la police en fonction de la génération 80
4.5 Boîte à moustache de l'ancienneté des polices 80
4.6 Répartition homme femme des données 81
4.7 Répartition homme femme du portefeuille à travers le temps 81
4.8 Boite à moustache du montant de rente garanti 82
4.9 Évolution du montant de rente garanti par génération 82
4.10 Évolution du montant de rente garantie par âge à la souscription et par sexe 83
4.11 Répartition des décès observés en fonction du sexe 84
4.12 Répartition des décès observés en fonction de l'âge. L'effectif sous risque est en pourcentage de l'effectif total et les décès en pourcentage du total des décès 84
4.13 Répartition Homme Femme des entrées en dépendance 85

4.14 Répartition des entrées en dépendance en fonction de l'âge. L'effectif sous risque est en pourcentage de l'effectif total et l'entrée en dépendance en pourcentage du total des entrées en dépendance 85
4.15 Répartition des entrées en dépendance en fonction de l'âge de la police en année 86
4.16 Répartition des dépendants en fonction du sexe 87
4.17 Durée de dépendance en fonction de l'âge à l'entrée en dépendance . 88
4.18 Répartitions des pathologies en fonction de l'âge à l'entrée 89

5.1 Représentation des taux de mortalité bruts et lissage de Whittaker Henderson 92
5.2 Représentation du rapport nombre de décès observés sur nombre de décès attendus avec la table TH-TF 93
5.3 Représentation des qx bruts et des qx abattus et des qx de la TH-TF 94
5.4 Représentation des logits des taux de de mortalité de la table TH-TF et de ceux de la population autonome 95
5.5 Ajustement des logits des taux de mortalité de la TH-TF sur les logits des taux de mortalité observés 96
5.6 Ajustement des logits des taux de mortalité de la TH-TF sur les logits des taux de mortalité observés 97
5.7 Représentation taux de mortalité observés, des taux de mortalité de la TH-TF et des taux ajustés par le modèle relationnel 98
5.8 Comparaison des taux de mortalité abattus par le SMR et des taux de mortalités ajustés par le modèle relationnel 99
5.9 Comparaison du nombre de décès attendus avec les taux de mortalité abattus par le SMR et ceux attendus avec les taux de mortalité ajustés par le modèle relationnel et le nombre de décès réels 100
5.10 Représentation des résidus pour les deux modèles 101
5.11 Extrapolation par Coale et Kisker, les qx de la TH-TF, les qx de la TH-TF abattus par le SMR et les qx de la TH-TF abattus par le modèle relationnel 102
5.12 Estimation de l'incidence par Kaplan Meier pour les Femmes et pour les Hommes 104
5.13 Estimation de l'incidence par Kaplan Meier 105
5.14 Représentation des taux d'incidence bruts et lissés par la méthode Whittaker Henderson 106
5.15 Représentation du nombre de dépendants estimés par les taux bruts et lissés par la méthode de Whittaker Henderson 107
5.16 Abattement des logits de la TH 109
5.17 Abattement de la TH, taux bruts d'incidence et taux de la TH 110

5.18 Abattement de la TH et lissage par Whittaker, taux bruts, taux de la TH . . . 111
5.19 Analyse des résidus . . . 112
5.20 Comparaison des deux ajustements, avec le logit des taux bruts, le logit des taux obtenus avec le premier ajustement et le logit des taux du second ajustement . . . 114
5.21 Comparaison des deux ajustements, avec les taux bruts, les taux obtenus avec le premier ajustement et les taux du second ajustement . . . 115
5.22 Analyse des résidus de la deuxième modélisation . . . 116
5.23 Taux d'incidence ajustés et taux d'incidence simulés (avec intervalle de confiance à 95% et 5%) . . . 118
5.24 Taux d'incidence ajustés et taux d'incidence simulés (avec intervalle de confiance à 95% et 5%) et taux bruts . . . 119
5.25 Espérance de survie partielle sans dépendance dans le modèle 1 et dans le modèle 2 . . . 120
5.26 Illustration de la possible dérive de l'incidence . . . 121
5.27 Estimation de la fonction de survie pour les différentes classes d'âge . 125
5.28 Estimation de la fonction de survie pour les différentes classes d'âge . 126
5.29 Estimation des taux bruts . . . 128
5.30 Interpolation et lissage des taux bruts . . . 129
5.31 Adéquation du nombre de sorties estimé par les taux lissés par rapport à celui estimé avec les taux bruts . . . 130
5.32 Résidus entre l'estimation lissée et l'estimation brute pour les 2 classes d'âge . . . 131
5.33 Loi de décès des dépendants en fonction de l'âge à l'entrée obtenue avec la première méthode . . . 133
5.34 Loi de décès des dépendants en fonction de l'âge à l'entrée obtenue avec la deuxième méthode . . . 135
5.35 Estimation de la fonction de survie . . . 136
5.36 Estimation des taux bruts . . . 137
5.37 Interpolation et lissage des taux bruts . . . 138
5.38 Résidus entre l'estimation lissée et l'estimation brute . . . 139
5.39 Arbre de régression . . . 140

6.1 Illustration de la cession du risque d'une compagnie d'assurance à une compagnie de réassurance . . . 145

7.1 Schema de la simulation de Monte Carlo . . . 156

8.1 Exemple de 1000 simulations des prestations pour l'année 2020 . . . 162
8.2 Exemple de 1000 simulations des cotisations pour l'année 2020 . . . 162
8.3 Convergence de la variance . . . 163

8.4 Comparaison des simulations du résultat pour l'estimation centrale et l'estimation Bis de la mortalité des dépendants 164
8.5 Exemple de 100 trajectoires des cotisations sur 25 ans 165
8.6 Exemple de 100 trajectoires des PRC sur 25 ans 166
8.7 Exemple de 100 trajectoires des prestations sur 25 ans 167
8.8 Exemple de 100 trajectoires des PMR sur 25 ans 168
8.9 Trajectoire du résultat en fonction des qax 169
8.10 Trajectoire du résultat en fonction des ix 170
8.11 Trajectoire du résultat en fonction des qix 171
8.12 Trajectoire du résultat en fonction du taux de rendement des actifs financiers 171
8.13 Résultat technique simulé avec fond de stabilité 173
8.14 VaR, TVaR et moyenne 174

Section II

Aspects théoriques

Partie 1

Le cadre

Dans cette première partie, nous montrerons le risque d'accentuation du vieillissement de la population. Nous verrons alors que ce vieillissement pourrait s'accompagner de perte d'autonomie. Nous définirons alors ce qu'est la dépendance, et nous verrons comment l'Etat tente de venir en aide aux personnes en souffrant.

1.1 Perspectives démographiques

Vieillissement de la population

Alors que la proportion de personnes âgées de plus 60 ans était de 22% en 2004, elle s'élève en 2013 à 24%, et devrait atteindre 35% en 2040 [1]. Cette progression s'explique par plusieurs phénomènes. D'une part, l'augmentation de l'espérance de vie, qui ne cesse de progresser en raison des évolutions médicales et sanitaires et qui permet au plus grand nombre d'atteindre un âge élevé. D'autre part, cette tendance s'explique par le vieillissement de la génération du baby boom (cf. Illustration 1.1), c'est à dire les personnes nées entre les années 1940 et 1970 dans les pays d'Europe du Nord, qui arriveront à l'âge de 80 ans à partir de 2020, et par des taux de natalité insuffisamment élevés.

1. source INED

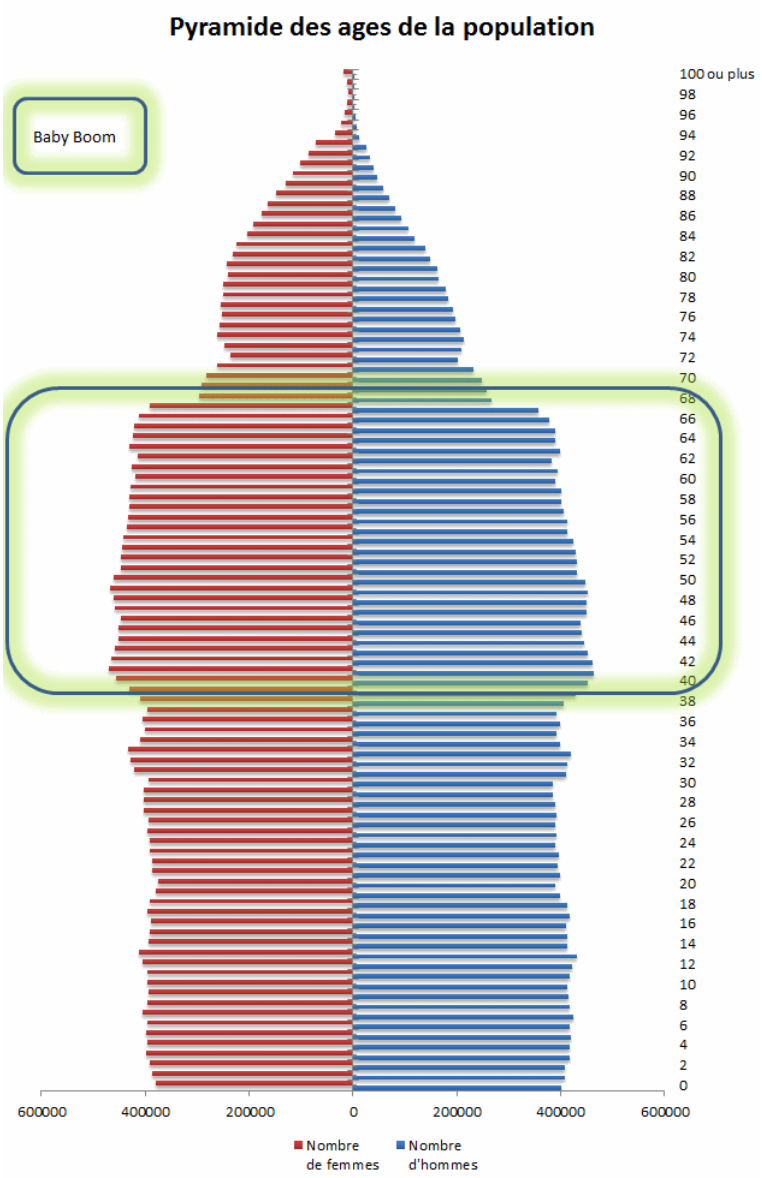

Illustration 1.1 : *Pyramide des âges de la population française en 2013*

Hausse de l'espérance de vie

Par ailleurs, le gain en espérance de vie à la naissance est considérable, passant d'environ 60 ans en 1946 à presque 79 en 2013 pour les hommes, soit une augmentation de 24% sur une période de 61 ans. L'évolution pour les femmes est similaire, avec une espérance de 65 ans en 1946, pour arriver à 85 ans en 2013 (cf. Illustration 1.3). Cette espérance de vie plus élevée des femmes entraîne une sur-représentativité de ces dernières dans la population aux âges élevés. Ainsi, 56% de la population âgée de plus de 60 ans est constituée de femmes, et pour la population âgée de plus de 80 ans, la proportion de personnes de sexe féminin atteint les 65% (cf. Illustration 1.2).

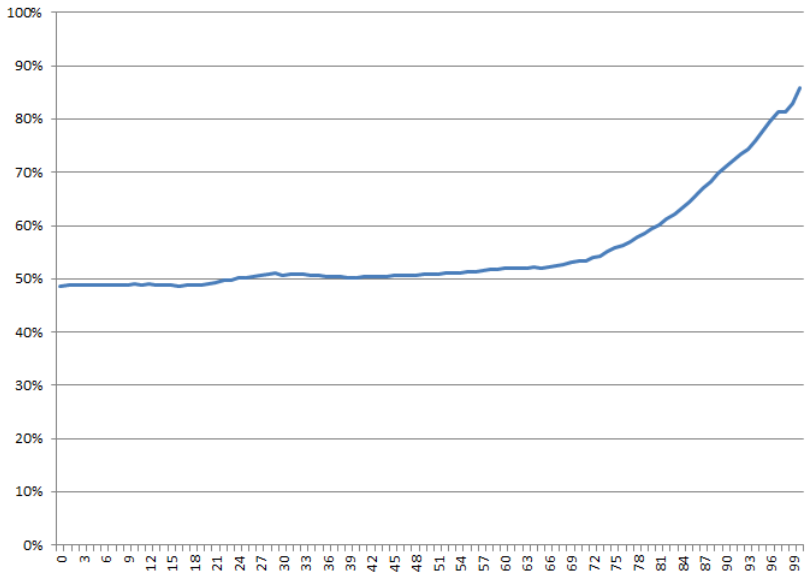

Illustration 1.2 : *Pourcentage de femmes dans la population en 2013 par âge*

Le constat est le même pour l'espérance de vie après 60 ans. Ainsi, nous estimions qu'un homme âgé de 60 ans en 1946 allait survivre en moyenne 15 ans, en 2013, nous estimons qu'une même personne a une espérance de survie de presque 23 ans, soit une évolution de 32%. L'évolution pour les femmes est encore plus forte, de l'ordre de 34% (27 ans en 2013 contre 18 en 1946).

Remarquons ainsi que l'espérance de survie des personnes âgées a davantage augmenté que l'espérance de survie à la naissance durant cette période. Notons également que l'espérance de vie à la naissance des femmes tend à marquer une inflexion et être rattrapée par l'espérance de vie des hommes, au vu de l'évolution de ces espérances à travers le temps (cf. Illustration 1.3).

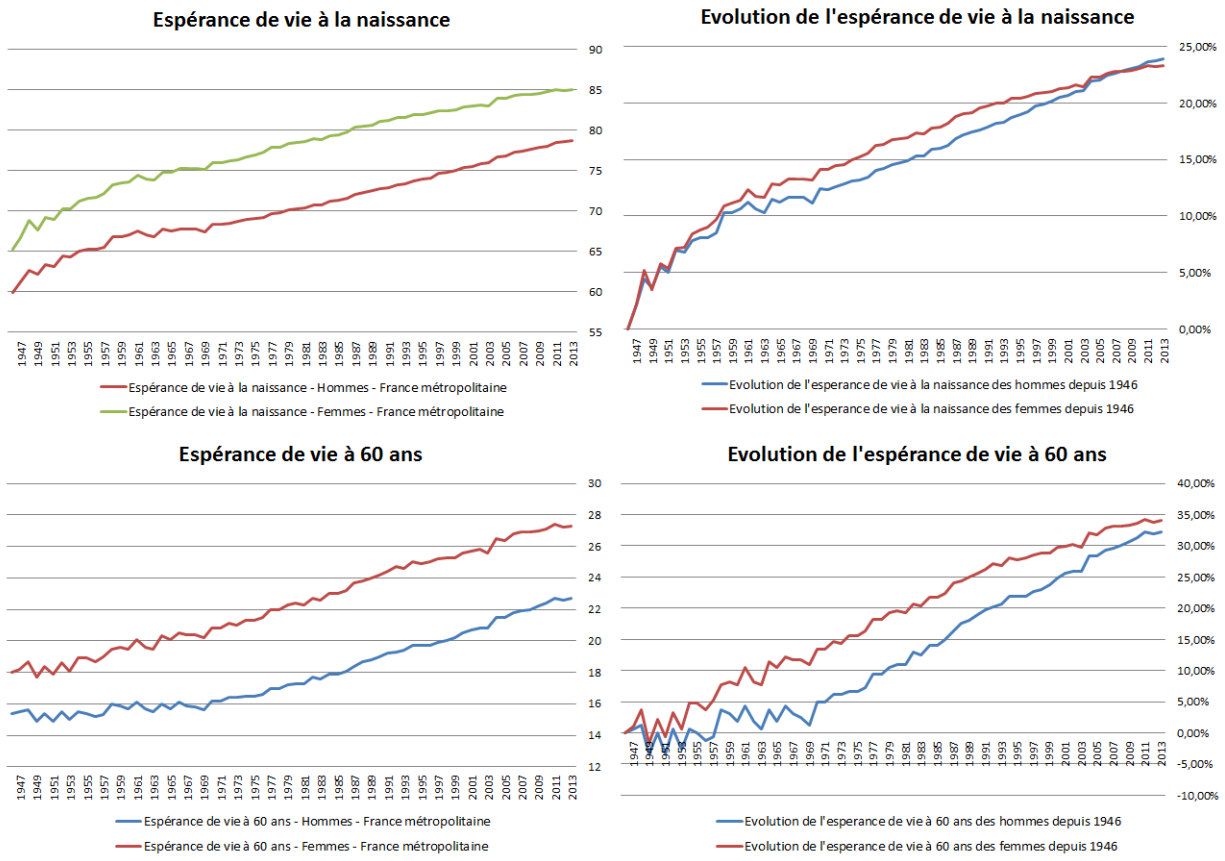

Illustration 1.3 : *Évolution de l'espérance de vie à travers le temps*

Une natalité insuffisante

La natalité, en hausse depuis un certain nombre d'années, mais insuffisante, entraîne une diminution du renouvellement de la population. Le "baby krach", comme certain le surnomme en référence au baby boom, fait référence à la natalité devenue relativement faible en Europe après les années 1970. L'indicateur conjoncturel de fécondité [2], est trop faible pour assurer le remplacement de la population. Ce dernier devrait s'élever, compte tenu de la mortalité en France, à 2.07. Or il atteindrait, selon des chiffres provisoires de l'INSEE [3], seulement 1.97 en 2013 (cf. Illustration 1.4). En légère hausse depuis les années 1990, il reste à un niveau bien inférieur que dans le passé (à titre d'exemple, il était de 3 en 1947, en plein baby boom). Malgré tout, ce dernier reste l'un des plus élevés en Europe.

2. Calcul du nombre moyen d'enfants auxquels les mères donneraient le jour si les générations futures avaient le même taux de fécondité par âge que les générations actuelles
3. Institut Nationale de la Statistique et des Etudes Economiques

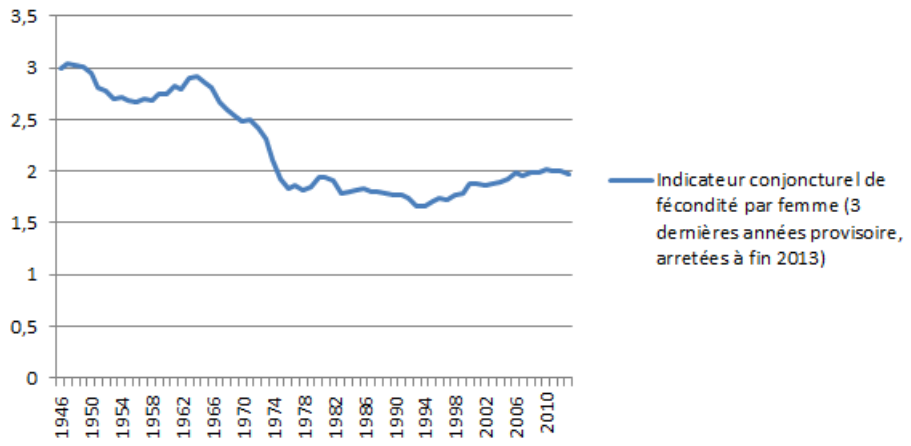

Illustration 1.4 : *Évolution du taux de fécondité conjoncturel à travers le temps*

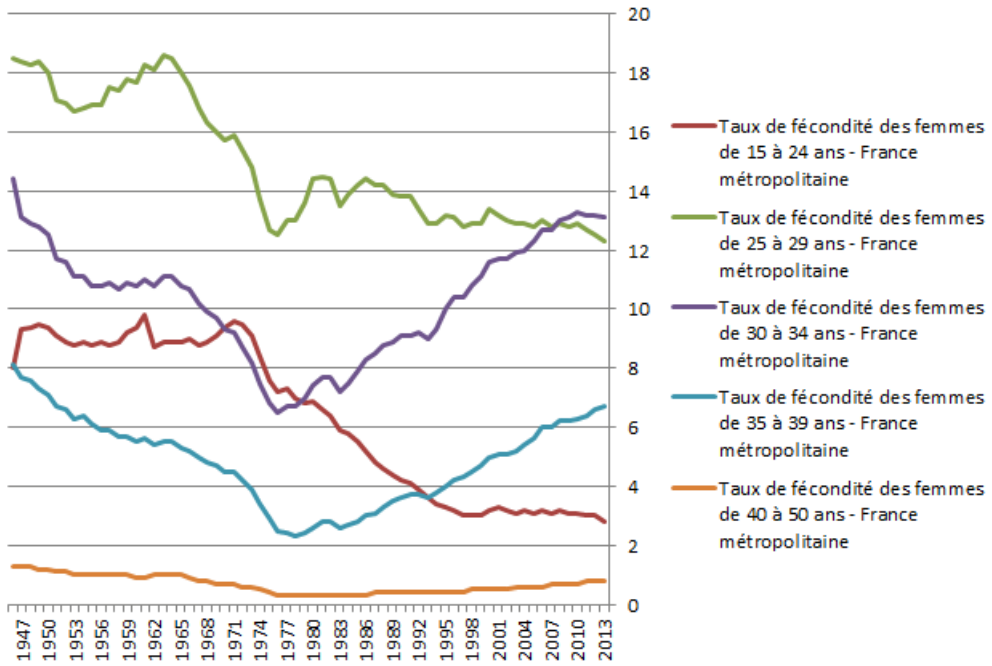

Illustration 1.5 : *Évolution du taux de fécondité à travers le temps*

Le taux de fécondité étant le rapport entre nombre de naissances vivantes de l'année et la population totale moyenne de l'année.

De plus en plus de centenaires

Le nombre de centenaires ne cesse d'augmenter. Passant de 3 760 en 1990 en France, à 14 941 en 2009, soit une hausse de 75% en 19 ans. L'INSEE projette qu'en 2060, le nombre de centenaires pourrait atteindre 198 645, selon un scénario central. Le scénario optimiste, anticipant une espérance de vie en hausse, estime ce nombre à 379 051 (cf. Illustration 1.6).

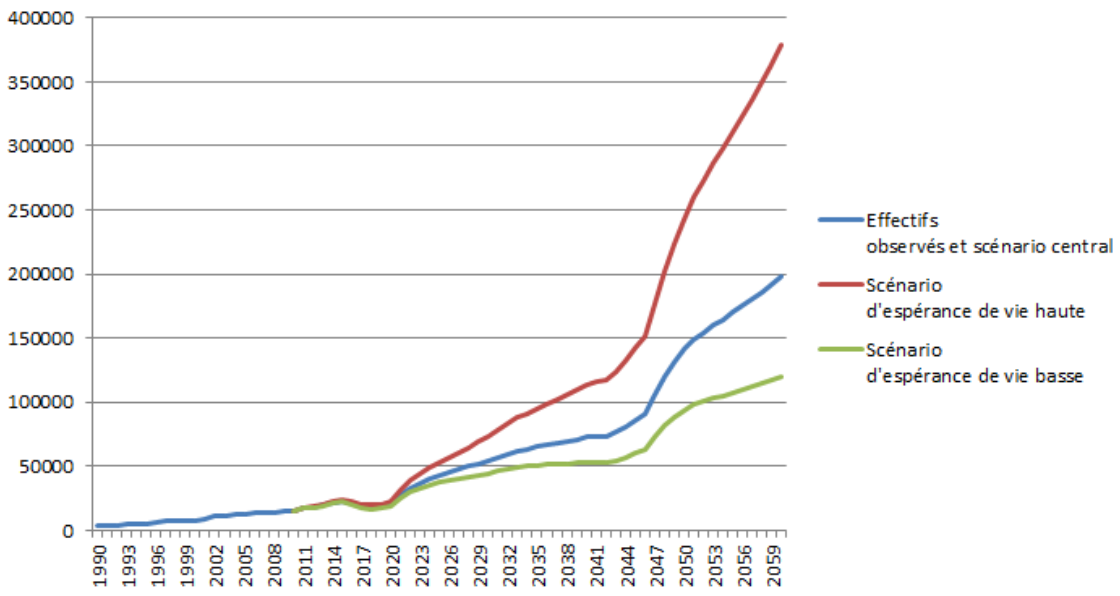

Illustration 1.6 : *Évolution du nombre de centenaires*

Vivre plus longtemps, mais dans quelles conditions ?

La question qui se pose est de savoir si cette population qui vit de plus en plus longtemps va rester en bonne santé, ou si le vieillissement de la population va s'accompagner de sénilité. Plusieurs scénarios sont envisagés :

- Allongement de l'espérance de vie sans allongement de l'espérance de vie sans incapacité (extension de la morbidité) [4] ;
- Allongement de l'espérance de vie avec allongement de l'espérance de vie sans incapacité (maintien de la morbidité) [5] ;
- Age moyen d'apparition des maladies générant des états de dépendance davantage retardé que l'âge de décès (compression de la morbidité) [6].

Cette introduction a pour but de montrer dans quelles mesures le vieillissement de la population est une problématique d'actualité, plus importante chaque jour, notamment lorsque l'allongement de l'espérance de vie est accompagné d'une augmentation du nombre de dépendants. Nous nous attacherons dans la partie suivante à apporter une définition plus complète de la dépendance et nous présenterons les méthodes permettant de la mesurer de la manière la plus objective possible.

4. Gruenberg et Kramer (1980)
5. Fries (1980)
6. Manton (1992)

1.2 Évaluation de l'état de dépendance

1.2.1 Définition

La dépendance est « l'état de la personne qui, malgré les soins qu'elle est susceptible de recevoir, a besoin d'être aidée pour l'accomplissement des actes essentiels de la vie et requiert une surveillance régulière ». [7]

Wood (1980) définit une personne dépendante comme "une personne qui en raison d'un déficit anatomique, ou d'un trouble physiologique, ne peut remplir des fonctions, effectuer des gestes essentiels à la vie quotidienne, sans le concours d'autres personnes...". Wood définit également un schéma menant à la dépendance, dans le but de prévenir celle-ci. Pour lui, la dépendance provient d'un handicap, le handicap découle d'une incapacité, l'incapacité est provoquée par une déficience, elle même déclenchée par une maladie.

Les déficiences peuvent être intellectuelles, psychiques, auditives, visuelles ou motrices.

Afin de pouvoir répondre de façon optimale aux besoins d'une personne en état de dépendance, il convient de mesurer celle-ci de manière objective grâce à des critères précis. Pour ce faire, un ensemble de méthodes permettant d'évaluer l'état de dépendance des individus a été mis en place. Ces méthodes sont utilisées non seulement par l'Etat pour le versement des aides aux personnes dépendantes, mais aussi par les organismes assureurs, dans le but d'évaluer s'il y a survenance du risque ou non. Le choix de ces méthodes est déterminant dans le sens où certaines d'entre elles peuvent se révéler plus souple dans l'évaluation de l'état de dépendance de la personne et ainsi surévaluer le nombre de dépendants, ce qui entraînerait des coûts plus élevés.

1.2.2 Indicateur de Katz

Cet indicateur, mis au point en 1969 par Lawton et Brody, est basé sur un questionnaire visant à évaluer la capacité du sujet à accomplir les tâches utilitaires quotidiennes. Il évalue la dépendance d'une personne en fonction des six activités de la vie quotidienne définies telles que :

- faire sa toilette ;
- s'habiller ;
- aller aux toilettes et les utiliser ;
- se coucher ou quitter son lit et s'asseoir ou quitter son siège ;
- contrôler ses selles et ses urines ;
- manger des aliments déjà préparés (cuisinés ou coupés).

7. Définition retenue par la loi du 24 janvier 1997

L'indicateur de Katz est une synthèse des besoins d'aide vers l'indépendance pour la réalisation de ces six activités de la vie quotidienne.

1.2.3 Grille AGGIR (Autonomie Gérontologie Groupe Iso-Ressource)

La grille nationale AGGIR (Autonomie Gérontologie Groupes Iso-Ressources) constitue un outil destiné à évaluer le degré de perte d'autonomie ou le degré de dépendance, physique et psychique, des demandeurs de l'Allocation Personnalisée d'Autonomie (APA) dans l'accomplissement de leurs actes quotidiens. L'évaluation se fait sur la base de dix-sept variables :

- dix variables dites "discriminantes" se rapportent à la perte d'autonomie physique et psychique et sont utilisées pour le calcul du GIR (Groupe Iso-Ressources) : cohérence - orientation - toilette - habillage - alimentation - élimination - transferts (se lever, se coucher, s'asseoir) - déplacement à l'intérieur - déplacement à l'extérieur - communication à distance ;
- sept variables dites "illustratives", concernant la perte d'autonomie domestique et sociale, n'entrent pas dans le calcul du GIR mais apportent des informations utiles à l'élaboration du plan d'aide : gestion personnelle de son budget et de ses biens - cuisine - ménage - transports - achats - suivi du traitement - activités de temps libre. A partir des réponses au questionnaire, les individus sont classés en 6 groupes GIR (Groupe d'Iso Ressources)

Les Groupes Iso Ressources (GIR), au nombre de six, sont constitués à partir des réponses à la grille AGGIR qui est un outil de mesure de l'autonomie à travers l'observation des activités qu'effectue seule la personne âgée :

- Le GIR 1 comprend les personnes confinées au lit ou au fauteuil et ayant perdu leur autonomie mentale, corporelle, locomotrice et sociale, qui nécessitent une présence indispensable et continue des intervenants.
- Le GIR 2 est composé de deux sous-groupes : d'une part, les personnes confinées au lit ou au fauteuil dont les fonctions mentales ne sont pas totalement altérées et qui nécessitent une prise en charge pour la plupart des activités de la vie courante ; d'autre part, celles dont les fonctions mentales sont altérées et qui ont conservé leurs capacités motrices. Le déplacement à l'intérieur est possible mais la toilette et l'habillage ne sont pas faits ou sont faits partiellement.
- Le GIR 3 regroupe les personnes ayant conservé leur autonomie mentale, partiellement leur autonomie locomotrice, mais qui nécessitent quotidiennement et plusieurs fois par jour des aides pour leur autonomie corporelle. Ainsi, la toilette et l'habillage ne sont pas faits partiellement. De plus, l'hygiène de l'élimination nécessite l'aide d'une tierce personne.
- Le GIR 4 comprend les personnes qui n'assument pas seules leur transfert

mais qui, une fois levées, peuvent se déplacer à l'intérieur du logement. Elles doivent être aidées pour la toilette et l'habillage. La plupart s'alimentent seules ; ce groupe comprend aussi celles qui n'ont pas de problèmes locomoteurs mais qu'il faut aider pour leurs activités corporelles et leur repas.

- Le GIR 5 est composé des personnes qui assurent seules leurs déplacements à l'intérieur du logement, s'alimentent et s'habillent seules. Elles peuvent avoir besoin une aide ponctuelle pour la toilette, la préparation des repas et le ménage.
- Le GIR 6 est composé des personnes qui n'ont pas perdu leur autonomie pour les actes discriminants de la vie quotidienne.

1.2.4 La grille Colvez

Cette grille d'appréhension de la dépendance (au sens du besoin d'aide) mesure la perte de mobilité. Elle classe les personnes en quatre groupes :

- niveau 1 : personnes confinées au lit ou au fauteuil (dépendance lourde) ;
- niveau 2 : personnes non confinées au lit ou au fauteuil mais ayant besoin d'aide pour la toilette et l'habillage (dépendance légère) ;
- niveau 3 : personnes ayant besoin d'aide pour sortir de leur domicile ou de l'institution où elles sont hébergées, mais n'appartenant pas aux niveaux 1 et 2 ;
- niveau 4 : autres personnes (considérées comme dépendantes).

Simple et relativement ancienne, la grille Colvez a permis de fournir des estimations sur la population âgée dépendante en France depuis le début des années 1990. Cependant la grille est limitée à la mesure de la dépendance physique et omet donc de déterminer la dépendance psychique.

1.2.5 L'indicateur EHPA

Utilisé la première fois en 1990, l'indicateur EHPA [8] permet d'appréhender à la fois la dépendance physique et la dépendance psychique. En effet, il croise les quatre groupes de la grille Colvez, qui mesure le besoin d'aide lié à la perte de mobilité, avec deux groupes de dépendance psychique constitués en fonction de l'importance de l'aide nécessitée par les problèmes d'orientation dans le temps et dans l'espace et les troubles du comportement de la personne. Nous définissons huit groupes EHPA :

- EHPA 11 : dépendance psychique et confinée au lit ou au fauteuil ;
- EHPA 12 : dépendance psychique et besoin d'aide pour la toilette et l'habillage ;
- EHPA 13 : dépendance psychique et besoin d'aide pour sortir du domicile ou de l'institution ;

8. Etablissement d'Hébergement pour Personnes Agées

- EHPA 14 : dépendance psychique et pas de dépendance physique ;
- EHPA 21 : peu ou pas de dépendance psychique et confiné au lit ou au fauteuil ;
- EHPA 22 : peu ou pas de dépendance psychique et besoin d'aide pour la toilette et l'habillage ;
- EHPA 23 : peu ou pas de dépendance psychique et besoin d'aide pour sortir du domicile ou de l'institution ;
- EHPA 24 : peu ou pas de dépendance psychique et pas de dépendance physique.

Les groupes EHPA 11, 12, 21 et 22 correspondent ainsi aux deux premiers niveaux de la grille Colvez (dépendance physique lourde).

Comme nous l'avons vu, la dépendance des individus peut ne pas être simplement physique. Elle peut aussi être psychique. C'est pourquoi des tests spécifiques à ce type de dépendance ont été mis en place. Décrivons succinctement un de ces tests, le test blessed.

1.2.6 Test Blessed

Le test Blessed mis en place par l'auteur du même nom en 1968 est un outil utilisé pour évaluer l'état de dépendance mentale d'un individu. Il doit être réalisé par un neurologue ou un psychiatre, par interrogation de la personne dépendante et de son entourage.

1.2.7 Importance du choix de l'indicateur

Selon les estimations de Colin et Coutton (2000) réalisées avec l'enquête HID, le nombre de personnes dépendantes serait, pour cette enquête et en fonction de l'indicateur :

- 1,4 million de personnes avec la grille Colvez (niveaux 1 à 3) ;
- 800 000 selon l'outil Aggir (GIR 1 à 4) ;
- 700 000 selon l'indicateur Katz (besoin d'aide pour au moins deux des 6 activités de la vie quotidienne retenues).

Au 1er janvier 2012, la DREES estime à 1,172 million le nombre de personnes âgées dépendantes au sens de l'APA (GIR 1 à 4, avec 111 256 personnes en GIR 1, 308 628 en GIR 2, 243 572 en GIR 3 et 508 544 en GIR 4).

L'Etat a pris conscience de ce risque grandissant de perte d'autonomie à grande échelle depuis un certain nombres d'années. Des conséquences néfastes d'un point de vue social et économique pourraient en résulter si aucune mesure n'était prise.

1.3 Enjeux économiques

1.3.1 Prise en charge par l'Etat

En France, les personnes dépendantes de plus de 60 ans reçoivent l'Aide Personnalisée à l'Autonomie (APA). L'outil AGGIR (Autonomie Gérontologique Groupe Iso Ressources) est retenu depuis la loi du 24 janvier 1997 comme grille nationale d'évaluation de la perte d'autonomie pour l'attribution de la Prestation Spécifique Dépendance (PSD) de 1997 à 2001 et de l'allocation personnalisée d'autonomie (APA) mise en place depuis 2002. Les personnes âgées qui sollicitent le bénéfice de l'allocation personnalisée d'autonomie (APA), sont classées dans les six groupes iso-ressources que compte la grille nationale, en fonction des aides à la personne nécessaires à leur survie. Seuls les quatre premiers GIR de la grille nationale ouvrent droit à l'APA, que les bénéficiaires se trouvent à domicile ou en établissement, à condition qu'ils répondent aux critères d'âge et de résidence. Les personnes âgées classées en GIR 5 et 6 peuvent néanmoins prétendre au versement des prestations d'aide ménagère servies par leur régime de retraite ou par l'aide sociale départementale. Cette aide prend ainsi en compte le niveau de dépendance de la personnes et ses revenus. Elle est affectée et plafonnée en fonction du GIR dans lequel se trouve la personne. Soit pour 2014, les différents plafonds :

- 1 312,67 €/mois pour le GIR 1 ;
- 1 125,14 €/mois en GIR 2 ;
- 843,86 €/mois en GIR 3 ;
- 562,57 €/mois en GIR 4.

Comme nous l'avons vu, l'APA est versée aussi bien à une personne restant à son domicile, qu'à une personne résidant en établissement, mais le montant versé diffère. Le montant moyen mensuel de l'APA à domicile était de 406 Euros et, en établissement, de 307 Euros au 30 juin 2009.

Toutefois cette aide n'est pas suffisante pour aider les personnes dépendantes.

Le coût moyen d'un EHPAD [9] est estimé à 2892 euros par mois [10] pour l'année 2012. Ainsi une partie importante reste à charge de la personne dépendante puisque ce dernier montant est bien supérieur aux plafonds cités précédemment. Signalons aussi que la durée de survie moyenne en dépendance est estimée à 4 ans (3,7 ans pour les hommes et 4,7 ans pour les femmes) [11].

1.3.2 Réformes politiques

Après l'annonce en 2007 de la grande réforme de la dépendance par le gouvernement Fillon, un texte de loi sur la perte d'autonomie a été adopté le 3 juin 2014 au

9. Etablissement d'Hébergement pour Personnes Agées Dépendantes
10. CNSA (Caisse Nationale de Solidarité pour l'Autonomie)
11. INSEE

Conseil des Ministres. Il doit entrer en vigueur en 2015. Ce texte de loi s'articule, selon Marisol Touraine, alors Ministre des affaires sociales et Mme Rossignol, en charge des personnes âgées et de l'autonomie, autour de trois piliers : "anticipation de la perte d'autonomie, adaptation de la société, accompagnement des personnes en perte d'autonomie". Pour ce projet de loi, 700 millions d'Euros ont été alloués.

Ainsi, premièrement les pouvoirs publiques vont tenter de lutter contre l'isolement des personnes âgées grâce notamment aux nouvelles technologies.

Deuxièmement, le nombre d'heures d'accompagnement allouées aux bénéficiaires de l'APA devrait être augmenté.

Troisièmement, un soutien devrait être apporté aux aidants, lorsqu'il s'agit des proches de la personne, se manifestant par une prise en charge temporaire de la personne dépendante. Par ailleurs, les métiers d'aide à la personne en cas de perte d'autonomie vont être valorisés. Ce point est important pour l'économie du pays et pour le bien-être social. En effet, ce type d'emploi permet d'une part de libérer les aidants, limitant ainsi les impacts sur leur vie professionnelle, et d'autre part de créer des emplois. Ces deux derniers effets permettent de lutter contre le chômage et d'améliorer la productivité.

Quant à l'adaptation de la société, elle comprend la création d'un volontariat civique senior tel que l'association des grands-parents à la gestion des crèches de leurs petits-enfants, et l'adaptation de 80 000 logements aux besoins d'une personne en perte d'autonomie.

Toutefois, cette réforme ne prend pas suffisamment en compte la problématique du financement de la partie à charge de la personne lors d'un hébergement en établissement spécialisé, puisqu'il est annoncé que la part à charge des personnes dépendantes sera diminuée par une revalorisation de l'APA, voire même supprimée pour les bénéficiaires du minimum vieillesse.

Le texte de loi prévoit aussi un encadrement plus important des assurances dépendances, qui devront répondre à un cahier des charges précis pour que leurs produits continuent de bénéficier d'abattements fiscaux (encadrement notamment de la définition de la dépendance).

C'est pourtant ce dernier point qui est central, le risque de fonte du patrimoine de la personne dépendante est grand et les enfants ont une "obligation alimentaire" envers leurs ascendants, qui comprend la participation aux soins médicaux.

Dans ce contexte, l'assurance dépendance privée représente un enjeu majeur dans la prise en charge du risque de dépendance et la prévention de risque économique lié au vieillissement de la population. Non seulement ce type d'assurance permet de participer au financement de la prise en charge par un établissement spécialisé, mais peut aussi permettre, si la personne le souhaite, d'aménager son domicile afin qu'elle y demeure. Cette couverture peut aussi permettre de rémunérer un personnel spécialisé dans l'aide à la personne, libérant ainsi les aidants familiaux.

L'effet économique pour la nation ne s'arrête pas là, la prévision du risque de dépendance par un assureur entraîne la mise en provision de capitaux importants. Ces capitaux pourraient permettre d'aider le financement de l'économie si ce type d'assurance devenait obligatoire. A titre d'indication, les sociétés d'assurances ont mis en réserve 3,1 Milliards d'Euros [12] pour faire face aux prestations futures à la fin de l'année 2011.

1.3.3 Évolution et coûts de la dépendance en France

Il y avait 1,2 million de personnes âgées dépendantes en France début 2012 [13]. Elles devraient être 2,3 millions en 2060 selon les projections réalisées par la DREES [14]. Actuellement évaluer à environ 1.05 point de PIB en France, soit 21,1 milliards d'Euros, le financement de la dépendance pourrait augmenter jusqu'à 1.77 points d'ici à 2060 (avec prise en compte des dépenses de santé, médico sociale et d'hébergement).

1.3.4 Problème de développement

Malgré les avantages qu'elle présente, cette garantie ne rencontre pas le succès escompté. Seuls 5,5 millions de contrats dépendance sont souscrits en France à l'heure actuelle. Les individus pourtant risquophobes ont du mal à considérer ce risque. A contrario par exemple d'un accident de voiture, ou des frais dentaires, qui sont des risques facilement envisageables à tout instant par chacun, les effets de la dépendance sont difficilement imaginés par des individus encore en bonne santé. Et surtout, ils n'ont pas envie de l'envisager, puisque l'état de dépendance est un état irréversible. Ce manque de prévoyance des individus entraîne des souscriptions très tardives, avec un âge moyen à l'adhésion de 62 ans. Les portefeuilles dépendance sont ainsi généralement relativement âgés. Cela a pour conséquence de rendre la souscription plus chère et de limiter la mutualisation que l'assureur pourrait effectuer entre personnes jeunes et personnes âgées.

Cette partie nous a permis de mieux comprendre comment est mesuré l'état de dépendance d'un individu. Nous avons vu que l'Etat vient en aide aux personnes dépendantes, avec certaines subventions, représentant un coût important pour la société, et dont l'évolution reste incertaine. Nous allons maintenant voir comment les assureurs viennent en aide à ces personnes, par le biais des couvertures dépendances. Nous commencerons par analyser la tarification d'une garantie dépendance. Pour ce faire, nous décrirons les particularités d'un contrat dépendance, puis nous analyserons la tarification de la garantie et le calcul des provisions.

12. Les contrats d'assurance dépendance en 2011 –FFSA-GEMA Juin 2012
13. Source DREES, tous GIR confondus
14. Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques

Nous présentons la tarification de la garantie et le calcul des provisions dans le but de comprendre la construction du produit et de comprendre d'où peuvent venir les éventuelles problèmes de pertes liés à cette garantie.

Partie 2

Tarification et provisionnement de la dépendance

2.1 Notions préliminaires

Nous considérons ici le cas d'un contrat dépendance individuelle, à titre de garantie principale. En effet, une couverture dépendance peut aussi être jointe en option de couverture santé ou via des produits d'assurance vie. La couverture dépendance peut aussi être de type collectif.

Définition de la dépendance La définition de la dépendance pour les assurances se fait généralement grâce aux AVQ [1] et/ou à la grille GIR ou grâce au AIVQ [2].

Type de prestations possibles La prestation peut être indemnitaire (en fonction des dépenses) ou forfaitaire (l'assuré choisit le montant de rente ou de capital qu'il percevra en cas de survenance du risque). La prestation de type indemnitaire n'est pas répandue en France. Elle se pratique essentiellement aux Etats-Unis (problèmes de dérives des coûts qui s'est produit).

La prestation est généralement sous forme de rente mensuelle, mais il est souvent possible d'y adjoindre des garanties complémentaires : capital en cas de dépendance pour l'aménagement du domicile, assistance pour aidant, frais d'obsèques, capital en cas de fractures...

Paiement de la cotisation La cotisation est le plus souvent calculée en fonction de l'âge à l'adhésion c'est à dire nivelée. Mais il peut aussi s'agir d'une prime de risque annuelle. La cotisation doit être unisexe depuis la Gender Directive [3].

1. Activités de la Vie Quotidienne
2. Activités Instrumentales de la Vie Quotidienne
3. Council Directive 2004/113/EC of 13 December 2004, en application pour les assurances depuis le 21 Décembre 2012

Âge de souscription L'âge de souscription varie en fonction des produits, mais est généralement compris entre 25 et 75 ans.

Présence d'une sélection médicale à la souscription En raison des problèmes d'aléa moral et d'anti-sélection, le plus souvent une sélection médicale est présente à la souscription.

L'aléa moral[4] est le fait qu'un agent se comporte différemment si ce n'est pas lui qui supporte le risque.

L'anti séléction est le fait qu'un assuré se sachant touché par un risque va chercher à souscrire une assurance pour annuler les conséquences financières de ce risque. Dans notre cas, il pourrait s'agir d'une personne au premier stade d'une maladie, qui cherche à se protéger en souscrivant une assurance dépendance.

La sélection médicale a pour but d'éviter les mauvais risques, c'est à dire évaluer si la probabilité d'entrée en dépendance de la personne au moment de la souscription n'est pas plus élevée que celle anticipée par l'assureur. Cette évaluation est souvent effectuée grâce à un questionnaire médical.

Si l'assuré est en bonne santé, la souscription se fait normalement, le cas échéant l'assuré présente un risque supplémentaire. Trois décisions sont alors possibles, l'assuré peut payer une surprime, l'assureur peut décider d'ajouter des exclusions au contrat ou encore la souscription peut être refusée.

L'article L113-4 du Code des Assurances stipule que la sélection médicale s'effectue à l'adhésion du Contrat et pour toute sa durée : " l'assureur ne peut plus se prévaloir de l'aggravation des risques quand, après en avoir été informé de quelque manière que ce soit, il a manifesté son consentement au maintien de l'assurance, spécialement en continuant à recevoir les primes ou en payant, après un sinistre, une indemnité. "[5].

En cas de fausse déclaration lors de cette sélection médicale, selon l'article L113-9 du Code des Assurances, l'assuré s'expose à la nullité de la garantie, si la mauvaise foi est avérée :

"L'omission ou la déclaration inexacte de la part de l'assuré dont la mauvaise foi n'est pas établie n'entraîne pas la nullité de l'assurance.

Si elle est constatée avant tout sinistre, l'assureur a le droit soit de maintenir le contrat, moyennant une augmentation de prime acceptée par l'assuré, soit de résilier le contrat dix jours après notification adressée à l'assuré par lettre recommandée, en restituant la portion de la prime payée pour le temps où l'assurance ne court plus.

Dans le cas où la constatation n'a lieu qu'après un sinistre, l'indemnité est réduite en proportion du taux des primes payées par rapport au taux des primes qui auraient été dues, si les risques avaient été complètement et exactement déclarés."[6]

4. Adam Smith's (1759) Theory of Moral Sentiments
5. Article L113-4 du Code des Assurances
6. Article L113-9 du Code des Assurances

Délai de franchise Il est possible d'inclure un délai de franchise entre la survenance de la dépendance et la prise en charge, ce qui permet de réduire le coût de la garantie pour l'assuré puisqu'une partie des dépendants meurt durant les mois qui suivent leurs entrées en dépendance.

Délai de carence La présence d'un délai de carence différent selon la pathologie permet de réduire l'effet d'anti-sélection.

Généralement, il n'y a pas de délai de carence pour la dépendance survenue suite à un accident, puisque cela ne peut être prévisible par l'assuré, mais il peut y avoir un délai de carence de 2 voire 3 ans pour la dépendance des suites d'une maladie.

Réduction Selon les produits, les contrats peuvent être mis en réduction, c'est à dire que la cessation de paiement des primes ne met pas fin à la couverture, mais le montant de rente est diminué. Toutefois cette réduction n'est possible qu'après avoir cotisé un nombre minimum d'années, généralement 8 ans au moins.

Revalorisation Les montants de rente souscrits peuvent être revalorisés en fonction de certains indices, prévus contractuellement, dans la limite des résultats financiers. En cas de mauvais résultats financiers ou techniques, l'assureur n'est pas toujours tenu de revaloriser les montants de rente. Il peut par contre y être contraint, selon les termes du contrat, même en cas de mauvais résultats.

Rôle de la réassurance La réassurance a tenu et tient encore une place importante dans l'assurance du risque de dépendance. En effet, de par son expérience sur le plan international et son accès à d'importantes bases de données, le réassureur a une plus grande connaissance du risque. Les réassureurs ont ainsi pu conseiller les assureurs lors la tarification de ces garanties nouvelles.

Pour ce type de couverture, la réassurance est généralement sous forme de quote-part, ce qui permet l'absence de dérive de risque et d'aléa moral de la part de l'assureur, puisque son résultat est exactement lié à celui du réassureur.

La réassurance permet aussi à la cédante [7] de céder du risque, de réduire la marge de solvabilité en limitant les besoins en capitaux (notamment les provisions).

2.2 Evaluation de la prime

2.2.1 Hypothèses de tarification

- Dans cette partie, la tarification va être effectuée pour une rente viagère de 1 Euro par mois ;
- Les décès et les entrées en dépendance surviennent en début d'année ;

7. l'entité cédant du risque par le biais de la réassurance

- Aucune rémission n'est possible, lorsque l'assuré entre en dépendance, nous considèrons qu'il y reste (état irréversible) ;
- Lorsque l'assuré entre en dépendance, la cotisation cesse d'être due.

La tarification de la garantie suppose la connaissance d'au minimum trois lois, à savoir la loi de mortalité des autonomes, la loi d'entrée en incidence des autonomes et la loi de mortalité des dépendants. En effet, chaque année l'assureur doit être en mesure d'estimer le nombre d'assurés non dépendants et donc versant des cotisations (ou du moins le nombre d'assurés sous risque, puisqu'il peut y avoir des cas de réduction pour lesquels nous ne recevons plus de primes), le nombre d'assurés devenant ou étant dépendants, c'est-à-dire ceux à qui il faudra payer une prestation.

Signalons que contrairement à d'autre type de couverture, par exemple la couverture décès ou viagère, le Code des Assurances n'impose pas l'utilisation d'une table spécifique pour la tarification de la dépendance.

Nous traitons ici le cas d'un seul type de dépendance, admettons GIR 1-2 donc dépendance totale. Nous verrons dans la suite comment nous intégrons une garantie pour de la dépendance partielle. Notons :

- q_x^a la probabilité pour un assuré non dépendant (autonome) de décéder entre l'age x et x+1 ;
- ${}_kp_x^a$ la probabilité pour un assuré autonome d'âge x de survivre encore k années (sans dépendance) ;
- i_x la probabilité pour un assuré autonome de devenir dépendant dans sa x ème année ;
- q_x^i la probabilité pour un assuré dépendant de décéder entre l'âge x et x+1 ;
- ${}_kp_x^i$ la probabilité pour un assuré dépendant d'âge x de vivre encore k années en état de dépendance.

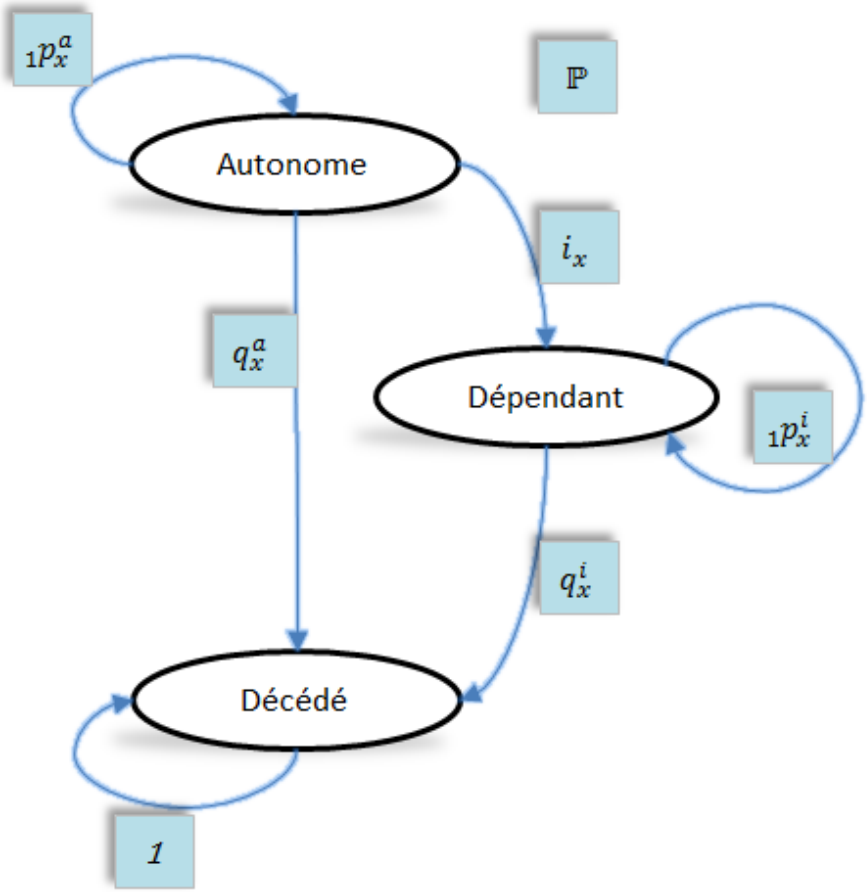

Illustration 2.1 : *Illustration de l'application des différentes probabilités dans l'année pour un individu d'âge x*

De plus, nous avons v, le taux d'actualisation égal à :

$$v = \frac{1}{1+i}$$

avec i le taux d'actualisation réglementaire. Ce taux est le rendement financier minimum sur lequel s'engage un assureur pour un contrat et qui anticipe, dans le calcul des cotisations ou des provisions mathématiques, le placement des flux futurs entrants pour l'assureur. Ce taux est indexé sur le taux moyen des emprunts d'Etat. L'article Article A132-1 et A132-1-1 du Code des Assurances en fixe les modalités de calcul.

Notons aussi a_x le capital constitutif d'une rente viagère annuelle à termes échus où :

$$a_x = \sum_{k>=1} v^k \cdot {}_kp_x$$

et celui d'une rente fractionnée en m échéances, $a_x^{(m)}$:

$$a_x^{(m)} = \frac{1}{m} \sum_{k>=1} v^{\frac{k}{m}} \cdot {}_{\frac{k}{m}}p_x$$

Engagement de l'assureur L'engagement pur de l'assureur se compose des prestations à payer lors de la survenance de la dépendance. Mais il est limité par d'éventuelles carences, ainsi que franchises.

Chaque année, pour un assuré d'âge x, l'espérance de l'engagement de l'assureur est égal à la probabilité que l'assuré tombe en dépendance, multipliée par le montant de prestation, conditionnée par le fait que l'assuré survive jusqu'à l'année x. Le montant de la prestation est égal au capital constitutif d'une rente viagère annuelle à termes échus, fractionnée en m paiements durant l'année. Notons Π_x cet engagement :

$$\Pi_x = \sum_{k>=0} v^k \cdot {}_kp_x^a \cdot i_{x+k} \cdot a_{x+k}^{i(m)}$$

Engagement de l'assuré L'engagement de l'assuré est quant à lui le paiement de primes tant que l'assuré est autonome. L'engagement de l'assuré est donc égal au capital constitutif d'une rente viagère à termes à échoir :

$$\ddot{a}_x^a = \sum_{k>=0} v^k \cdot {}_kp_x^a$$

Equité actuarielle A la souscription, l'engagement de l'assureur est égal à l'engagement de l'assuré. La prime périodique ω_x, vérifie alors, pour 1 de rente :

$$\omega_x \cdot \ddot{a}_x^a = \Pi_x$$

d'où

$$\omega_x = \frac{\Pi_x}{\ddot{a}_x^a}$$

Ajout d'un délai de carence Si nous considérons maintenant la prise en compte d'un délai de carence α, qui vient en diminution de l'engagement de l'assuré, la prime pure unique avec délai de carence Π_x^α est alors égale à :

$$\Pi_x^\alpha \cdot (1 - \sum_{k>=0}^{\alpha-1} v^k \cdot {}_kp_x^a \cdot i_{x+k}) = \Pi_{x+\alpha} \cdot v^\alpha \cdot {}_\alpha p_x^a$$

d'où

$$\Pi_x^\alpha = \frac{\Pi_{x+\alpha} \cdot v^\alpha \cdot {}_\alpha p_{x+\alpha}^a}{(1 - \sum_{k>=0}^{\alpha-1} v^k \cdot {}_kp_x^a \cdot i_{x+k})}$$

Nivellement de la prime En cas de primes nivelées, la prime périodique ω_x^α vérifie :

$$\omega_x^\alpha \cdot \ddot{a}_x^a - \omega_x^\alpha \cdot \sum_{k>=0}^{\alpha-1} (k+1) \cdot v^k \cdot {}_kp_x^a \cdot i_{x+k} = \Pi_{x+\alpha} \cdot v^\alpha \cdot {}_\alpha p_x^a$$

d'où

$$\omega_x^\alpha = \frac{\Pi_{x+\alpha} \cdot v^\alpha \cdot {}_\alpha p_x^a}{\ddot{a}_x^a - \sum\limits_{k>=0}^{\alpha-1} (k+1) \cdot v^k \cdot {}_k p_x^a \cdot i_{x+k}}$$

Ajout de plusieurs types de carences Bien souvent, l'assureur distingue plusieurs types de carences. Supposons cette définition des carences, communément utilisée dans les contrats dépendance. Premièrement, si la dépendance survient en raison d'un accident, il n'y a pas de délai de carence, l'assureur prend en charge immédiatement l'assuré. Deuxièmement, en cas de maladie autre que démence, le délai de carence est égal à α_1. Troisièmement, en cas de démence, le délai de carence est égal à α_2 de telle sorte que $\alpha_1 < \alpha_2$ (généralement respectivement 2 et 3 ans). En outre, nous devons faire une hypothèse quant à la probabilité pour un assuré de devenir dépendant pour les 3 causes précitées. Nous notons δ_1 la probabilité pour un assuré de devenir dépendant suite à une maladie autre que démence et δ_2 la probabilité de devenir dépendant pour cause de démence.

Ainsi la prime unique est égale à :

$$\Pi_x^{\alpha_1,\alpha_2} = \frac{(1-\delta_1-\delta_2) \cdot \Pi_x + \delta_1 \cdot \Pi_{x+\alpha_1} \cdot v^{\alpha_1} \cdot {}_{\alpha_1} p_x^a + \delta_2 \cdot \Pi_{x+\alpha_2} \cdot v^{\alpha_2} \cdot {}_{\alpha_2} p_x^a}{1 - (\delta_1 \cdot \sum\limits_{k>=0}^{\alpha_1-1} v^k \cdot {}_k p_x^a \cdot i_{x+k} + \delta_2 \cdot \sum\limits_{k>=0}^{\alpha_2-1} v^k \cdot {}_k p_x^a \cdot i_{x+k})}$$

Et la prime périodique :

$$\omega_x^{\alpha_1,\alpha_2} = \frac{(1-\delta_1-\delta_2) \cdot \Pi_x + \delta_1 \cdot \Pi_{x+\alpha_1} \cdot v^{\alpha_1} \cdot {}_{\alpha_1} p_x^a + \delta_2 \cdot \Pi_{x+\alpha_2} \cdot v^{\alpha_2} \cdot {}_{\alpha_2} p_x^a}{\ddot{a}_x^a - (\delta_1 \cdot \sum\limits_{k>=0}^{\alpha_1-1} (k+1) \cdot v^k \cdot {}_k p_x^a \cdot i_{x+k} + \delta_2 \cdot \sum\limits_{k>=0}^{\alpha_2-1} (k+1) \cdot v^k \cdot {}_k p_x^a \cdot i_{x+k})}$$

Tarification de garantie avec dépendance partielle et totale Il est fréquent que les assureurs proposent des garanties distinguant dépendance totale (GIR 1 à 2 par exemple) et dépendance partielle (GIR 3), en proposant un montant de rente plus faible lors de la dépendance partielle. Cette pratique permet d'offrir une meilleure couverture aux assurés, qui bénéficient ainsi d'une couverture en cas de dépendance partielle et totale, tout en limitant son coût, puisqu'à priori être en dépendance partielle a un coût plus faible qu'être en dépendance totale.

L'approche la plus robuste en terme de probabilités et la plus simple consiste à considérer qu'il s'agit de deux contrats.

En effet, il demeure une tâche peu aisée d'évaluer des lois spécifiques concernant la dépendance partielle, par manque d'observation et de recul sur cette garantie spécifique. Il en est de même pour la loi de passage de dépendance partielle vers dépendance totale. Ainsi, la tarification de ce produit consiste en l'addition de deux

garanties : une garantie couvrant la dépendance partielle et totale et offrant un montant de rente réduit (par exemple 70%) en cas de dépendance partielle ou totale et une autre garantie couvrant la dépendance totale, offrant un montant de rente complémentaire en cas de survenance du risque (dépendance totale), afin d'obtenir un montant de rente de 100% en cas de dépendance totale.

Ainsi, dans cette approche, nous divisons la population assurée en deux catégories :

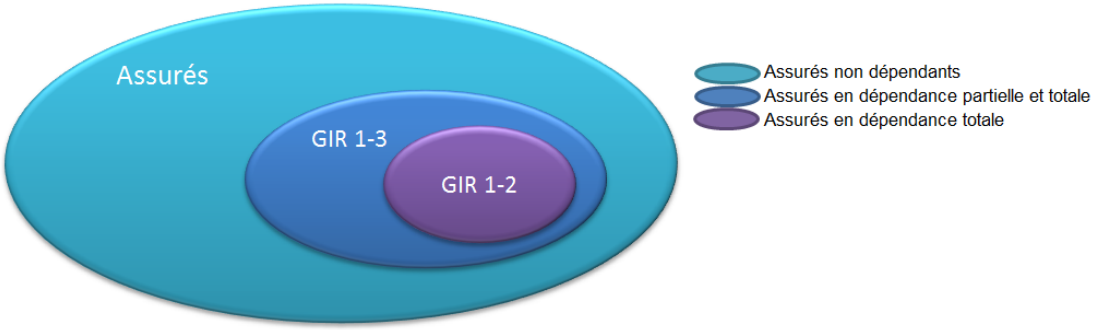

Illustration 2.2 : *Segmentation de la population*

Nous avons alors, par exemple pour la prime nivelée ω_x^{DP+DT}, dans le cas où 70% de la rente est garanti en cas de dépendance partielle :

$$\omega_x^{DP+DT} = \omega_x^{DP} + (100\% - 70\%) \cdot \omega_x^{DT}$$

où ω_x^{DP} représente la prime nivelée calculée précédemment, avec les probabilités i_x et q_{ix} concernent le GIR 1 à 3 et ω_x^{DP} avec i_x et q_{ix} concernant le GIR 1 à 2.

Les primes concernant les garanties avec carence sont calculées de la même façon, de même que le sont les provisions pour risques croissants, et les provisions mathématiques de rente, à savoir en effectuant une tarification comme s'il s'agissait de deux contrats comme précités.

Ce calcul est une approximation, il ne prend pas en compte les probabilités de passage entre la dépendance partielle et la dépendance totale. Toutefois dans un but de simplification, il s'agit d'une pratique courante.

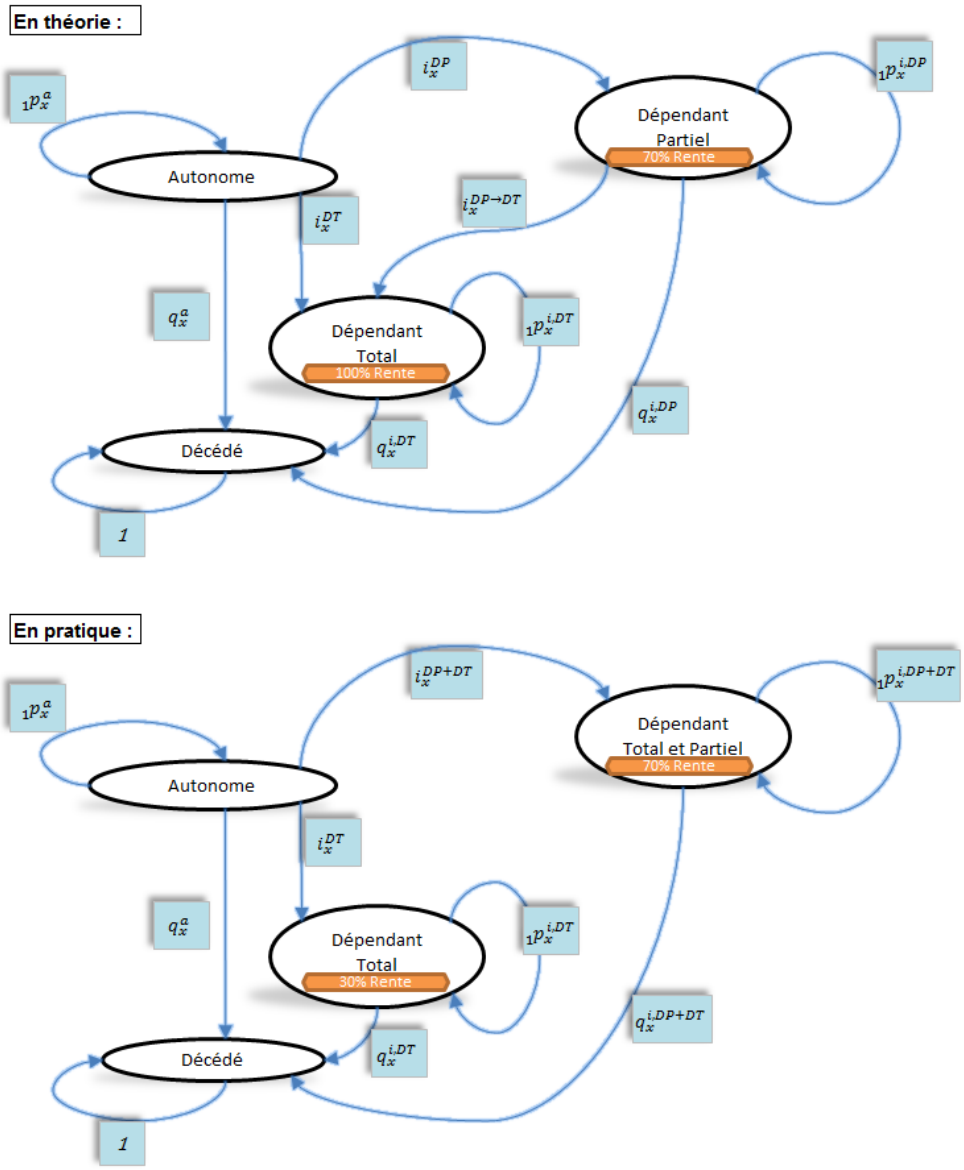

Illustration 2.3 : *Tarification d'une garantie Dépendance partielle et totale*

Nous abordons maintenant le thème des provisions, c'est à dire la "réserve" que l'assureur doit constituer afin d'être en mesure d'honorer les engagements qu'il a pris.

2.3 Calcul des provisions

Les provisions doivent permettre à l'assureur d'être en mesure à tout instant de respecter ses engagements, dans notre cas, payer une rente à un assuré devenant dépendant. Elles correspondent à la reconnaissance de la dette probable qu'à un assureur envers ses assurés et constituent un élément de solvabilité majeur en vue de garantir une continuité d'activité aux assurés.

Il s'agit d'une obligation réglementaire : "Les engagements réglementés dont les entreprises[8] mentionnées à l'article L. 310-1 ou au 1 du III de l'article L. 310-1-1 doivent, à toute époque, être en mesure de justifier l'évaluation sont les suivants :
1 Les provisions techniques suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés, des entreprises réassurées et bénéficiaires de contrats ..."[9]

Nous distinguons la provision mathématique de rente, qui est la provision effectuée pour payer la rente lorsque l'assuré devient dépendant et la provision pour risque croissant, qui est la provision destinée à faire face au fait que l'assuré peut devenir dépendant.

2.3.1 Provision mathématique de rente

La provision pour sinistre à payer ou provision mathématique de rente est la provision constituée lorsque le sinistre survient. Il s'agit d'une provision devant permettre à l'assureur de verser à l'assuré la rente viagère dont il est bénéficiaire. C'est donc le capital constitutif d'une rente viagère à termes échus. Appelons ${}_tPMR_x$ la provision mathématique de rente de l'année t pour un assuré d'âge x à l'adhésion. Nous avons

$${}_tPMR_x = a_{x+t}^{i,(m)}$$

où

$$a_{x+t}^{i,(m)} = \sum_{k>=0} v^{\frac{k}{m}} \cdot {}_{\frac{k}{m}}p_{x+t}^{i}$$

2.3.2 Provision pour Risques Croissants

La provision pour risques croissants est la provision calculée pour les contrats à primes nivelées dans le but de prendre en compte le risque de survenance de la dépendance croissante avec l'âge de l'assuré.

En effet, en début de contrat, l'assureur perçoit un montant de prime plus élevé que le risque probable qu'il supporte. Cette provision est ainsi constituée par la différence entre l'engagement de l'assureur et l'engagement de l'assuré, en faveur de l'assureur en début de contrat (cf. Illustration 2.4). Puis vient un moment où la prime perçue par l'assureur devient inférieure au risque qu'il supporte. La provision qu'il a alors constituée en début de contrat lui permet de combler cette inégalité. Notons ${}_tPRC_x$ la provision pour risque croissant de l'année t pour un assuré d'âge x à l'adhésion.

8. Voir annexes
9. Article R331-1 du Code des Assurances

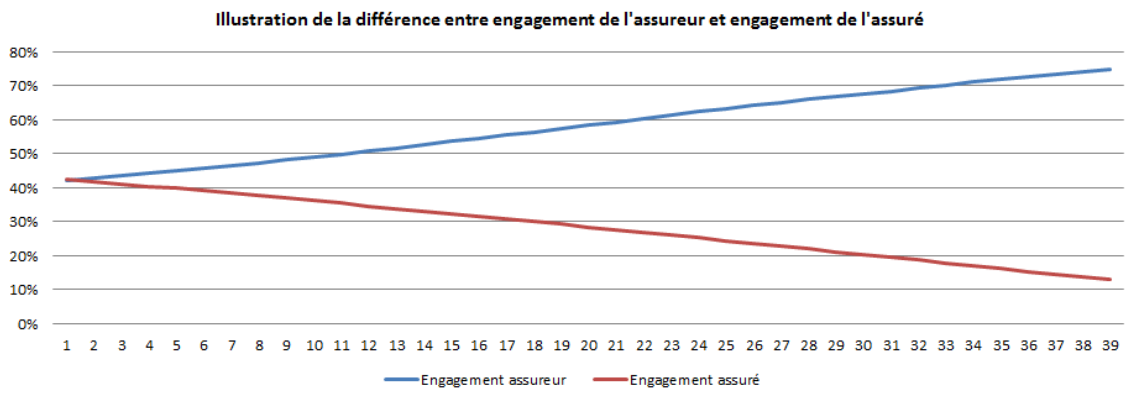

Illustration 2.4 : *Engagement probable de l'assureur et engagement probable de l'assuré*

Dans le cas sans carence, nous avons :

$$ {}_tPRC_x = \sum_{k>=0} v^k \cdot {}_kp^a_{x+t} \cdot i_{x+k+t} \cdot a^{i(m)}_{x+k+t} - \omega_x \cdot \sum_{k>=0} v^k \cdot {}_kp^a_x = \Pi_{x+t} - \omega_x \cdot \ddot{a}^a_{x+t} $$

Dans le cas des carences α_1 et α_2 comme mentionné précédemment, nous avons :
Si t<Min(α_1,α_2) alors :

$$ \begin{aligned} {}_tPRC_x = \quad & (1 - \delta_1 - \delta_2) \cdot \Pi^{\alpha_1,\alpha_2}_{x+t} \\ & + \delta_1 \cdot \Pi^{\alpha_1,\alpha_2}_{x+\alpha_1-t} \cdot v^{\alpha_1-t} \cdot {}_{\alpha_1-t}p^a_{x+t} \\ & + \delta_2 \cdot \Pi^{\alpha_1,\alpha_2}_{x+\alpha_2-t} \cdot v^{\alpha_2-t} \cdot {}_{\alpha_2-t}p^a_{x+t} \\ & + \delta_1 \cdot \sum_{k>=0}^{\alpha_1-1} (k+1+t) \cdot \omega^{\alpha_1,\alpha_2}_x \cdot v^k \cdot {}_kp^a_{x+t} \cdot i_{x+k+t} \\ & + \delta_2 \cdot \sum_{k>=0}^{\alpha_2-1} (k+1+t) \cdot \omega^{\alpha_1,\alpha_2}_x \cdot v^k \cdot {}_kp^a_{x+t} \cdot i_{x+k+t} \\ & - \omega^{\alpha_1,\alpha_2}_x \cdot \ddot{a}^a_{x+t} \end{aligned} $$

Si α_1<t<α_2 alors :

$$ \begin{aligned} {}_tPRC_x = \quad & (1 - \delta_1 - \delta_2) \cdot \Pi^{\alpha_1,\alpha_2}_{x+t} \\ & + \delta_2 \cdot \Pi^{\alpha_1,\alpha_2}_{x+\alpha_2-t} \cdot v^{\alpha_2-t} \cdot {}_{\alpha_2-t}p^a_{x+\alpha_2+t} \\ & + \delta_2 \cdot \sum_{k>=0}^{\alpha_2-1} (k+1+t) \cdot \omega^{\alpha_1,\alpha_2}_x \cdot v^k \cdot {}_kp^a_{x+t} \cdot i_{x+k+t} \\ & - \omega^{\alpha_1,\alpha_2}_x \cdot \ddot{a}^a_{x+t} \end{aligned} $$

Si $t > \alpha_2$, alors :

$$ {}_tPRC_x = \Pi^{\alpha_1,\alpha_2}_{x+k} - \ddot{a}^a_{x+k} \cdot \omega^{\alpha_1,\alpha_2}_x $$

2.4 Valeur de réduction

Si, après avoir payé ses primes durant un certain nombre d'années, l'assuré ne peut plus ou ne veut plus payer ses primes, alors les garanties peuvent être maintenues, même en l'absence de paiement de primes, mais avec un montant de rente réduit en cas de sinistre. Le montant de la rente réduite est déterminé en fonction du nombre de primes payées et de l'âge de l'assuré à la date de réduction. L'assureur est autorisé à appliquer une pénalité lors de la réduction. La réduction consiste à convertir la provision pour risques croissants constituée à la date de demande de réduction en une prime unique.

Soit R^d le montant de rente réduite. R^d vérifie :

$$\Pi_{x+k}^{\alpha_1,\alpha_2} \cdot R^d = R \cdot (\Pi_{x+k}^{\alpha_1,\alpha_2} - \ddot{a}_{x+k}^{a} \cdot \omega_x^{\alpha_1,\alpha_2})$$

d'où

$$R^d = R \cdot \frac{\Pi_{x+k}^{\alpha_1,\alpha_2} - \ddot{a}_{x+k}^{a} \cdot \omega_x^{\alpha_1,\alpha_2}}{\Pi_{x+k}^{\alpha_1,\alpha_2}}$$

$$R^d = (1 - \frac{\ddot{a}_{x+k}^{a} \cdot \omega_x^{\alpha_1,\alpha_2}}{\Pi_{x+k}^{\alpha_1,\alpha_2}}) \cdot R$$

Et ainsi le coefficient de réduction est égal à :

$$(1 - \frac{\ddot{a}_{x+k}^{a} \cdot \omega_x^{\alpha_1,\alpha_2}}{\Pi_{x+k}^{\alpha_1,\alpha_2}}) \cdot K$$

où $K < 1$ est une pénalité.

Nous avons présenté dans cette partie les méthodes de tarification d'une couverture dépendance, et le calcul des provisions en découlant pour l'assureur. Comme nous avons pu le voir, ce processus nécessite la connaissance de lois sur la mortalité des dépendants, la mortalité des autonomes et les probabilités de passage en dépendance. Dans la partie suivante nous décrierons des techniques permettant de calculer ces probabilités. Elles ne nous serviront pas à tarifier un produit dépendance, notre objectif principal étant de mesurer la rentabilité d'un produit dépendance qui a déjà été tarifié. Les lois que nous estimerons nous permettrons ainsi de projeter le résultat de ce produit.

Partie 3

Importance de l'estimation des lois d'expérience

3.1 Analyse de survie

La problématique est de mesurer statistiquement la probabilité de survie ou de maintien d'un individu dans un état donné. La durée de survie désigne le temps écoulé jusqu'à la survenue d'un évènement.
Par exemple, dans le milieu médical, nous pouvons chercher à vérifier l'effet d'un médicament sur la durée de vie d'un patient, ou dans le domaine de l'industrie, évaluer la durée de vie de pièces électroniques. Dans notre cas, c'est l'évaluation de la durée de vie d'un être humain qui va nous intéresser. L'estimation peut être paramétrique ou non, c'est à dire que nous pouvons supposer une loi a priori sur la mortalité des individus ou non.

Ces estimations ont pour but l'obtention de probabilités concernant la mortalité des assurés autonomes (non dépendants), la mortalité des assurés dépendants et l'entrée en dépendance des individus.

Elles nous serviront à projeter la mortalité et l'incidence des individus du portefeuille à étudier de la manière la plus proche possible de la réalité. En effet, nous serons ainsi en mesure de déterminer chaque année le nombre d'assurés autonomes qui décèdent, le nombre d'assurés autonomes entrants en dépendance et le nombre d'assurés dépendants qui décèdent.

L'objectif de cette partie est ainsi d'exposer des méthodes permettant d'obtenir des estimations de ces différentes probabilités de la façon la plus précise possible.

Dans un premier temps, nous présenterons des notions générales d'analyse de survie. Puis nous décrirons les problèmes auxquels nous pouvons être confrontés lors d'une telle analyse sur un portefeuille. Enfin nous développerons l'estimation non paramétrique et l'estimation paramétrique de ces probabilités.

3.1.1 Notions préliminaires

Pour faire le lien avec les probabilités précitées, notons :

$$_tp_x = \mathbb{P}(T > x + t | T > x)$$

la probabilité de survie d'un individu d'âge x durant t années encore, et

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

la probabilité de décès d'un individu d'âge x dans les k prochaines années.
Soit :

- l_x : le nombre de "survivants" à l'âge x
- d_x : le nombre de "décès" entre l'âge x et x+1

tel que :

$$l_{x+1} = l_x - d_x$$

Considérons N individus. Ce ne sont pas les observations de la durée de vie $T_1, ..., T_N$ pour chaque individu $i \in [1, N]$ qui sont utilisées mais la durée de vie observée des individus qui est égale à $\text{Min}(T_i, L_i)$ où L_i est l'âge atteint lors de l'observation si l'individu est toujours en vie. Si $T_i < L_i$, il s'agit d'une sortie pour survenance du risque, mais si $L_i > T_i$, il s'agit d'une censure à droite. Nous expliquerons ce qu'est et comment prendre en compte les censures par la suite. Intéressons nous d'abord à la situation où l'observation est complète.

Fonction de survie

Considérons T la variable aléatoire mesurant la "durée de vie" tel que T $\in \mathbb{R}_+^*$ Notons S(t) la fonction de survie d'un individu i.e. $S(t) = P(T > t)$, la probabilité qu'un individu d'âge x survive plus longtemps qu'un certain temps t, avec $t \in [0, \infty)$ et $S(t) \in [0, 1]$

Décrivons quelques propriétés de S(t) :

- $S(0) = 1$;
- $\lim_{t\to\infty} S(t) = 0$;
- $S(t_a) \geq S(t_b) \Leftrightarrow t_a \leq t_b$;
- $S(t) = 1 - F(t) = \int_t^\infty f(\tau)d\tau$, où F(t) est la fonction de distribution de T et f sa densité, si elle existe.

L'objectif est d'estimer la fonction S.

Densité de probabilité

Si F admet une dérivée au point t, alors :

$$f(t) = \lim_{h\to 0} \frac{\mathbb{P}(t \leq T < t + h)}{h}$$

f(t) représente la probabilité de mourir dans un court instant après t. Nous avons :

$$f(t) = F'(t) = -S'(t)$$

Taux de hasard

Introduisons aussi $\mu(t)$, la fonction de hasard ou de "décès" instantané, qui caractérise la probabilité de mourir dans un intervalle de temps très court après t, conditionnée par le fait d'être arrivé en vie jusque t. Elle est définie par :

$$\begin{aligned}
\mu(t) &= \lim_{h\to 0} \frac{\mathbb{P}(t \leq X < t+h | X \geq t)}{h} \\
&= \frac{f(t)}{S(t)} \\
&= -\frac{\frac{d}{dt}S(t)}{S(t)} \\
&= -\frac{d}{dt} ln(S(t))
\end{aligned}$$

(dans l'hypothèse où S est dérivable).

Nous avons donc :

$$\frac{d}{dt}S(t) = -\mu(t) \cdot S(t)$$

Après résolution de l'équation différentielle linéaire simple, nous avons :

$$\begin{aligned}
S(t) &= S(0) \cdot exp\left\{\int_0^t -\mu(y)dy\right\} \\
&= exp\left\{\int_0^t -\mu(y)dy\right\}
\end{aligned}$$

puisque S(0)=1

Taux de hasard cumulé

De même $\Lambda(t)$ est la fonction de hasard intégrée telle que :

$$\Lambda(t) = \int_0^t \mu(s)ds$$

Distribution empirique

Nous construisons la distribution empirique

$$\hat{F}_{\hat{T}_1,\ldots,\hat{T}_N}(y) = \hat{F}_N(y) = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathbb{1}_{\hat{T}_i \leq y}$$

Supposons que $T_1, ..., T_N$ soient des variables aléatoires de distribution identique F. Alors $\hat{F}_{\hat{T}_1,...,\hat{T}_N}(y)$ est un estimateur non biaisé pour F(y). En effet :

$$\begin{aligned}
\mathbb{E}(\hat{F}_{\hat{T}_1,...,\hat{T}_N}(y)) &= \mathbb{E}(\frac{1}{N} \cdot \sum_{i=1}^{N} \mathbb{1}_{T_i \leq y}) \\
&= \frac{1}{N} \cdot \sum_{i=1}^{N} \mathbb{E}(\mathbb{1}_{T_i \leq y}) \\
&= \frac{1}{N} \cdot \sum_{i=1}^{N} \mathbb{P}(T_i \leq y) \\
&= F(y)
\end{aligned}$$

Quant à la variance :

$$\begin{aligned}
\mathbb{V}(\hat{F}_{\hat{T}_1,...,\hat{T}_N}(y)) &= \mathbb{E}((\hat{F}_{\hat{T}_1,...,\hat{T}_N}(y))^2) - F^2(y) \\
&= \frac{1}{N^2} \cdot \sum_{i=1,j=1}^{N} \mathbb{E}(\mathbb{1}_{T_i \leq y} \cdot \mathbb{1}_{T_j \leq y}) - F^2(y) \\
&= \frac{1}{N^2} \cdot \sum_{i=1}^{N} \mathbb{E}(\mathbb{1}_{T_i \leq y}) + \frac{1}{N^2} \cdot \sum_{i \neq j=1}^{N} \mathbb{E}(\mathbb{1}_{T_i \leq y} \cdot \mathbb{1}_{T_j \leq y}) - F^2(y) \\
&= \frac{1}{N} \cdot F(y) + (\frac{1}{N^2} \cdot \sum_{i \neq j=1}^{N} F^2(y)) - F^2(y) \\
&= \frac{1}{N} F(y)(1 - F(y))
\end{aligned}$$

Ce qui prouve que l'estimateur est asymptotiquement sans biais.

Intéressons nous à la durée moyenne de vie d'un individu :

$$\mathbb{E}(T) = \int_0^\infty t \cdot f(t)dt = \int_0^\infty t \cdot \frac{d}{dt} F(t)dt = \int_0^\infty t \cdot S(t)dt$$

Notons $S_x(t) = {}_tp_x = \mathbb{P}(T > x + t | T > x)$. Nous avons alors :

$$S_x(t) = {}_tp_x = exp\left\{\int_0^t -\mu(x+y)dy\right\}$$

De la même façon, la durée de vie moyenne résiduelle d'un individu, T_x d'âge x est donnée par :

$$\mathbb{E}(T_x) = \int_x^\infty t \cdot S(t)dt$$

Notons $r(t) = \mathbb{E}(T - t | T > t)$

Par ailleurs la variance de la durée de vie peut s'écrire :

$$\mathbb{V}(T) = 2 \cdot \int_0^\infty \mathbb{E}(t)dt - \mathbb{E}(T)^2$$

3.1.2 Modèles non paramétriques

Analyse de survie

Nous cherchons à estimer la probabilité qu'un individu passe durant sa x ème année d'un état à l'autre. Intuitivement nous pouvons supposer que le passage dans l'état étudié est une épreuve de Bernoulli de paramètre i_x pour l'individu d'âge x. Ainsi, si nous considérons un effectif sous risque à l'âge x noté n_x, où chaque individu est indépendant l'un de l'autre, le nombre de personnes dans l'état étudié est une variable aléatoire suivant une loi Binomiale de paramètre (n_x, i_x). En notant d_x le nombre d'entrée dans l'état étudié l'année x, nous avons :

$$\mathbb{P}(d_x = k) = C_{n_x}^k \cdot i_x^k \cdot (1 - i_x)^k$$

Or pour chaque année d_x et n_x sont connus. Il suffit donc de rechercher le paramètre i_x qui vérifie cette équation. Pour ce faire, nous utilisons la méthode du maximum de vraisemblance.

Avec la vraisemblance de la loi Binomiale précédemment citée :

$$L(n_x^1, ..., n_x^{n_x}; i_x) = \prod_{k=1}^{n_x} C_{n_x^k}^k \cdot i_{x^k}^k \cdot (1 - i_x)^k$$

Intéressons nous d'abord à la situation où l'observation est complète, c'est à dire que l'individu est observé depuis sa "naissance" jusqu'à sa "mort". Soit les observations $\hat{T}_1, ..., \hat{T}_{N_0}$ Considérons des intervalles de temps non forcément équidistants $\{\tau_0, \tau_1, ...\}$ telles que $\tau_i < \tau_j$ si $i < j$

Dans notre cas, $\{\tau_0 = 0, \tau_1 = 1, ..., \tau_\omega = \omega\}$ où ω sera l'âge de fermeture de la table.

Nous définissons :

- $\hat{D}_i = \sum_{1 \leq k \leq N_0} \mathbb{1}_{\tau_i \leq \hat{T}_k \leq \tau_{i+1}}$ qui représente le nombre de décès entre τ_i et τ_{i+1} ;
- $\hat{N}_i = \sum_{k \geq 1} \mathbb{1}_{\hat{T}_k \geq \tau_i}$ qui représente le nombre de sujets encore vivants après τ_i.

L'estimateur conditionnel de $\mathbb{P}(T \leq x + 1 | T > x) = \hat{q}_{\tau_i} = \frac{\hat{D}_i}{\hat{N}_i}$, obtenu par maximisation de la vraisemblance citée précédemment.

Cet estimateur est non biaisé sous la filtration $\left\{\hat{T} \geq \tau_i\right\}$. En effet :

$$
\begin{aligned}
\mathbb{E}(\hat{q}_{\tau_i}|\hat{T}_1, ..., \hat{T}_{N_0} \geq \tau_i) &= \mathbb{E}(\frac{\hat{D}_i}{\hat{N}_i}|\hat{T}_1, ..., \hat{T}_{N_0} \geq \tau_i) \\
&= \frac{1}{\hat{N}_i} \cdot \mathbb{E}(\hat{D}_i|\hat{T}_1, ..., \hat{T}_{N_0} \geq \tau_i) \\
&= \frac{1}{\hat{N}_i} \cdot \mathbb{E}(\sum_{1 \leq k \leq N_0} \mathbb{1}_{\tau_i \leq \hat{T}_k \leq \tau_{i+1}}|\hat{T}_1, ..., \hat{T}_{N_0} \geq \tau_i) \\
&= \frac{1}{\hat{N}_i} \cdot \sum_{1 \leq k \leq N_0} \mathbb{E}(\mathbb{1}_{\tau_i \leq \hat{T}_k \leq \tau_{i+1}}|\hat{T}_1, ..., \hat{T}_{N_0} \geq \tau_i) \\
&= \frac{1}{\hat{N}_i} \cdot \hat{N}_i \cdot \hat{q}_{\tau_i} \\
&= \frac{1}{\hat{N}_i} \cdot \hat{N}_i \cdot \frac{\hat{D}_i}{\hat{N}_i} \\
&= \frac{1}{\hat{N}_i} \cdot \hat{N}_i \cdot \hat{q}_{\tau_i} \\
&= \hat{q}_{\tau_i}
\end{aligned}
$$

La variance conditionnelle de $\hat{q}_{\tau_i}$ est égale à :

$$
\mathbb{V}(\hat{q}_{\tau_i}|T \geq \tau_i) = \frac{1}{\hat{N}_i} \cdot \hat{q}_{\tau_i} \cdot (1 - \hat{q}_{\tau_i})
$$

Nous retrouvons bien la variance d'une loi Binomiale de paramètre $\hat{N}_i$ et $\hat{q}_{\tau_i}$.

3.1.3 Hypothèse pour les sous intervalles

Ces probabilités peuvent aussi être considérées pour les sous intervalles, entre deux instants par exemple. Plusieurs hypothèses permettent de déterminer ces probabilités.

Hypothèse de linéarité

Soit s $\in [0, 1]$, faisons l'hypothèse que :

$$
l_{(1-s)\cdot\tau_i + s\cdot\tau_{i+1}} = (1 - s) \cdot l_{\tau_i} + s \cdot l_{(1-s)\tau_i + s\cdot\tau_{i+1}}
$$

Ainsi :

$$
\begin{aligned}
S(s \cdot (\tau_{i+1} - \tau_i)) &= \frac{l_{(1-s)\cdot\tau_i + s\cdot\tau_{i+1}}}{l_0} \\
&= (1 - s) \cdot S(\tau_i) + s \cdot S(\tau_{i+1})
\end{aligned}
$$

$$
\begin{aligned}
{}_{s\cdot(\tau_{i+1}-\tau_i)}q_{\tau i} &= \mathbb{P}(\tau_i \leq T \leq s\cdot(\tau_{i+1}-\tau_i)|T \geq \tau_i) \\
&= \frac{S(\tau_i) - S(s(\tau_{i+1}-\tau_i))}{S(\tau_i)} \\
&= \frac{S(\tau_i) - (1-s)\cdot S(\tau_i) + s\cdot S(\tau_{i+1})}{S(\tau_i)} \\
&= s\cdot(1 - \frac{S(\tau_{i+1})}{S(\tau_i)}) \\
&= s\cdot\ {}_{\tau_{i+1}-\tau_i}q_{\tau_i}
\end{aligned}
$$

L'hypothèse faite ici est que $\mu(x) = \mu(x+\epsilon)$ pour ϵ assez petit.

Par ailleurs :

$$S_x(t) = {}_tp_x = exp\left\{\int_0^t -\mu(x+y)dy\right\}$$

$$\Rightarrow -ln({}_tp_x) = \int_0^t \mu(x+y)dy$$

$$\Rightarrow -ln(p_x) = \mu(x)$$

Puisque $\mu(x) = \mu(x+\epsilon)$ pour ϵ assez petit. Sous cette hypothèse :

$$-ln(1-q_x) = \mu(x)$$

Hypothèse d'exponentialité

Cette fois-ci nous supposons :

$$l_{(1-s)\cdot\tau_i+s\cdot\tau_{i+1}} = l^s_{\tau_{i+1}} + l^{1-s}_{\tau_i}$$

Ou de manière équivalente :

$$
\begin{aligned}
l_{(1-s)\cdot\tau_i+s\cdot\tau_{i+1}} &= l^s_{\tau_{i+1}} + l^{1-s}_{\tau_i} \\
&= l_{\tau_i}(\frac{l_{\tau_{i+1}}}{l_{\tau_i}})^s \\
&= l_{\tau_i}(1-q_{\tau_i})^s
\end{aligned}
$$

Ainsi nous avons :

$$
\begin{aligned}
{}_{s\cdot(\tau_{i+1}-\tau_i)}q_{\tau i} &= \frac{l_{\tau_i} - l_{s\cdot(\tau_{i+1}-\tau_i)}}{l_{\tau_i}} \\
&= 1 - (1-q_{\tau_i})^s
\end{aligned}
$$

3.1.4 Censure et troncature

La censure est le fait qu'une observation soit incomplète pour une cause aléatoire. Cette cause aléatoire doit être indépendante de l'évènement de sortie étudié. La troncature est le fait qu'une observation soit incomplète du fait par exemple, qu'elle ait commencé avant le début de l'étude. Prenons le cas de l'étude de la durée de survie d'une personne atteinte du Sida, l'observation sera alors tronquée si l'inoculation a été effectuée avant le début de l'étude. Elle sera censurée s'il n'est plus possible de suivre l'individu.

Le plus souvent, la censure est une censure à droite. Dans ce cas, le sujet est suivi, puis sort de l'étude avant que l'évènement que nous tentons d'étudier ne survienne. Dans le cas de notre portefeuille, en fonction de l'étude à effectuer, cela peut être dû à une résiliation ou au fait que l'assuré n'a pas subi l'évènement au moment de l'étude. Même si ces observations ne sont pas complètes, nous ne pouvons pas les exclure. En effet, ne pas les prendre en compte biaiserait les résultats et augmenterait de manière non justifiée la variance de notre estimation.

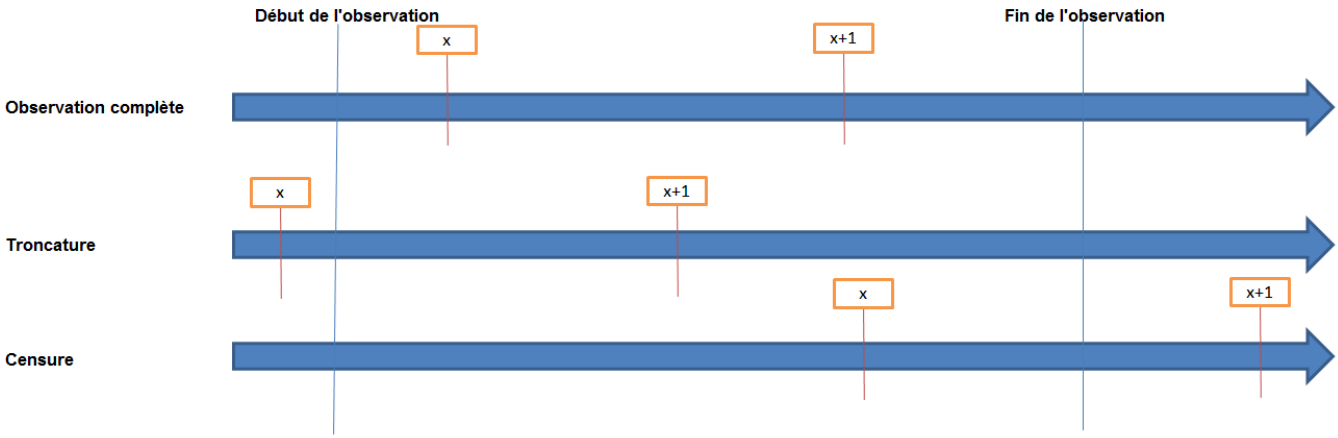

Illustration 3.1 : *Illustration de l'observation complète, de la troncature et de la censure avec x l'âge d'un assuré*

Nous présenterons trois estimateurs permettant de déterminer des fonctions de survie en présence de censure. Il s'agit de trois estimateurs classiques, à savoir l'estimateur de Kaplan Meier, l'estimateur de Nelson Aalen et l'estimateur actuariel. Dans la pratique, les deux premiers requièrent une base de données avec des dates de sorties précisément renseignées. Le dernier est moins précis mais plus utile lorsqu'une observation précise du portefeuille est impossible.

Estimateur de Kaplan-Meier

L'étude étant faite sur un portefeuille actif, les données sont forcément incomplètes. En effet, les informations les plus récentes du portefeuille présentent toujours

des individus pour lesquels la sortie n'est pas survenue. De même, certains individus sont sortis du portefeuille avant la fin de la date d'étude, mais pour des raisons autres que la sortie étudiée (par exemple résiliation). Tous ces assurés véhiculent une information qu'il est opportun de prendre en compte afin de ne pas biaiser l'étude. L'estimateur de Kaplan-Meier, mis au point en 1958, à la base pour étudier une population de cancéreux (Meier) et la durée de vie de tube à vide (Kaplan), tient compte de ces sorties.

L'idée de l'estimateur de Kaplan Meier est que pour survivre après un temps t, il faut être en vie juste avant cet instant t, et ne pas mourir au temps t.

Soit un échantillon de taille n.

Soit d_j le point de troncature du jème individu, soit x_j son observation non censurée et u_j son observation éventuellement censurée.

Soient $y_1 < y_2 < ... < y_k$ les k valeurs prises par x_j

Soit s_j le nombre de fois où l'observation non censurée y_j apparaît dans l'échantillon.

L'effectif sous risque r_j est défini comme :
r_j = le nombre d'individus observables à l'âge j + ceux qui décèdent à l'âge j ou plus tard + ceux qui sont censurés à cette âge ou plus tard

$$r_j = \sum_{i \in \theta_j} \mathbb{1}_{x_i \geq y_j} + \sum_{i \in \theta_j} \mathbb{1}_{u_i \geq y_j} - \sum_{i \in \theta_j} \mathbb{1}_{d_i \geq y_j}$$

Nous avons aussi r_j égal à l'effectif sous risque en j-1 + ceux dont la troncature intervient entre y_j et y_{j-1} - ceux dont la sortie s'effectue entre cette période, pour réalisation de l'évènement ou censure. En cas d'ex æquo, une entrée en portefeuille précède une censure qui précède elle-même une survenance du risque.

$$r_j = r_{j-1} + \sum_{i \in \theta_j} \mathbb{1}_{y_{j-1} \leq d_i < y_j} - \sum_{i \in \theta_j} \mathbb{1}_{y_{j-1} \leq x_i < y_j} - \sum_{i \in \theta_j} \mathbb{1}_{y_{j-1} \leq u_i < y_j}$$

Nous avons :

$$S_n(t) = \begin{cases} 1 & , \quad 0 \leq t < y_1 \\ \prod_{i=1}^{j-1} \frac{r_i - s_i}{r_i} & , \quad y_{j-1} \leq t < y_j, j = 2, ...k \\ \prod_{i=1}^{k} \frac{r_i - s_i}{r_i} \text{ ou } 0 & , \quad t \geq y_k \end{cases}$$

Le choix entre 0 ou $\frac{r_i - s_i}{r_i}$ provient du fait qu'en raison de la censure possible survenant avant le dernier décès, la fonction de survie pourrait ne pas valoir 0 puisqu'il serait possible que les individus censurés survivent après cette dernière date. Une option est alors de maintenir la valeur de la fonction de survie après ce dernier décès à sa dernière valeur. Une autre option est de lui attribuer la valeur 0. Nous choisirons dans la suite l'option la plus prudente, c'est à dire le maintien de la fonction de survie à sa dernière valeur, laissant présager une survie plus longue.

Nous obtenons ainsi la fonction de survie de Kaplan Meier, avec la probabilité de "sortie" $\frac{s_i}{r_i}$ à chaque instant ou âge i.

Variance de l'estimateur de Kaplan Meier :

$$\begin{aligned}\hat{\mathbb{V}}(S_n(y_j)) &= \frac{S(y_j)^2}{S(y_0)^2}\{\prod_{i=1}^{j}(1+\frac{S(y_{i-1})-S(y_i)}{r_iS(y_i)})-1\}\\ &\simeq \frac{S(y_j)^2}{S(y_0)^2}\{\sum_{i=1}^{j}\frac{S(y_{i-1})-S(y_i)}{r_iS(y_i)}\}\\ &\simeq S_n(y_j)^2\sum_{i=1}^{j}\frac{s_i}{r_i(r_i-s_i)}\end{aligned}$$

cette dernière approximation étant l'estimateur de Greenwood.

Cet estimateur permet d'établir un intervalle de confiance au seuil 1-α autour de l'estimation de la fonction de survie :

$$S_n(y_j)^+ = S_n(y_j) + z_{1-\alpha/2}\sqrt{\hat{\mathbb{V}}(S_n(y_j))}$$

$$S_n(y_j)^- = S_n(y_j) - z_{1-\alpha/2}\sqrt{\hat{\mathbb{V}}(S_n(y_j))}$$

avec $z_{1-\frac{\alpha}{2}}$ tel que $\mathbb{P}(U \leq \alpha) = 1 - \frac{1-\alpha}{2}$ où U suit une loi Normale N(0,1)

Toutefois, il est possible que cet intervalle de confiance déborde de l'intervalle des valeurs admissibles pour une fonction de survie, à savoir [0,1]. Il est possible d'obtenir un intervalle de confiance, via l'utilisation d'une fonction bijective, de borner cet intervalle de confiance. En prenant la fonction logarithme, nous obtenons :

$$S_n(y_j)^+ = S_n(y_j)^\theta$$

$$S_n(y_j)^- = S_n(y_j)^{\frac{1}{\theta}}$$

avec

$$\theta = \exp[\frac{z_{1-\alpha/2}}{log(S_n(y_j))}\sqrt{\sum_{i=1}^{j}\frac{s_i}{r_i(r_i-s_i)}}]$$

Exemple simplifié de l'utilisation de Kaplan Meier sur un échantillon de la base de données, pour l'estimation des taux de décès Voir le Tableau 3.1 pour un échantillon de la base sur lequel nous effectuons le calcul présenté dans le Tableau 3.2.

Tableau 3.1 : *Échantillon fictif de la base de données*

Année observation	**Age à la souscription**	**Année entrée**	**Age sortie**	**Motif sortie**
2010	70	2000	77	Décès
2010	70	2000	80	
2010	70	2000	77	Résiliation
2010	70	2000	80	
2010	70	2000	75	Décès
2010	70	2000	77	Décès
2010	70	2000	74	Décès
2010	70	2000	80	
2010	70	2000	75	Décès
2010	60	2000	70	
2010	60	2000	70	
2010	60	2000	61	Résiliation
2010	60	2000	70	
2010	60	2000	64	Décès
2010	60	2000	69	Décès
2010	60	2000	64	Décès
2010	60	2000	70	
2010	60	2000	70	
2010	50	2000	60	
2010	50	2000	60	
2010	50	2000	60	
2010	50	2000	54	Résiliation
2010	50	2000	56	Décès
2010	50	2000	56	Décès
2010	50	2000	60	
2010	50	2000	60	
2010	50	2000	60	

Estimateur de Nelson Aalen de la fonction de hasard cumulée

Ici le décès instantané est défini comme la proportion de la population qui sort au temps t, c'est à dire $\frac{s_i}{r_i}$. La fonction de hasard cumulée, H(t) est alors la somme des décès instantanés jusqu'au temps t.

$$\hat{H}(t) = \begin{cases} 0 & , \quad 0 \leq t < y_1 \\ \sum\limits_{i=1}^{j-1} \frac{s_i}{r_i} & , \quad y_{j-1} \leq t < y_j, j = 2, ...k \\ \sum\limits_{i=1}^{k} \frac{s_i}{r_i} & , \quad t \geq y_k \end{cases}$$

Tableau 3.2 : *Détermination de l'effectif sous risque, du nombre de décès et de censure pour chaque âge et calcul de la fonction de survie de Kaplan Meier et des taux de mortalité bruts*

Age	**Effectif sous risque**	**Nombre de décès**	**Nombre de censure**	**S(t)**	**qt**	**S(t)**	**qt**
50	9	0	0	100%	0%	1-0/9	1-S(51)/S(50)
51	9	0	0	100%	0%	1-0/9*S(50)	1-S(52)/S(51)
52	9	0	0	100%	0%	1-0/9*S(51)	1-S(53)/S(52)
53	9	0	0	100%	0%	1-0/9*S(52)	1-S(54)/S(53)
54	8	0	1	100%	0%	1-0/8*S(53)	1-S(55)/S(54)
55	8	0	0	100%	33%	1-0/8*S(54)	1-S(56)/S(55)
56	6	2	0	67%	0%	1-2/6*S(55)	1-S(57)/S(56)
57	6	0	0	67%	0%	1-0/6*S(56)	1-S(58)/S(57)
58	6	0	0	67%	0%	1-0/6*S(57)	1-S(59)/S(58)
59	6	0	0	67%	0%	1-0/6*S(58)	1-S(60)/S(59)
60	9	0	0	67%	0%	1-0/9*S(59)	1-S(61)/S(60)
61	8	0	1	67%	0%	1-0/8*S(60)	1-S(62)/S(61)
62	8	0	0	67%	0%	1-0/8*S(61)	1-S(63)/S(62)
63	8	0	0	67%	33%	1-0/8*S(62)	1-S(64)/S(63)
64	6	2	0	44%	0%	1-2/6*S(63)	1-S(65)/S(64)
65	6	0	0	44%	0%	1-0/6*S(64)	1-S(66)/S(65)
66	6	0	0	44%	0%	1-0/6*S(65)	1-S(67)/S(66)
67	6	0	0	44%	0%	1-0/6*S(66)	1-S(68)/S(67)
68	6	0	0	44%	20%	1-0/6*S(67)	1-S(69)/S(68)
69	5	1	0	36%	0%	1-1/5*S(68)	1-S(70)/S(69)
70	9	0	0	36%	0%	1-0/9*S(69)	1-S(71)/S(70)
71	9	0	0	36%	0%	1-0/9*S(70)	1-S(72)/S(71)
72	9	0	0	36%	0%	1-0/9*S(71)	1-S(73)/S(72)
73	9	0	0	36%	13%	1-0/9*S(72)	1-S(74)/S(73)
74	8	1	0	31%	33%	1-1/8*S(73)	1-S(75)/S(74)
75	6	2	0	21%	0%	1-2/6*S(74)	1-S(76)/S(75)
76	6	0	0	21%	67%	1-0/6*S(75)	1-S(77)/S(76)
77	3	2	1	7%	100%	1-2/3*S(76)	

et la fonction de survie de Fleming Harrington [1] peut être obtenue par :

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

Variance de l'estimateur de Nelson Aalen :

$$\hat{\mathbb{V}}(\hat{H}(y_j)) = \hat{\mathbb{V}}(\sum_{i=1}^{j-1} \frac{s_i}{r_i}) = \sum_{i=1}^{j-1} \frac{\hat{\mathbb{V}}(s_i)}{r_i^2} = \sum_{i=1}^{j-1} \frac{s_i}{r_i^2}$$

Et de la même façon un intervalle de confiance de type gaussien est construit autour de l'estimateur de Nelson Aalen :

$$\hat{H}(t)^+ = \hat{H}(t) + z_\alpha \sqrt{\hat{\mathbb{V}}(\hat{H}(t))}$$

1. FLEMING T.R., HARRINGTON D.P. [1991] Counting processes and survival analysis, Wiley Series in Probability and Mathematical Statistics. New-York : Wiley.

$$\hat{H}(t)^{-} = \hat{H}(t) - z_{\alpha}\sqrt{\hat{\mathbb{V}}(\hat{H}(t))}$$

avec $z_{\frac{\alpha}{2}}$ tel que $\mathbb{P}(U \leq \alpha) = 1 - \frac{\alpha}{2}$ où U suit une loi Normale N(0,1)

Estimateur actuariel

Cette méthode est utilisée lorsqu'un grand nombre de données est disponible. Elle se base sur des données agrégées et fait l'hypothèse que les sorties se font de manière aléatoire à l'intérieur des intervalles de temps considérés, la force de mortalité est supposée constante dans ces intervalles. Ces hypothèses engendrent un biais dans l'estimation lorsque l'intervalle de temps est grand. Nous ne détaillons pas plus cet estimateur qui ne sera pas utilisé dans cette étude, puisque nous disposons de données suffisamment précises.

3.1.5 Modèles paramétriques

Les modèles paramétriques reposent sur une hypothèse sur la distribution des durées de survie. Ainsi, le modèle paramétrique peut être formulé en précisant la forme de la loi de durée. Nous étudierons deux modèles couramment utilisés, le modèle exponentiel et le modèle de Gompertz Makeham.

Modèle exponentiel

Il s'agit d'un modèle où nous supposons que le taux instantané de sortie est constant avec le temps. Ainsi si

$$\mu(t) = \lambda$$

et comme

$$S(t) = exp\left\{\int_0^t -\mu(y)dy\right\}$$

Nous avons

$$S(t) = e^{\lambda t}$$

Modèle de Gompertz-Makeham

Dans le modèle de Gompertz-Makeham [2], nous supposons que :

$$\mu(t) = \alpha + \beta \cdot \gamma^t$$

2. MAKEHAM W. [1860] « On the law of mortality. » Journal of the Institute of Actuaries, 13, 325-358.

où α est un paramètre non dépendant de l'âge, et $\beta \cdot \gamma^t$ modélise le vieillissement exponentiel par rapport au temps.
La fonction de survie est alors égale à :

$$S(t) = \exp(-\alpha t - \frac{\beta}{log(\gamma)}(\gamma^t - 1))$$

avec $\gamma > 1$ et $\beta > 0$ afin d'assurer une croissance des taux de mortalité en fonction du temps.

L'obtention des paramètres est effectuée par maximisation de la vraisemblance des observations (par exemple avec le procédé de Newton Raphson évoqué par la suite).

En résumé, l'estimation non paramétrique est une méthode robuste, facilement interprétable, permettant l'estimation des fonctions de survie lorsque peu d'information sur la distribution est disponible. Elle présente l'avantage de ne faire aucune hypothèse quant à la fonction de survie et d'être simple à mettre en place. L'estimateur actuariel permet d'effectuer des estimations fiables lorsqu'un grand nombre de données est disponible. Toutefois il est moins précis que les deux autres estimateurs présentés. L'estimateur de Nelson Aalen a pour but d'estimer la fonction de hasard cumulée. Or dans notre étude notre variable d'intérêt sera la fonction de survie. Le décompte des taux de sortie utilisé dans cet estimateur est le même que celui utilisé pour l'estimateur de Kaplan Meier. L'estimateur de Flemington Harring permet à partir de la fonction de hasard cumulée de Nelson Aalen d'obtenir la fonction de survie. Ainsi puisque nous nous intéressons à la fonction de survie, autant immédiatement utiliser l'estimateur de Kaplan Meier, qui permet d'obtenir cette dernière directement. Néanmoins, cet estimateur présente un inconvénient lié à la sous estimation de la fonction de survie dans la queue [3]. En effet, considérons la situation où aucune censure n'intervient après le dernier décès de la population étudiée. Ainsi l'estimateur de Kaplan Meier vaudra 0 pour cette période, ainsi que pour toutes les périodes suivantes. Or supposons qu'un nombre important de censures interviennent juste avant le dernier décès. Il est tout à fait envisageable que ces individus censurés auraient pu survivre après le dernier des décès et donc la fonction de survie n'aurait pas été nulle. L'estimateur de Kaplan Meier sous estime ainsi la fonction de survie dans la queue de la distribution. Une solution est de ne pas attribuer la valeur 0 à la fonction de survie après le dernier décès, mais de conserver la dernière valeur. Nous procéderons ainsi dans la suite.

L'estimation paramétrique suppose que la force de mortalité varie avec le temps. Elle présente l'avantage de la flexibilité : à condition de pouvoir spécifier un modèle et un certain nombre de paramètres, l'estimation de la fonction de survie est aisée.

3. John P. Klein, Prem Goel

Toutefois le problème de ces modèles réside dans cette condition. Il est en effet difficile de justifier de manière scientifique que le modèle spécifié est adapté. Par ailleurs, en présence de censure, l'estimation des paramètres d'un modèle paramétrique est compliqué.

Ainsi, dans les estimations que nous aurons à effectuer, nous utiliserons l'estimateur de Kaplan Meier pour sa fiabilité lors de la détermination de la fonction de survie pour des effectifs peu importants.

Grâce à cette technique, nous pourrons notamment estimer les probabilités de survie, de décès ou d'entrée en dépendance. Néanmoins, l'estimation de ces probabilités grâce aux méthodes non paramétriques présente bien souvent des irrégularités. Nous verrons dans la partie suivante comment interpoler ces données lorsque l'estimation pour certains âges n'a pas été possible, puis comment lisser ces taux.

3.2 Lissage et interpolation des données brutes

3.2.1 Interpolation

Lors de l'estimation des taux, l'estimation par la méthode de Kaplan Meier nous permet uniquement d'obtenir des taux d'incidences pour des intervalles de temps durant lesquels une sortie est survenue. Ainsi, en l'absence d'une exhaustivité totale des données, nous ne disposons pas des taux pour chaque âge ou chaque intervalle de temps considéré. Il nous faut donc essayer d'estimer ces données.

Spline cubique

L'idée de l'interpolation sur les splines est basée sur l'utilisation par les ingénieurs d'un outil pour tracer des courbes lisses entre un certain nombre de points. Le principe mathématique de l'interpolation par spline est le même.
Le terme cubique est employé, puisqu'il s'agit de l'ajustement d'un polynôme de degré 3. Nous considérons une fonction par morceau du type :

$$S(x) \begin{cases} \mathrm{s}_1(x) \text{ si } x_1 \leq x < x_2 \\ \mathrm{s}_2(x) \text{ si } x_2 \leq x < x_3 \\ \vdots \\ \mathrm{s}_{n-1}(x) \text{ si } x_{n-1} \leq x < x_n \end{cases}$$

où $s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i, \forall$ i = 1,2,...,n-1
x_i sont les points de passage de l'interpolation
et où a_i, b_i, c_i, et d_i sont des poids.

L'interpolation par spline repose sur 4 hypothèses :

1. La fonction par morceau S(x) doit interpoler tous les points x_i tel que i=1,...,n. Notons alors $S(x_i) = y_i$.
2. S(x) doit être continue sur l'intervalle $[x_1, x_n]$;
3. S'(x) doit être continue sur l'intervalle $[x_1, x_n]$;
4. S"(x) doit être continue sur l'intervalle $[x_1, x_n]$;

Après résolution du problème sous ces contraintes, nous obtenons :

$$\begin{cases} a_i = \frac{s''_{i+1}(x_{i+1}) - s''_i(x_i)}{6(x_i - x_{i-1})} \\ b_i = \frac{s''_i(x_i)}{2} \\ c_i = \frac{S(x_{i+1}) - S(x_i)}{x_i - x_{i-1}} - \frac{s''_{i+1}(x_{i+1}) + 2(s''_i(x_i))}{6}(x_i - x_{i-1}) \\ d_i = S(x_i) \end{cases}$$

Notons que ce système ne permet pas d'estimer les polynômes pour les deux extrémités. Il faut dont fixer deux autres conditions pour que ce système puisse être résolu.

Par exemple, pour l'interpolation par spline cubique naturelle, les dérivées secondes sont fixées, égales à 0 pour le premier et le dernier sous intervalle :

$$s''_n(x_n) = s''_1(x_1) = 0$$

Cette hypothèse indique qu'il n'y a plus d'accélération de la courbure aux extrémités d'où le terme naturel, contrairement aux splines circulaires par exemple où une courbe est attendue et où il n'y a plus aucun intervalle en bord.

Nous ne présentons pas ici l'interpolation linéaire ou "spline linéaire", représentant l'alternative la plus simple en matière d'interpolation. Cette méthode consiste en utilisant une relation affine de la forme f(x) = ax+b passant par les points à déterminer.

L'interpolation par spline présente l'avantage de non seulement passer par les points à interpoler, mais aussi de tenir compte de la courbure des données. Nous utiliserons ainsi principalement les splines cubiques lorsque nous aurons besoin d'interpoler des données.

Les données estimées présentent bien souvent des irrégularités aléatoires. Plusieurs techniques permettent de réduire ou d'annuler l'effet de ces facteurs aléatoires. C'est le but du lissage. Deux familles de lissages seront présentées par la suite. Nous

présenterons quelques exemples de lissage non paramétrique puis nous exposerons des exemples de lissage paramétrique.

3.2.2 Lissage non paramétrique

Dans un premier temps nous présenterons le lissage par moyenne mobile. Ces techniques permettent de mettre en évidence la tendance sous-jacente de la série de données estimées en extrayant, soit les composantes saisonnières, soit les résidus aléatoires. Dans un second temps, nous soumettrons une analyse de la méthode de lissage proposée par Whittaker.

Il s'agit de méthodes de lissage dont l'usage est très répandu et simple à mettre en place.

Lissage par moyenne mobile

L'idée est de résumer l'information passée par le biais de sets de données plus petits. Des moyennes sur des sous groupes de données sont réalisées. Notons $X_1, ..., X_N$ un échantillon. Soit $\tilde{X_1}, ..., \tilde{X_{N'}}$ la série lissée.
Dans le cas d'une moyenne mobile simple de période m, nous observons :

$$\tilde{X}_t = \frac{1}{2 \cdot m + 1} \cdot \sum_{k=-m}^{m} y_{t+k}$$

Il s'agit donc d'une moyenne sur $2 \cdot m + 1$ valeurs qui permet d'estimer un premier point à la m ème valeur de la série et qui "glisse" jusqu'à la N-m ème valeur de la série. Notons déjà une première limite : l'utilisation ne permet par de considérer les m premières valeurs de la série. Cette méthode a aussi l'inconvénient, comme toute méthode de moyenne, d'être sensible aux valeurs extrêmes.

Considérons maintenant le lissage par moyenne mobile exponentiel, qui présente l'avantage d'appliquer un poids aux données anciennes qui décroît exponentiellement. La formule peut être obtenue de manière récursive en prenant

$$\tilde{X}_1 = X_1$$

et

$$\tilde{X}_t = \alpha \cdot X_{t-1} + (1 - \alpha) \cdot \tilde{X}_{t-1}$$

où α représente la vitesse avec lequel le poids diminue avec $\alpha \in [0, 1]$
Un autre avantage de cette technique est qu'elle permet de prendre en compte la première valeur.

Lissage par la méthode de Whittaker Henderson

Cette méthode de lissage a été introduite par E. T. Whittaker en 1923. Henderson en 1924 et 1925 a également participé à cette théorie, surtout en montrant comment elle pouvait être mise en pratique. Cette méthode de lissage consiste à rechercher le meilleur compromis entre la fidélité aux données brutes mesurées par la somme des écarts quadratiques, F, et la régularité estimée par les écarts des différences énièmes, S (en pratique, nous nous limitons aux différences premières et secondes) sur les taux lissés.

Nous avons un tableau $(X_i, h_i)_{i=1,...,n}$ et le but est de trouver $(X_i, g_i)_{i=1,...,n}$ telle que g_i soit fidèle à h_i et (X_i, g_i) soit lisse où X_i représente par exemple l'âge et h_i la probabilité de décès à l'âge X_i.

1. Critère de fidélité aux données brutes : le critère des moindres carrés est utilisé.
   Le but est de mininiser F(g), la somme des écarts aux carrés entre h_i et g_i pondérés par un poids w_i. Nous avons $F(g) = \sum_{i=1}^{n} w_i(h_i - g_i)^2$

2. (ii) Critère de lissage (smoothness) :
   Soit $S_k(g) = \sum_{i=1}^{n-k} (\Delta^k g)^2$
   où $(\Delta g)_i = \frac{g(X_{i+1})-g(X_i)}{X_{i+1}-X_i}$
   et $(\Delta^{k+1} g)_i = \frac{(\Delta^k g)_{i+1}-(\Delta^k g)_i}{X_{i+1}-X_i} = \Delta((\Delta^k g)_i)$

Le problème est alors de trouver, avec α mesurant la relation entre fidélité et lissage, $\bar{g} = (g_1, ..., g_n)$ tel que $M = F(\bar{g}) + \alpha S_k(\bar{g})$ soit minimum. Soit $\bar{g}$ tel que :

$$\frac{\partial M}{\partial \bar{g}} = 0$$

Le problème peut être écris sous forme matricielle :

$$M = (\bar{h} - \bar{g})^{tr} W (\bar{h} - \bar{g}) + \bar{g}^{tr} (\alpha D^{(k)tr} D^{(k)}) \bar{g}$$

avec

$$\mathrm{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{pmatrix}$$

$$\bar{g} = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix}$$

$$\bar{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

$$\text{et } D^{(k)} = \begin{pmatrix} \Delta^k g_1 \\ \vdots \\ \Delta^k g_{n-z} \end{pmatrix}$$

Alors :

$$\begin{aligned}(\bar{h}-\bar{g})^{tr}W(\bar{h}-\bar{g}) + \bar{g}^{tr}(\alpha D^{(k)tr}D^{(k)})\bar{g} &= \bar{h}^{tr}W\bar{h} - 2\bar{g}^{tr}W\bar{h} + \bar{g}^{tr}(W+\alpha D^{(k)tr}D^{(k)})\bar{g} \\ &= (\bar{g}-\bar{X})^{tr}(W+\alpha D^{(k)tr}D^{(k)})(\bar{g}-\bar{X}) \\ &+ \textit{terme independant de } \bar{g}\end{aligned}$$

avec $\bar{X}$ devant vérifier :

$$\begin{aligned}&-2\bar{g}^{tr}(W+\alpha D^{(k)tr}D^{(k)})\bar{X} = -2\bar{g}^{tr}W\bar{h} \\ &\Leftrightarrow (W+\alpha D^{(k)tr}D^{(k)})\bar{X} = W\bar{h}\end{aligned}$$

d'où $\bar{g^*} = \bar{X} = (W+\alpha D^{(k)tr}D^{(k)})^{-1}W\bar{h}$
(solution de Gauss-Seidel)

3.2.3 Lissage paramétrique

Cette méthode est basée sur l'hypothèse que la mortalité (s'agissant des taux instantanés ou des taux de décès annuels) est fonction de l'âge. L'objectif étant d'obtenir le meilleur ajustement possible avec les taux à lisser avec le moins de paramètres possibles, il faut donc concilier qualité de l'ajustement et nombre de paramètres. Le risque serait de spécifier un trop grand nombre de paramètres, ce qui rendrait le modèle parfaitement adapté aux données étudiées mais dénué de toute qualité d'estimation de la mortalité d'une autre population pourtant proche.

Lissage de Gompertz-Makeham

La fonction de type Gompertz-Makeham de la forme :

$$GM_\alpha(x) = \alpha + \beta \cdot \gamma^x = \mu(x)$$

nous avions montré que :

$$S(t) = \exp(-\alpha t - \frac{\beta}{log(\gamma)}(\gamma^t - 1))$$

La détermination peut être effectuée par différentes méthodes. Renshaw, principalement, a proposé la détermination des r+s paramètres de cette fonction par la méthode des modèles linéaires généralisés. L'utilisation de cette technique se justifie parfaitement par l'absence de normalité des taux de mortalité.

Rappel sur la régression linéaire simple Faire de la régression linéaire de y sur x, c'est s'intéresser à la dépendance entre y et x. y peut dépendre de x, mais il peut aussi dépendre d'une infinité d'autres facteurs. Le modèle sous-jacent à la régression simple est :

$$Y = f(x) + \varepsilon$$

où :

Y est la variable à expliquer (par exemple la mortalité d'un individu)
x la variable explicative ou le prédicteur (par exemple l'âge de l'individu)
ε est l'erreur commise dans l'approximation de Y par f(x), aussi appelé résidu

En général nous disposons d'un certain nombre n d'observations :

$$Y_i = f(x_i) + \varepsilon_i$$

avec $i = 1, ..., n$

Quand nous parlons de régression linéaire, cela revient à supposer f affine, que chaque ε_i est indépendant et identiquement distribué $\forall i$ (i.i.d.), centré et de même variance σ^2 (hypothèse d'homoscedasticité, i.e. toutes les mesures ont été faites indépendamment des autres, dans les mêmes conditions)

Si nous considérons le modèle $Y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$ avec $i = 1, ...n$ et $\mathbb{E}\varepsilon_i = 0$, $\mathbb{V}\varepsilon_i = \sigma^2$ et ε_i i.i.d., sans pour autant supposer la loi de ε_i, le problème est alors d'estimer les paramètres β_0 et β_1, déterminer σ^2, créer des intervalles de confiances pour Y_i et valider la modélisation.

Si la linéarité entre les paramètres est supposée, nous utilisons la méthode des moindres carrés et nous cherchons à minimiser la somme des écarts quadratiques entre Y_i et $\beta_0 + \beta_1 \cdot x_i$. Nous cherchons donc à minimiser :

$$\sum_{i=1}^{n}(Y_i - \beta_0 + \beta_1 \cdot x_i)$$

ce qui revient à minimiser la grandeur que nous noterons δ_n^2 :

$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta_0 + \beta_1 \cdot x_i)$$

En annulant successivement les dérivés par β_0 et β_1 de δ_n^2, des estimateurs de β_0 et β_1 sont alors trouvés :

$$\hat{\beta}_0 = \overline{Y} - \beta_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i Y_i - \overline{Y}\overline{x}}{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - (\overline{x})^2} = \frac{Covariance_{empirique}(x, Y)}{Variance_{empirique}(x)} = \frac{C_{x,Y}}{S_x^2}$$

Ainsi l'équation de la droite des moindres carrés est obtenue :

$$\overline{Y} + \frac{C_{x,Y}}{S_x^2}(x_i - \overline{x})$$

qui est la version empirique de la meilleure approximation affine de Y.

En considérant le modèle précédent, $Y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i$, avec $\mathbb{E}(\varepsilon_i) = 0$ et $\mathbb{V}(\varepsilon_i) = \varepsilon_i$, d'après le théorème de Gauss-Markov, $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des estimateurs sans biais à variance minimale parmi tous les estimateurs sans biais linéaires.

Intéressons nous maintenant à l'estimation de la variance des résidus :

$$\sigma^2 = \mathbb{V}(Y_i - \beta_0 - \beta_1 x_i)$$

Soient

$$E_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

les résidus empiriques, nous avons le résultat suivant :

$$\hat{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(E_i - \overline{E_n})^2}{n-2}$$

où $\overline{E_n}$ est la moyenne empirique des résidus empiriques.

Tout ceci a été effectué sans hypothèses sur la loi des ε_i. Cette régression permet l'estimation de valeurs inconnues (e.g. la mortalité lissée) sous forme de combinaison linéaire d'un prédicteur - par exemple, l'âge. Cette linéarité implique qu'un certain changement du prédicteur entraîne un certain changement de la variable estimée de la même ampleur. Or ce type de réponse de la variable au prédicteur n'est valable que dans certains cas. La réponse au prédicteur peut dans certains cas être plutôt exponentielle ou de type logistique (la réponse attendue ne peut être que 0 ou 1).

Rappel sur les modèles linéaires généralisés L'intérêt des modèles linéaires généralisés est, précisément de généraliser la prédiction au cas où la variable réponse suit une distribution quelconque. Le recours à une fonction de lien entre la variable réponse et le prédicteur, afin que la réponse varie linéairement avec le prédicteur, n'est donc plus indispensable.

Sous l'hypothèse que $Y_1, ..., Y_n$ sont i.i.d. de densité f, nous obtenons, par maximisation de la log vraisemblance de f :

$$\mathbb{E}Y = b'(\theta)$$

$$\mathbb{V}Y = b''(\theta)a(\phi)$$

Dans ce cas, l'hypothèse est faite que la loi de Y est dans la famille exponentielle, i.e. la densité de Y peut être écrite sous la forme

$$f(y, \theta, \phi) = exp(\frac{y\theta - b(\theta)}{a(\phi)} - c(y, \phi))$$

Par exemple dans le cas de la loi Binomiale de paramètre n,p où $\mathbb{P}(Y = y) = C_n^y p^y (1-p)^{n-y}$, $\mathbb{P}(Y = y)$ peut être réécris sous la forme :

$$exp(ylog(\frac{p}{1-p}) + nlog(1-p) + log(C_n^y))$$

où peut être identifier :

$$exp(y\underbrace{log(\frac{p}{1-p})}_{\theta} + \underbrace{nlog(1-p)}_{\text{-b}(\theta)} + \underbrace{log(C_n^y)}_{\text{c(y,}\phi\text{)}})$$

Nous avons alors :

$$\theta = log(\frac{p}{1-p}) \Leftrightarrow \frac{p}{1-p} = exp(\theta) \Leftrightarrow p = \frac{exp(\theta)}{1 - exp(\theta)}$$

$$b(\theta) = -nlog(1-p) = nlog(1 + exp(\theta))$$

$$a(\phi) = 1$$

Il faut aussi définir le prédicteur linéaire

$$\eta_i = x_i'\beta$$

où

$$x_i' = (x_{i_0}, ..., x_{i_p})$$

sont les variables exogènes associées à l'individu i et

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

le vecteur des paramètres inconnus à estimer.

Définissons aussi la fonction de lien g :

$$g(\mu_i) = \eta_i = x_i'\beta$$

$$\Leftrightarrow \mu_i = g^{\leftarrow}(x_i'\beta)$$

Ensuite, β et ϕ sont estimés par maximum de vraisemblance. L'objectif est donc d'annuler la dérivée de la log vraisemblance de f en les β et en ϕ. La résolution se fait ensuite par la méthode de Newton Raphson.

Nous obtenons :

- Soit $u(\beta)$ le vecteur gradient dont l'élément en position j est égal à :

$$\frac{\delta}{\delta\beta_j} log(L)$$

- Soit $\mathscr{H}(\beta)$ la matrice Hessienne dont l'élément en position (j,k) est égal à : $\frac{\delta^2}{\delta\beta_j\delta\beta_k} log(L)$

Par définition du maximum de vraisemblance :

$$\begin{aligned} u(\hat{\beta}) &= 0 \\ &\approx u(\beta^*) + \mathscr{H}(\beta^*)(\hat{\beta} - \beta^*) \text{ par développement limité} \end{aligned}$$

$$\Rightarrow \hat{\beta} = \beta^* - \mathscr{H}^{\leftarrow}(\beta^*)u(\beta^*)$$

Avec la méthode de Newton Raphson, nous partons d'une valeur $\hat{\beta}^{(0)}$ et nous obtenons itérativement :

$$\hat{\beta}^{(r+1)} = \hat{\beta}^{(r)} - \mathscr{H}^{\leftarrow}(\hat{\beta}^{(r)})u(\hat{\beta}^{(r)})$$

Retour au lissage de Gompertz Makeham Supposons que E_x soit l'effectif sous risque l'année x, c'est à dire le nombre de personnes présente dans l'étude à l'âge x et étant soumis, par exemple, au risque de décès entre les âges x et x+1 avec

la probabilité de décès égale à q_x. Supposons aussi que la probabilité de décès de chaque individu compris dans la population E_x est indépendant de la probabilité de décès des autres individus de cette population. Alors le nombre de décès entre les âges x et x+1, D_x, est une variable aléatoire suivant une loi Binomiale. Nous avons

$$D_x \sim Bin(E_i, q_x)$$

Soit

$$q_x = \frac{GM_\alpha^{r,s}}{1 + GM_\alpha^{r,s}}$$

en utilisant la transformation logit comme fonction de lien pour q_x, nous trouvons que :

$$\frac{q_x}{1 - q_x} = GM_\alpha^{r,s}(x)$$

c'est à dire que $\frac{q_x}{1-q_x}$ correspond au prédicteur linéaire.

En résumé, le lissage par moyenne mobile présente l'avantage d'être simple à utiliser, mais présente l'inconvénient d'être sensible aux valeurs extrêmes, auxquelles nous serons confrontées.

Le lissage par la méthode de Whittaker Henderson quant à lui présente une bonne alternative : en effectuant un arbitrage entre fidélité aux données et régularité, cette méthode permet d'éliminer d'éventuelles fluctuations injustifiées des taux, tout en conservant une proximité avec les taux non lissés. Nous utiliserons préférentiellement ce type de lissage, avec un paramètre lambda permettant d'obtenir un lissage suffisant. Le choix de ce paramètre est arbitraire, dans la suite de l'étude, nous le choisirons le plus faible possible afin d'obtenir un lissage, tout en étant fidèle aux données.

Un problème non résolu par le lissage demeure : bien souvent, nous ne disposons pas d'observations de personnes très âgées. Toutefois, il n'est pas impossible que de telles personnes soient présentes dans le portefeuille à l'avenir. Il faut donc tenter d'extrapoler les taux obtenus aux grands âges, grâce notamment aux techniques présentées dans la partie suivante.

3.3 Fermeture des tables

Les techniques de fermeture de table permettent d'extrapoler les taux de mortalité ou d'incidence aux âges élevés. Dans la suite nous présenterons deux de ces

méthodes, à savoir la méthode de Coaler et Kisker et la méthode de Denuit et Goderniaux.

Nous présentons ces deux méthodes car elles présentent une évolution des taux de mortalité future différente : l'une estimant des taux de mortalité inférieure à l'autre pour les âges élevés.

3.3.1 Méthode de Coale et Kisker

Le modèle de Coale et Kisker [4] admet l'hypothèse que l'exponentiel du taux de mortalité aux âges très élevés diminue de manière linéaire avec l'âge. Ce modèle comporte deux paramètres, k et le taux de mortalité pour l'âge de fermeture de la table (μ_{110} dans l'article, $\mu_{110} = 1$ pour les hommes et $\mu_{110} = 0.8$ pour les femmes pour éviter le croisement de la mortalité homme femme aux âges élevés). Le but de ce modèle est de prédire le taux de mortalité après 85 ans connaissant le taux de mortalité autour de 85 ans.

Soit :

$$k_x = ln(\frac{\mu_x}{\mu_{x-1}})$$

(avec par exemple dans l'hypothèse de linéarité dans les sous intervalles,

$$\mu_x = -ln(1 - q_x)$$

)

Dans l'hypothèse où μ_{65} est connu, alors :

$$\mu_x = \mu_{65}\ exp(\sum_{y=66}^{x} k_y)\ \forall\ x = 66, 67, ...$$

Puisque dans ce modèle l'hypothèse est faite que k_x est linéaire pour les âges supérieurs à 85 ans, nous prenons

$$k_x = k_{85} + (x - 85)s$$

avec s désignant la pente de k_x pour ces âges. Cela implique que :

$$\mu_x = \mu_{84} \exp(\sum_{y=85}^{x} k_{85} + (y - 85)s)$$

Si nous admettons que x=110, alors nous obtenons :

$$s = -\frac{ln(\frac{\mu_{84}}{\mu_{110}}) + 26k_{85}}{325}$$

4. Coale and Kisker (1990)

Nous avons alors :

$$k_x = k_{85} - (x-85)\frac{ln(\frac{\mu_{84}}{\mu_{110}}) + 26k_{85}}{325}$$

et comme

$$\mu_x = \mu_{x-1}\exp(k_x)$$

nous obtenons :

$$\mu_x = \mu_{x-1}\exp(k_{85} - (x-85)\frac{ln(\frac{\mu_{84}}{\mu_{110}}) + 26k_{85}}{325})$$

Pour éviter les fluctuations, au lieu d'utiliser μ_{84}, Coale et Kisker proposèrent d'utiliser une moyenne arithmétique de 81 à 88 ans. Notons alors :

$$\mu_{84}^{'} = \frac{ln(\frac{\mu_{88}}{\mu_{81}})}{7}$$

3.3.2 Méthode de Denuit et Goderniaux

La méthode de Denuit et Goderniaux[5] consiste à ajuster par la méthode des moindres carré le modèle log-quadratique suivant :
$\ln \hat{q}_x = a + bx + cx^2 + \epsilon_x$ avec ϵ_x iid suivant une loi Normale N(0,σ^2)
En faisant l'hypothèse que $q_{130} = 1$ et $q_{130}^{'} = 0$ afin d'obtenir une mortalité aux âges élevés concave.
Ces deux hypothèses impliquent que :

$$a + bx + cx^2 = c(130^2 - 260x + x^2)$$

Afin de choisir l'âge de raccord, le coefficient de détermination est maximisé sous contrainte que cet âge de raccord soit compris entre 50 et 85 ans. Le coefficient de détermination, noté R^2 est égale à :

$$R^2 = 1 - \frac{\sum_x (q_x - \tilde{q}_x)^2}{\sum_x (q_x - \frac{\sum_x q_x}{n})^2}$$

où n est le nombre d'observations, q_x sont les taux sur lesquels sont estimés les paramètres du modèle de Denuit et Goderniaux, $\tilde{q}_x$ les taux estimés par le modèle, avant l'âge de clôture.

Pour Quashie et Denuit (2005), cette dernière méthode semble plus performante que la première dans l'estimation d'engagements viagers. Il semble alors prudent d'utiliser la méthode de Coale et Kisker pour l'évaluation d'engagements liés aux

5. Denuit et Goderniaux (2005)

décès et la méthode de Denuit et Goderniaux pour évaluer des engagements viagers.

La méthode de Coale et Kisker est basée sur l'hypothèse de diminution linéaire avec l'âge de l'exponentiel du taux de mortalité des individus très âgés. La méthode de Denuit et Goderniaux quant à elle suppose une mortalité aux âges élevés concave. Dans les deux cas, il s'agit d'hypothèses assez difficile à valider empiriquement car peu d'observations aux âges élevés sont disponibles. Il s'agit de modèles dépendant de l'opinion de l'utilisateur quant à l'évolution de la mortalité future.

Par manque de données, il est parfois impossible d'estimer les probabilités désirées par les méthodes présentées jusqu'ici. Toutefois, il est possible d'ajuster des probabilités connues et fiables, afin qu'elles prennent en compte la réalité du portefeuille. Nous présenterons deux de ces techniques dans la partie suivante.

3.4 Abattement par expérience

Une autre idée pour rendre la modélisation plus proche de la réalité est d'abattre la mortalité estimée par des tables de mortalité nationales, par exemple les tables de l'INSEE. Dans cette partie, nous verrons deux méthodes permettant de réaliser de tels ajustements sur les tables de mortalité. La première méthode, l'approche SMR consiste à calculer d'un taux permettant d'abattre la table de référence. La deuxième, l'approche relationnelle de type Brass, permet d'ajuster la table de référence via une relation affine des logits des taux.

3.4.1 Approche SMR

Définissons le Ratio de Mortalité Standardisé [6] (SMR) qui est le ratio entre le nombre de décès observés dans une population durant une certaine période d'observation, et le nombre de décès attendus dans une telle population, tels qu'estimés par une référence externe.

Soit $Do_{x,t}$ le nombre de décès observés à l'âge x et à la période t,
$Da_{x,t}$ le nombre de décès attendus à l'âge x et à la période t,
$E_{x,t}$ le nombre de personnes sous risque à l'âge x à l'instant t,
Soit q_x^{ref} la probabilité de décès entre les âges x et x+1 obtenu grâce à une référence externe.

Considérons le cas sur une seule période, nous avons :

$$SMR = \frac{\sum\limits_{x=x^*} Do_x}{\sum\limits_{x=x^*} Da_x} = \frac{\sum\limits_{x=x^*} Do_x}{\sum\limits_{x=x^*} E_x \cdot q_x^{ref}}$$

6. Liddel (1984)

Nous obtenons les probabilités réduites (ou majorées), $\tilde{q}_x$, en multipliant les probabilités de référence q_x^{ref} par le coefficient SMR.

Nous avons :

$$\tilde{q}_x = SMR \cdot q_x^{ref}$$

Le choix de x^* va être important, de plus l'ajustement avec le SMR ne sera valide qu'à partir de x^*.

3.4.2 Méthode semi paramétrique avec utilisation d'un modèle relationnel de type Brass

Dans ce modèle, l'hypothèse est faite que le logit des taux propres à la compagnie peuvent évoluer de manière linéaire avec le logit des taux d'une référence externe, à une constante près. L'utilisation du logit est nécessaire pour effectuer une régression sur des taux de mortalité.

Nous avons donc le modèle :

$$logit(q_{x^*}) = \alpha + \beta \cdot logit(q_{x^*}^{ref})$$

avec α et β étant deux paramètres, l'un affectant tous les âges de la même façon, l'autre ayant un impact différent selon l'âge.

En effet, les taux de mortalité étant compris entre 0 et 1, il n'est pas possible d'ajuster ces valeurs, puisque qu'aucune droite obtenue par régression ne peut être bornée entre 0 et 1 (cf. Illustration 3.2). Le logit permet d'obtenir des valeurs comprises entre $]-\infty, +\infty[$, plus susceptible d'être ajusté par une droite.

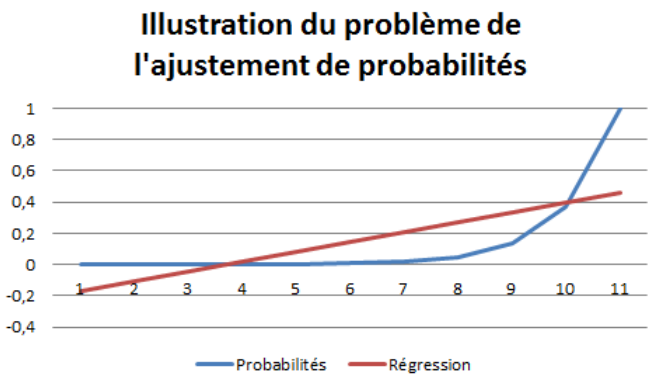

Illustration 3.2 : *Problème de la régression sur des probabilités*

La solution est obtenue en minimisant :

$$\sum_{x=x^*} |E_x(\hat{q}_x - \tilde{q}_x)|$$

avec $\hat{q}_x$ la probabilité de décès observée (i.e $\hat{q}_x = \frac{Do_x}{N_x}$) et $\tilde{q}_x$ la probabilité de décès estimée.

Puisque :

$$logit(\tilde{q}_{x^*}) = \hat{\alpha} + \hat{\beta} \cdot logit(q_{x^*}^{ref}) \Leftrightarrow log(\frac{\tilde{q}_{x^*}}{1 - \tilde{q}_{x^*}}) = \hat{\alpha} + \hat{\beta} \cdot log(\frac{q_{x^*}^{ref}}{1 - q_{x^*}^{ref}})$$

$$\Rightarrow \frac{\tilde{q}_{x^*}}{1-\tilde{q}_{x^*}} = \exp(\hat{\alpha} + \hat{\beta} \cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}}))$$

$$\Rightarrow \tilde{q}_{x^*} = \frac{\exp(\hat{\alpha}+\hat{\beta}\cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}}))}{1+exp(\hat{\alpha}+\hat{\beta}\cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}}))}$$

$$\Rightarrow \tilde{q}_{x^*} = \frac{1}{1+\exp(-(\hat{\alpha}+\hat{\beta}\cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}})))}$$

Nous cherchons donc à minimiser :

$$\sum_{x=x^*} |E_x(\frac{D_x}{E_x} - \frac{1}{1 + \exp(-(\hat{\alpha} + \hat{\beta} \cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}})))})|$$

en α et β.

En considérant les modèles précédents, nous obtenons alors de nouvelles estimations s'approchant le plus possible de la réalité. Le ratio SMR présente l'avantage d'être facile à mettre en place et d'être facilement interprétable. Néanmoins, il n'est pas certain qu'un seul ratio SMR reflète complètement la sur ou sous mortalité de la population étudiée comparée à la population de référence. Il se peut que le ratio SMR surestime la mortalité sur un certain intervalle et la sousestime pour d'autre intervalle. L'ajustement semi paramétrique est plus fin lors de la prise en compte de la surmortalité ou sous mortalité d'une population par rapport à une autre en la caractérisant par un facteur non dépendant de l'âge, par exemple les accidents, et un facteur lié à l'âge, par exemple la santé.

La partie suivante analyse les différentes méthodes de validation de ces ajustements.

3.5 Validation des estimations

Une fois la table créée, il nous faut un critère objectif afin de pouvoir juger si l'ajustement effectué est acceptable ou non. Il faut ainsi définir comment est définie l'acceptabilité d'une estimation, paramétrique ou non paramétrique, ou un lissage. Nous nous appuierons principalement sur la distance entre l'estimation des taux bruts et l'ajustement effectué de ces taux.

3.5.1 Analyse de la qualité de l'ajustement

Objectif

L'objectif est de savoir si l'estimation faite est acceptable, c'est à dire si l'écart entre l'estimation et la réalité n'est pas trop élevé.

Nous reprenons la validation d'un modèle tel que proposé par Planchet et Tomas dans une note technique [7] pour l'Institut des Actuaires. Dans ces travaux, nous une méthode de validation d'un modèle sur trois niveaux est exposée. Un premier niveau de proximité entre les observations et le modèle, un deuxième de régularité des taux et un troisième de plausibilité de ces taux. Nous ne reprendrons pas l'ensemble des tests proposés.

Validation de l'ajustement du modèle aux données

Les trois tests suivants permettent de juger de la qualité globale de l'ajustement.

Mesure du χ^2 Utilisé dans Forfar et al (1988), ce test permet de mesurer la qualité d'adéquation du modèle.

$$\chi^2 = \sum_{(x,t)} \frac{(D_x - E_x * \tilde{q}_x)^2}{E_x \tilde{q}_x (1 - \tilde{q}_x)}$$

Le but étant d'avoir un χ^2 minimum.

Mean Average Percentage Error Le MAPE [8] est une mesure de la précision de l'ajustement du modèle aux données. C'est la moyenne des valeurs absolues des déviations par rapport aux observations.

$$MAPE = 100 \cdot \frac{\sum_x |\frac{\frac{D_x}{E_x} - \tilde{q}_x}{\frac{D_x}{E_x}}|}{\sum_x D_x}$$

7. Note de travail II1291-11 v1.4
8. Felipe et al.(2002)

SMR Nous définissons le SMR, cette fois-ci à titre de test afin de déterminer si le modèle surestime ou sous-estime la mortalité.

$$SMR = \frac{\sum_x D_x}{\sum_x E_x \cdot \tilde{q}_x}$$

Si le SMR est plus grand que 1, alors le modèle sous-estime le nombre de décès.

Analyse graphique

L'analyse graphique permet d'identifier des valeurs ajustées qui s'écarteraient de manière trop importante de l'estimation initiale, sans pour autant perturber les résultats aux trois tests précédents.

Résidus bruts Les résidus bruts permettent de constater visuellement si certaines valeurs de l'ajustement s'écartent des observations. Toutefois la distance n'étant pas normalisée, ces résidus ne permettent pas de juger si l'écart est important ou non autre que relativement avec les autres écarts.

$$r_x = \frac{D_x}{E_x} - \tilde{q}_x$$

Résidus de Pearson Le résidu de Pearson se calcule comme le résidu brut, rapporté à l'écart-type estimé de $E_x \cdot \tilde{q}_x$.

$$r_x^P = \frac{(D_x - E_x \cdot \tilde{q}_x)}{\sqrt{\mathbb{V}(E_x \cdot \tilde{q}_x)}}$$

Les valeurs aberrantes sont celles supérieures en valeur absolue à 2 Le choix de cette valeur 2 correspond à l'hypothèse d'approximation des résidus par une loi Normale centrée réduite.

Résidus de déviance

$$r_x^D = signe(D_x - E_x \cdot \tilde{q}_x)\sqrt{Deviance_x}$$

Avec la déviance étant une mesure de la qualité d'ajustement du modèle. Elle est obtenue en émettant l'hypothèse que $D_x \simeq P(E_x \cdot q_x)$, ce qui permet d'écrire la déviance comme suit :
Si $D_x > 0$:

$$Deviance_x = 2(D_x \cdot ln(\frac{D_x}{E_x \cdot \tilde{q}_x}) - (D_x - E_x\tilde{q}_x))$$

Si $D_x = 0$

$$Deviance_x = 2 \cdot E_x\tilde{q}_x$$

Analyse de la régularité

L'objectif est de visualiser si les données ont été trop ou pas assez lissées. Pour ce faire nous nous servons du test du signe.

Test du signe Ce test analyse la fréquence avec laquelle la différence entre les observations et le modèle change de signe.
Soit l'hypothèse H0 : la médiane entre le nombre de signes positifs et le nombre de signes négatifs est nulle.
Soit n_+ et n_- le nombre de signes respectivement positifs et négatifs, avec $n = n_+ + n_-$, nous avons :

$$\sigma^{signe} = \frac{|n_+ - n_-| - 1}{\sqrt{n}}$$

En faisant l'hypothèse - forte vu le faible nombre de données utilisées - que cette statistique suit une loi Normale, H0 est rejetée si :

$$|\sigma^{signe}| > N_{1-\frac{\alpha}{2}}(0,1)$$

où $N_{1-\frac{\alpha}{2}}(0,1)$ est le quantile au seuil $1 - \frac{\alpha}{2}$ d'une loi Normale centrée réduite.

Analyse de la cohérence

Il s'agit d'analyser si l'estimation est vraisemblable et cohérente. Nous pourrons par exemple analyser la durée de survie restante ou l'âge médian à la mort.

Dans la suite, nous serons amener à analyser la pertinence de la segmentation de différentes populations dans l'étude des différentes probabilités. Nous présenterons ainsi dans la partie suivante comment analyser la pertinence d'une telle segmentation dans le cadre de données censurées.

3.5.2 Tests d'hypothèses : différenciation de courbe de survie

Objectif

Dans le cadre des tests de rang, l'objectif est de tester deux hypothèses :

- H0 : les deux courbes de survie ont des profils identiques, le risque de sortie est le même dans les deux groupes ;
- H1 : les deux courbes de survie ont des profils différents.

Soit N la taille de l'échantillon total. En reprenant les notations précédentes, nous avons $r_{i,j}$ l'effectif sous risque en i et $s_{i,j}$ le nombre de sortie en i du groupe j.

Les individus sous risques sont estimés avant les sorties, ce qui implique que :

$$r_{i+1,j} = r_{i,j} - s_{i,j}$$

Ainsi, si l'hypothèse H0 est vérifiée, $s_{i,j}$ suit une loi hypergéométrique H(r_i, s_i, $\frac{r_{i,j}}{r_i}$), qui est bien la loi du nombre de décès dans le groupe j, r_{ij} parmi un effectif r_i, avec la proportion de $r_{i,j}$ présente au départ égale à $\frac{r_{i,j}}{r_i}$

Ainsi, nous obtenons :

$$\mathbb{E}(s_{i,j}) = s_i \frac{r_{i,j}}{r_i}$$

et

$$\mathbb{V}(s_{i,j}) = s_i \frac{r_i - s_i}{r_i - 1} \frac{r_{i,1} r_{i,2}}{r_i^2}$$

La somme des normalisations de $s_{i,j}$, asymptotiquement gaussien, peut alors être construite et pondérée par un poids w_i
Nous obtenons :

$$\Phi_j = \frac{[\sum_{i=1}^{N} w_i (s_{i,j} - s_i \frac{r_{i,j}}{r_i})]^2}{\sum_{i=1}^{N} w_i^2 s_i \frac{r_i - s_i}{r_i - 1} \frac{r_{i,1} r_{i,2}}{r_i^2}}$$

qui suit asymptotiquement une loi du Khi-2 de paramètre 1.

Test du Log Rank

Dans le cas du test du Log-Rank [9], chaque observation a le même poids : $w_i = 1 \ \forall i$.

Une formule approchée de la statistique peut être obtenue dans le cas de l'étude de deux groupes, avec $D_i^{théorique}$ le nombre de sorties dans le groupe i, déterminé avec la loi que nous cherchons à comparer et $D_i^{observé}$ le nombre de sorties réelles du groupe i :

$$\Phi = \frac{(D_1^{théorique} - D_1^{observé})^2}{D_1^{théorique}} + \frac{(D_2^{théorique} - D_2^{observé})^2}{D_2^{théorique}}$$

qui suit sous H0 une loi du Khi^2 de degré de liberté 1 (ou nombre de groupes - 1 si nous comparons plus de deux courbes)

Test de Gehan

Dans le cas du test de Gehan [10], les observations sont pondérées par l'effectif sous risque : $w_i = r_i$. Le test de Gehan accorde donc un poids plus élevé aux décès précoces qu'aux décès tardifs, puisque le poids est décroissant et donc accorde plus

9. Mantel, Nathan (1966)
10. Gehan (1965)

d'importance aux courtes durées de survie.

Le fait que le test du log rank accorde à chaque observation le même poids le rend plus puissant pour la comparaison de fonctions de survie dont le taux de décès est proportionnel. Le poids plus important accordé au décès précoce par le test de Gehan le rend utile lorsque l'effet important pour l'étude se trouve en début de distribution. Par exemple, considérons deux groupes dont l'un présente une mortalité particulièrement sévère pour les premières périodes d'observation, alors que l'autre présente une mortalité plus faible. Si toutefois la mortalité est identique dans les deux groupes par la suite, le test du log rank pourrait nous faire rejeter l'hypothèse de différenciation des courbes de survie. Néanmoins, si le phénomène qui nous intéresse est la forte surmortalité d'un des groupes survenant rapidement après le début de l'étude, alors le test de Gehan sera plus adapté car aura moins tendance à rejeter l'hypothèse de différenciation des fonctions de survie.

Section III

Applications : Détermination des lois du portefeuille

Partie 4

Étude des données

4.1 Statistiques générales

Dans cette partie, nous appliquerons aux données disponibles les méthodes d'estimation, de lissage ou d'abattement présentées précédemment. Nous étudierons ainsi le comportement des assurés du portefeuille pour lequel nous devons projeter le résultat, afin de pouvoir observer l'évolution future du portefeuille avec le plus de réalisme possible. Pour l'étude de la survie des dépendants, nous y ajoutons un deuxième portefeuille similaire, afin d'attendre un nombre d'observations de dépendants égal à 950.

Au 01/01/2014, la durée d'observation est de 14 ans et le portefeuille est toujours ouvert à la commercialisation à cette date. L'état de dépendance est défini en fonction des GIR et des AVQ.

Pour des raisons de confidentialité, nous ne communiquerons pas plus d'informations sur ces données.

Après avoir effectué un tri de cohérence dans les données dans le but de nous assurer que nous disposons de données fiables, et ce selon les techniques proposées dans la partie suivante intitulée "Vérification de la qualité des données", nous conservons pour cette étude environ 50 000 assurés autonomes (ou l'ayant été) et la totalité des assurés en dépendance. Nous commencerons pas présenter le tri des données puis étudierons les caractéristiques de celles-ci.

4.1.1 Vérification de la qualité des données

La qualité des données est primordiale lors de l'estimation de taux de mortalité ou d'incidence. La présence de doublons pourrait créer un biais dans l'estimation des taux de mortalité ou d'incidence. Les erreurs de gestion pourraient quant à elles entraîner des résultats aberrants.

Nous présentons dans la suite la nature des problèmes pouvant être rencontrés sur les bases de données et comment ils peuvent être décelés.

Doublons

Les doublons, c'est à dire la présence, au moins en double, d'un même individu dans la base de donnée peut entraîner une erreur dans l'estimation des taux de sortie. En effet, la présence répétée d'un individu, qu'il soit sorti ou encore dans le portefeuille, entraîne soit une sur-évaluation du nombre de sortie (en cas de sortie), soit une sur-évaluation de l'exposition au risque (en cas de présence dans le portefeuille).

Puisque nous disposons de bases de données relativement bien renseignées, le travail de recherche des doublons est facilité. Grâce à diverses requêtes, nous vérifions qu'il n'existe aucun assuré ayant le même nom et prénom ne soit présent dans la base de données. Toutefois nous restons vigilants sur le fait qu'il peut y avoir des doublons dont le nom n'est pas exactement orthographié de la même façon, provenant par exemple d'une erreur de saisie non corrigée. Ainsi, nous vérifions manuellement l'ensemble des assurés ayant une date de naissance identique.

En outre, nous effectuons une vérification sur les numéros de contrat, afin de vérifier qu'aucun numéro ne soit identique.

Erreurs de gestion

Diverses erreurs de gestion peuvent apparaître dans les bases de données. Par exemple, l'enchaînement des dates peut être incohérent. Nous vérifions ainsi que :

- Les dates de sortie (décès, résiliation, incidence) sont bien supérieures aux dates d'entrée dans la base de données ;
- Les dates de naissance sont bien inférieures aux dates d'entrées dans le portefeuille ;
- Les âges de souscription sont bien compris dans l'intervalle imposé par les conditions générales du produit.

Ces techniques permettent d'éliminer une grande partie des problèmes pouvant être présents sur notre base de données. L'étude statistique du portefeuille réalisée par la suite va permettre d'étudier les caractéristiques de la base de données, mais aussi de vérifier que les ordres de grandeur observés sont bien ceux communément observés sur ce type de données et donc participer à la vérification de la qualité des données.

4.1.2 Étude de la population autonome du portefeuille

Dans cette partie nous allons étudier les caractéristiques des données utilisées pour l'étude.

Âge à la souscription Comme nous l'avions vu précédemment, le problème de ce type de contrat est une adhésion tardive de la part des assurés. Sur ces données,

ce problème est bien présent, avec un âge moyen, toutes générations confondues, de 63 ans, qui correspond bien avec l'âge moyen de souscription du marché, qui est de 62 ans [1].

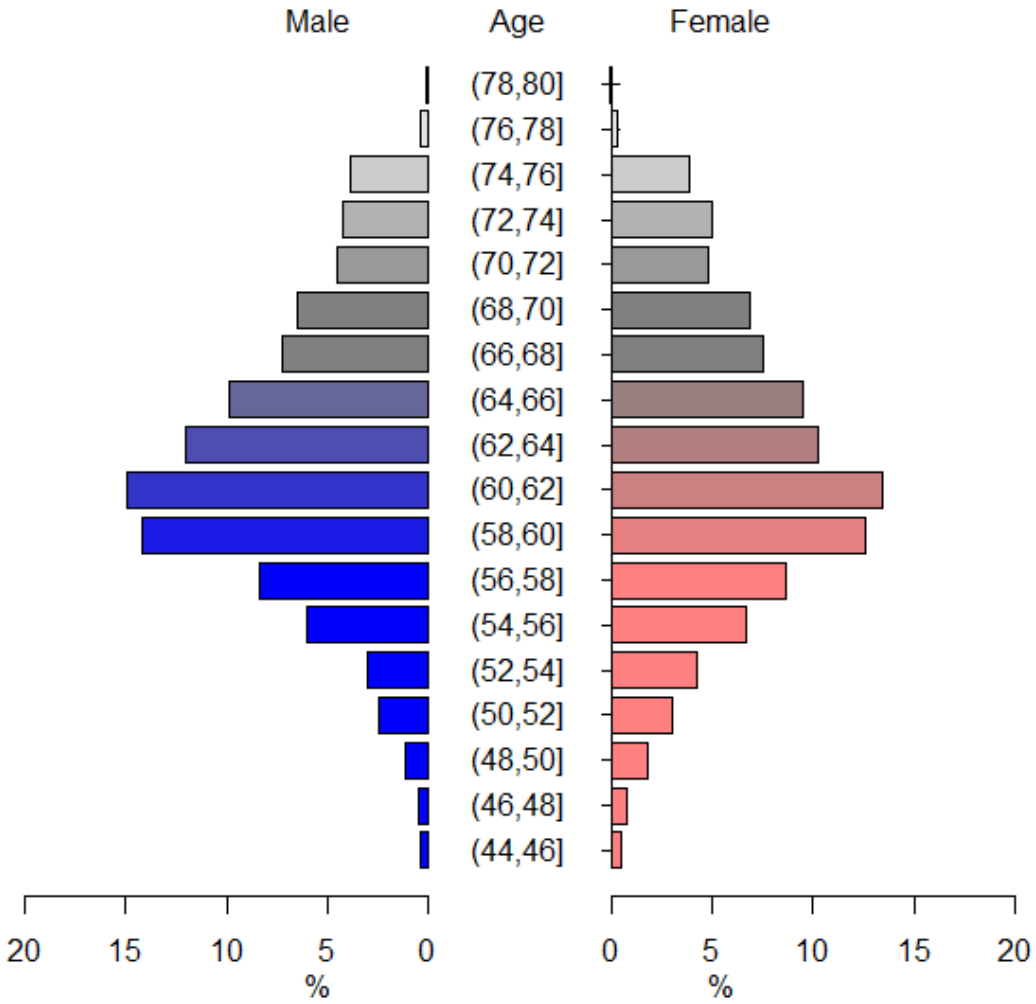

Illustration 4.1 : *Pyramide des âges à la souscription*

Nous pouvons constater une légère asymétrie à gauche de la distribution. La moyenne est proche de la médiane (cf. Illustration 4.2).

1. F.F.S.A.

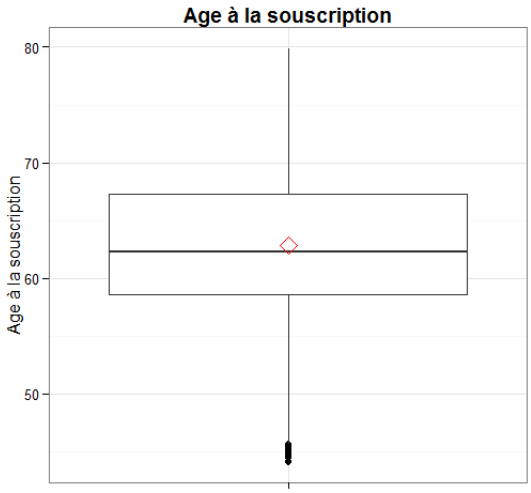

Illustration 4.2 : *Boîte à moustache de l'âge à la souscription (le losange étant la moyenne)*

Toutefois, nous constatons une évolution de cet âge de souscription au cours des différentes générations, peut être due à une progressive prise de conscience du problème de la dépendance par la population, notamment avec un âge à la souscription historiquement faible en 2003. Ce pic peut être du à la canicule, ayant fait prendre conscience à la population de la solitude et du manque d'accompagnement des personnes âgées (cf. Illustration 4.3).

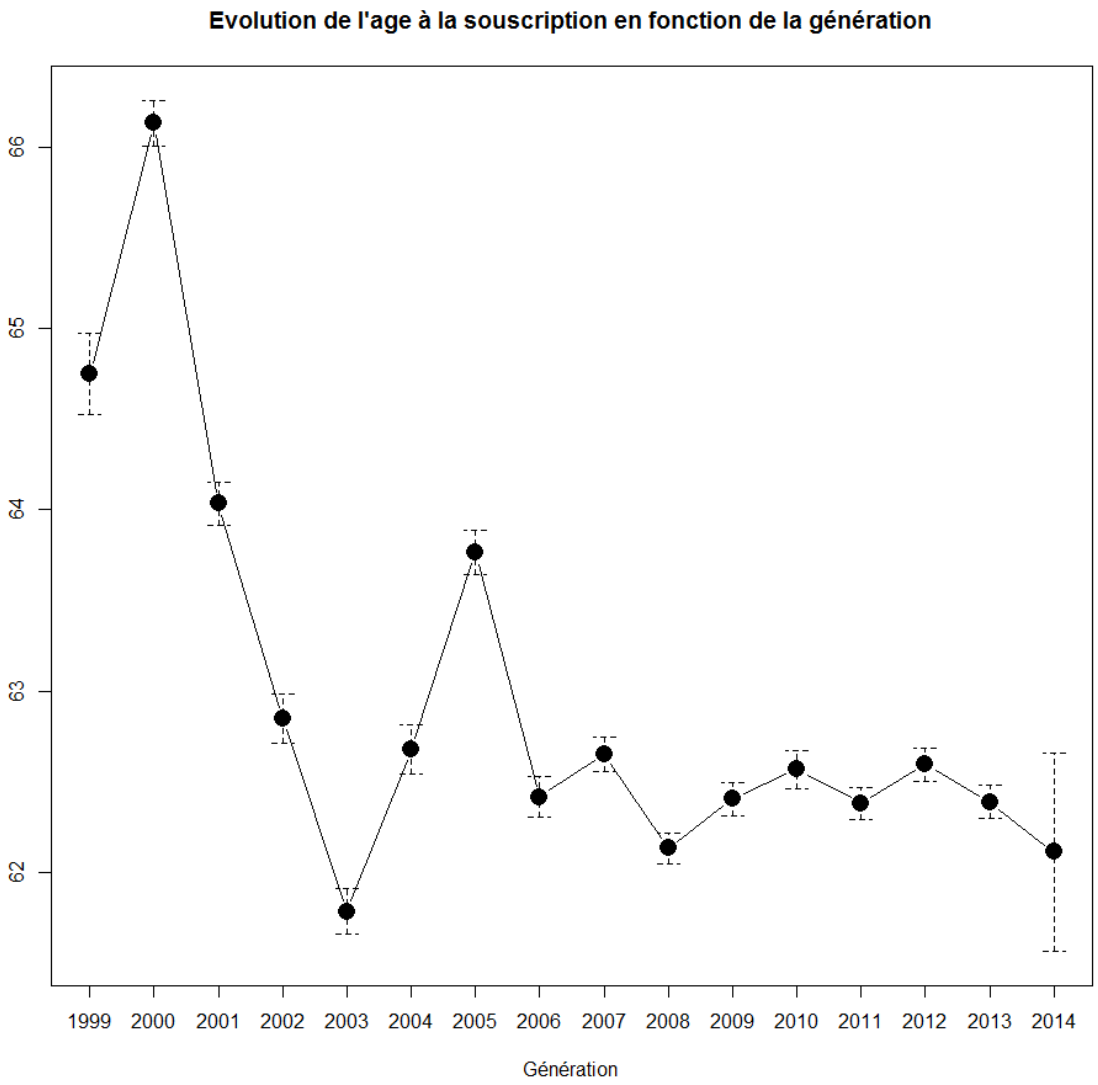

Illustration 4.3 : *Évolution de l'âge à la souscription au cours des générations de polices*

Durée d'activité des polices Par ailleurs, nous pouvons constater que globalement les polices restent actives assez longtemps. Par exemple, pour celles souscrites en 1999, la durée d'activité moyenne est de 148 mois (cf. Illustration 4.5) alors que la durée d'activité maximale de ces polices, s'il n'y avait eu aucune chute, aucun décès et aucune entrée en dépendance, serait de 169 mois (cf. Illustration 4.4).

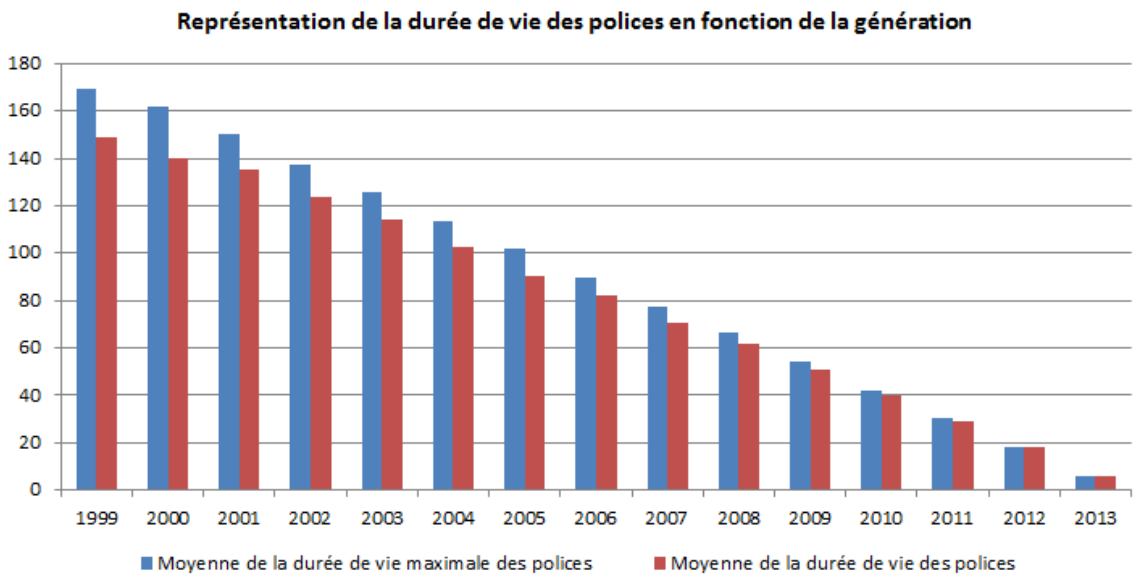

Illustration 4.4 : *Durée de vie de la police en fonction de la génération*

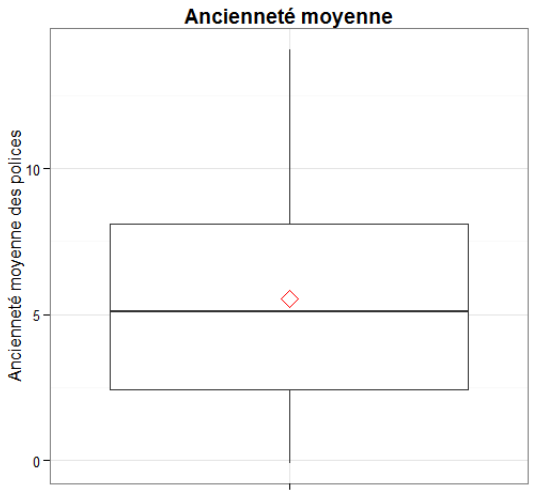

Illustration 4.5 : *Boîte à moustache de l'ancienneté des polices*

Répartition Homme Femme Les données concernent majoritairement des femmes, à plus de 63% toutes générations confondues (cf. Illustration 4.6).

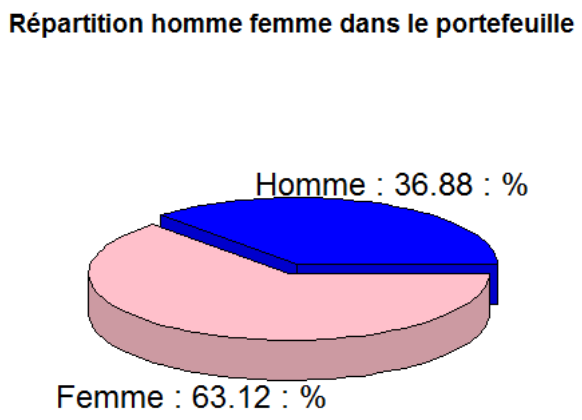

Illustration 4.6 : *Répartition homme femme des données*

Par ailleurs cette répartition est stable à travers les générations (cf. Illustration 4.7).

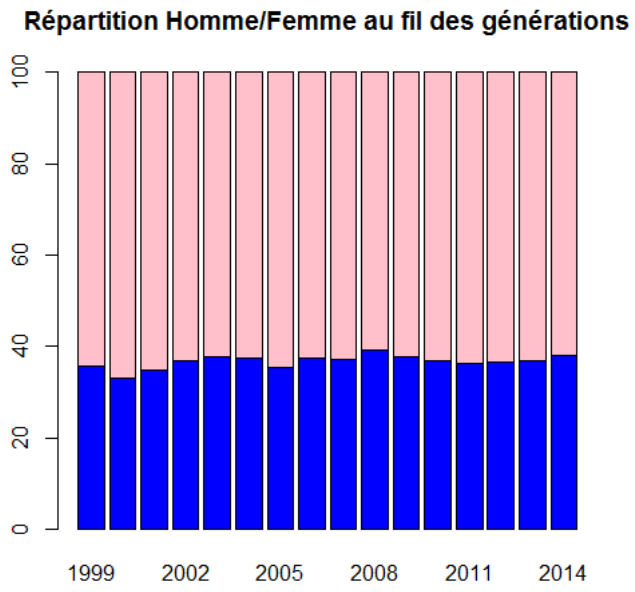

Illustration 4.7 : *Répartition homme femme du portefeuille à travers le temps*

Montant de rente souscrit Quant au montant de rente annuelle souscrit, il reste relativement faible, avec une moyenne à 6 761€ et une médiane à 5 987€ en comparaison des quelques valeurs extrêmes, avec des montants de rentes pouvant atteindre 25 000€ par an (cf. Illustration 4.8). Ce montant de rente moyen est dans la moyenne nationale : la F.F.S.A. estime que la rente versée au titre des contrats à adhésion individuelle est de 584 € par mois en moyenne, soit 7 008€ par an.

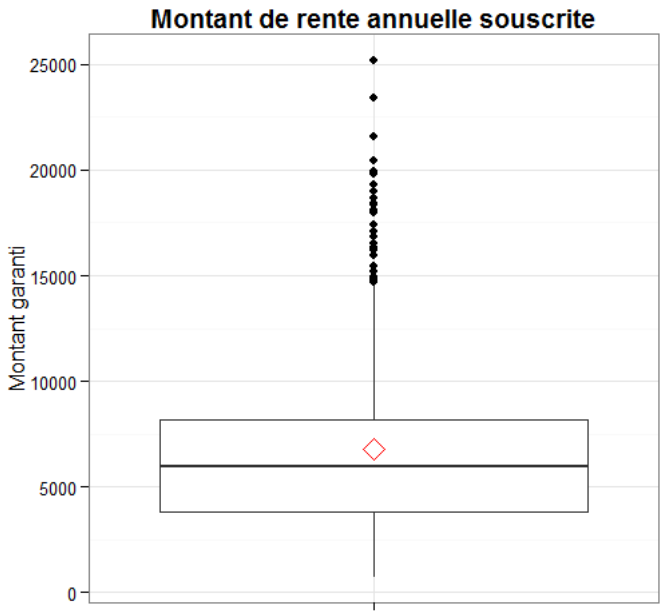

Illustration 4.8 : *Boite à moustache du montant de rente garanti*

Le montant moyen de rente souscrit fluctue à travers les générations, sans réelles tendances (cf. Illustration 4.9).

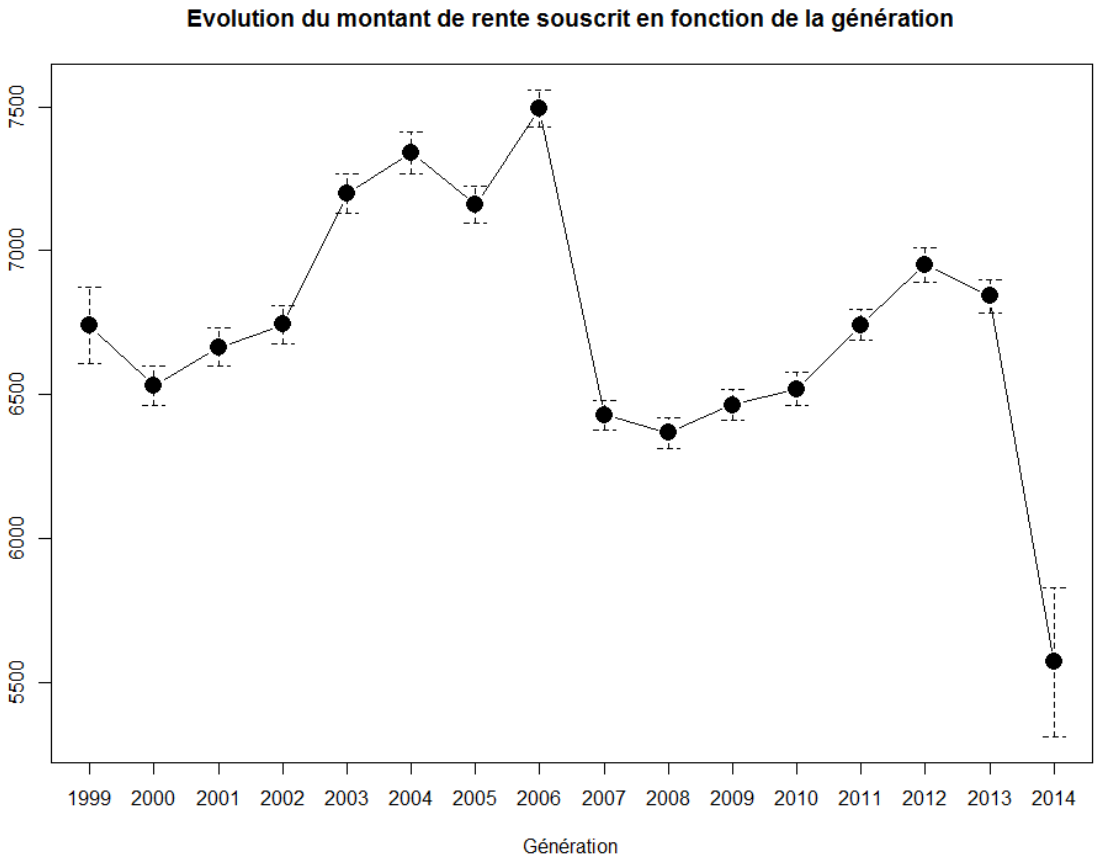

Illustration 4.9 : *Évolution du montant de rente garanti par génération*

Toutefois, nous constatons que le montant de rente souscrit est décroissant avec

l'âge à la souscription (cf. Illustration 4.10).

Excepté pour les personnes entre 44 et 51 ans, l'écart entre les montants souscrits par les femmes et les hommes n'est pas significatif.

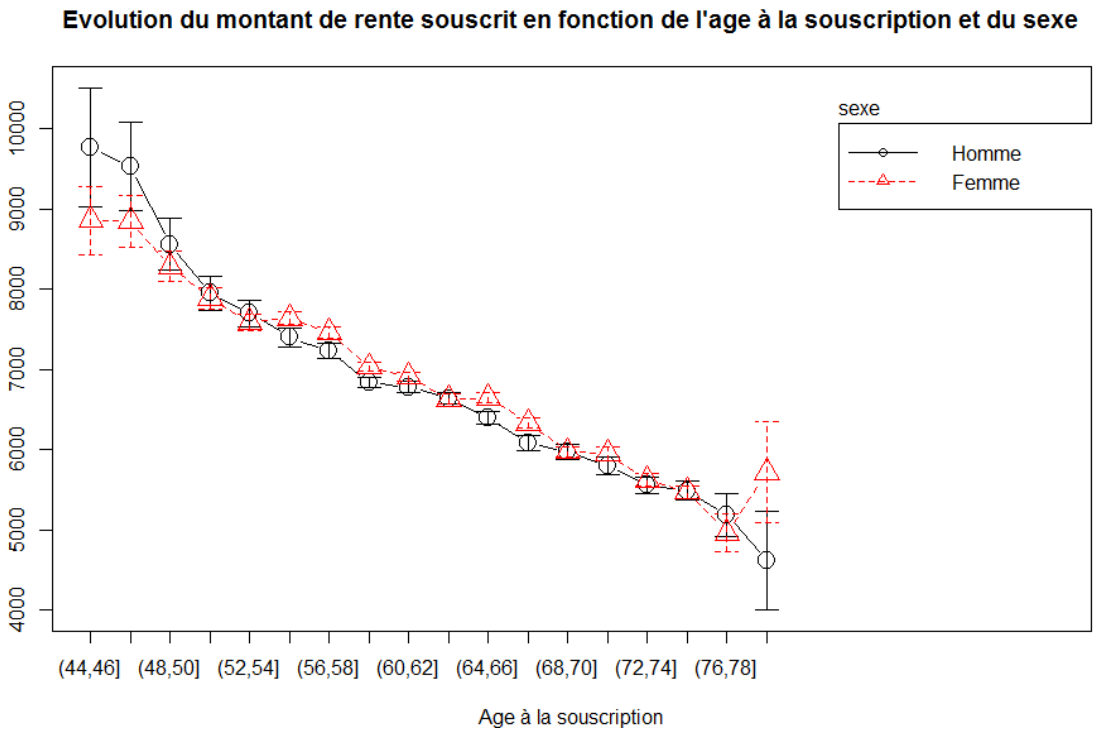

Illustration 4.10 : *Évolution du montant de rente garantie par âge à la souscription et par sexe*

Cette décroissance du montant de rente s'explique peut être par des pratiques commerciales, les réseaux de distributions proposant à des prospects plus âgés des montants de rente plus faibles pour leur octroyer des primes plus faibles, et inversement pour des prospects plus jeunes afin de ne pas dépasser un seuil de prime psychologique. Nous pouvons parler d'un "isobudget" où l'abritrage se fait entre prime et rente perçue.

4.2 Statistiques descriptives

4.2.1 Statistiques sur les autonomes

Nous étudions à présent la répartition des décès des autonomes, en fonction du sexe, de l'âge, et du temps passé depuis la souscription.

Répartition Homme Femme des décès Les décès des autonomes observés concernent majoritairement les hommes, avec 56%. L'âge moyen au décès est de 72 ans (71 ans pour les hommes et 73 ans pour les femmes) (cf. Illustration 4.11).

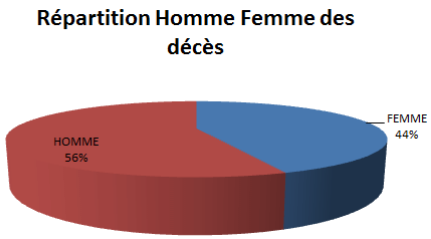

Illustration 4.11 : *Répartition des décès observés en fonction du sexe*

Répartition des décès en fonction de l'âge Quant à la répartition des décès en fonction de l'âge, 97% des décès observés sont répartis entre 57 ans et 87 ans (cf. Illustration 4.12).

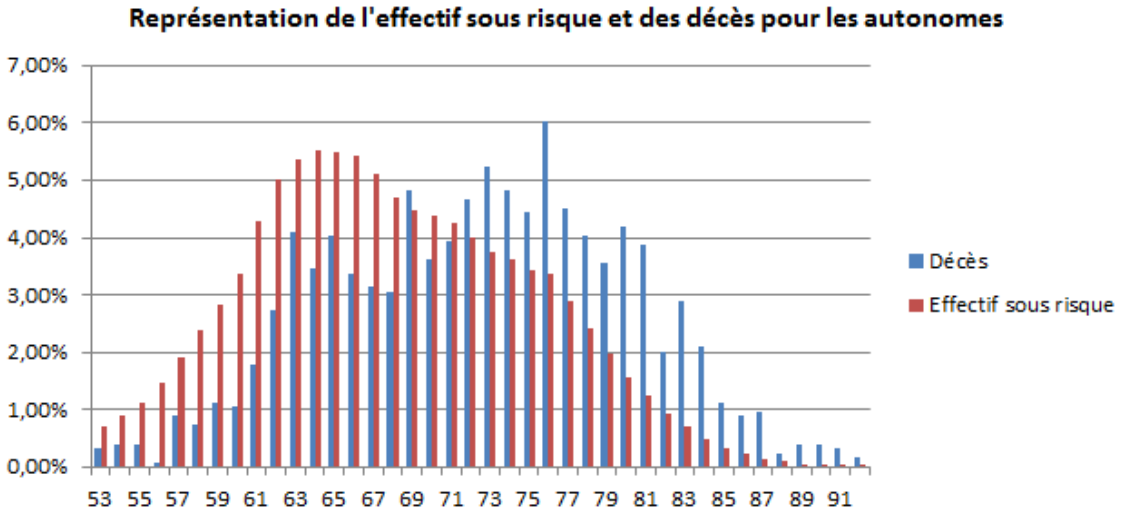

Illustration 4.12 : *Répartition des décès observés en fonction de l'âge. L'effectif sous risque est en pourcentage de l'effectif total et les décès en pourcentage du total des décès*

Il apparaît alors difficile d'estimer des taux de décès pour les autonomes en dehors de cette plage.

Toutefois, la répartition des décès en fonction de l'âge semble cohérente, il semble alors justifié d'étudier la mortalité des autonomes en fonction de leur âge.

4.2.2 Statistiques sur l'incidence en dépendance

Nous étudions dans cette partie les caractéristiques des individus entrant en dépendance partielle et totale.

Répartition Homme Femme de l'incidence Les observations d'entrée en dépendance concernent principalement des femmes, avec 55% de représentations dans les données observées (cf. Illustration 4.13). Cette observation n'indique pas que les

femmes entrent plus en dépendance que les hommes, au contraire, puisque le portefeuille est composé à 63% de femmes, donc statistiquement il y a plus d'hommes que de femmes qui entrent en dépendance.

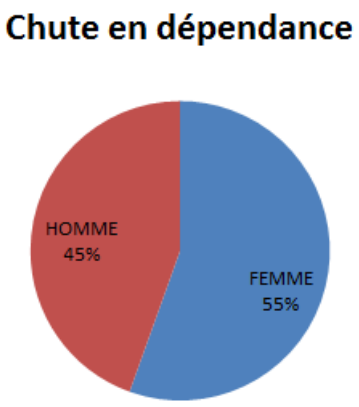

Illustration 4.13 : *Répartition Homme Femme des entrées en dépendance*

Répartition de l'incidence en fonction de l'âge Par ailleurs, la plage d'observation des entrées en dépendance s'étendant de 57 ans à 91 ans, il faudra être prudent lors de l'estimation des taux en dehors de l'intervalle 63-87 ans puisque 90% des entrées en dépendance sont comprises dans cet intervalle (cf. Illustration 4.14).

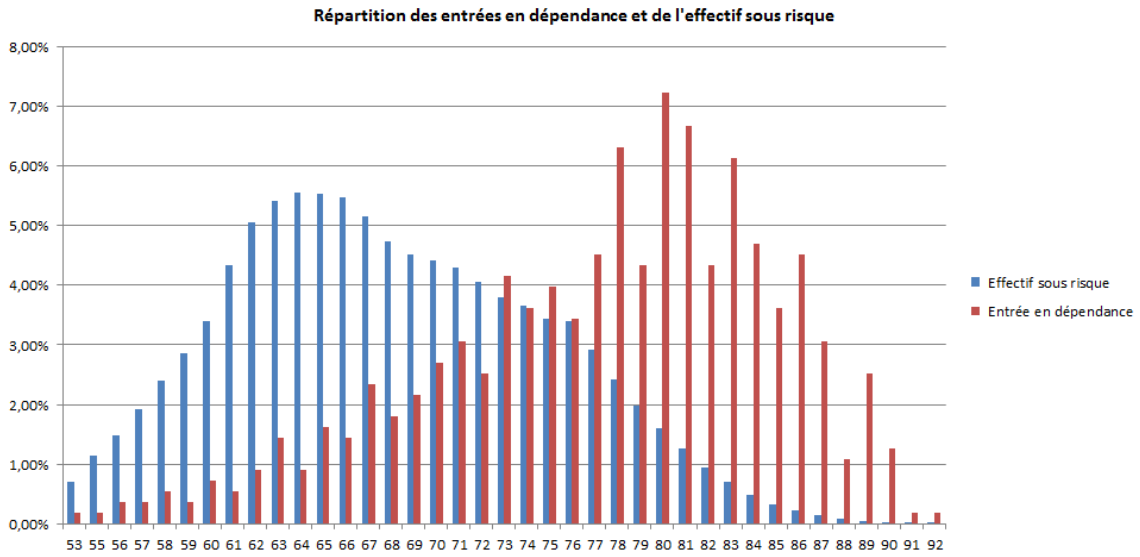

Illustration 4.14 : *Répartition des entrées en dépendance en fonction de l'âge. L'effectif sous risque est en pourcentage de l'effectif total et l'entrée en dépendance en pourcentage du total des entrées en dépendance*

Répartition des entrées en dépendance en fonction du temps passé depuis la souscription de la police La sélection médicale semble aussi avoir son effet. Nous observons une faible entrée en dépendance lors des premières années de souscription du contrat (cf. Illustration 4.15). Ainsi, la sélection médicale, qui permet de sélectionner des assurés en bonne santé, ne présentant pas un risque aggravé d'entrer

en dépendance (du moins par rapport à celui que l'assureur a prévu), semble avoir un réel impact.

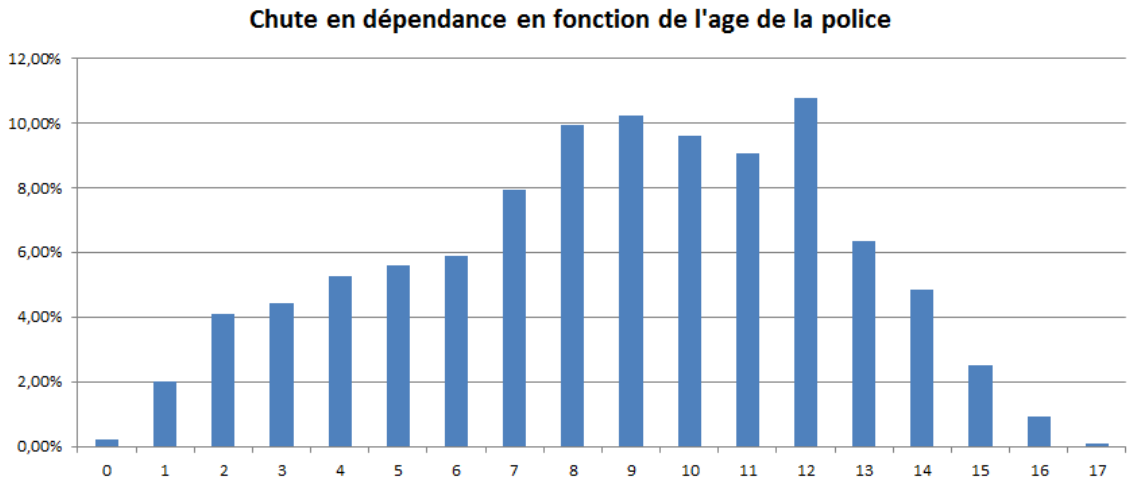

Illustration 4.15 : *Répartition des entrées en dépendance en fonction de l'âge de la police en année*

4.2.3 Statistiques sur les dépendants

Dans cette partie, nous étudions les caractéristiques des individus se trouvant en dépendance totale et partielle.

Âge à l'entrée en dépendance L'âge moyen d'entrée en dépendance est relativement tardif : en moyenne 77 ans (75 pour les hommes et 78 pour les femmes).

Si la dépendance partielle est distinguée de la dépendance totale, l'âge moyen à l'entrée en dépendance partielle est de 76 ans, quant à celui de l'entrée en dépendance totale, il est de 77 ans, tous sexes confondus.

Répartition Homme Femme des dépendants 60% des personnes observées en dépendance sont des femmes. Plus précisément, 51% des observations en dépendance partielle concernent des femmes, et 62% en dépendance totale (cf. Illustration 4.16).

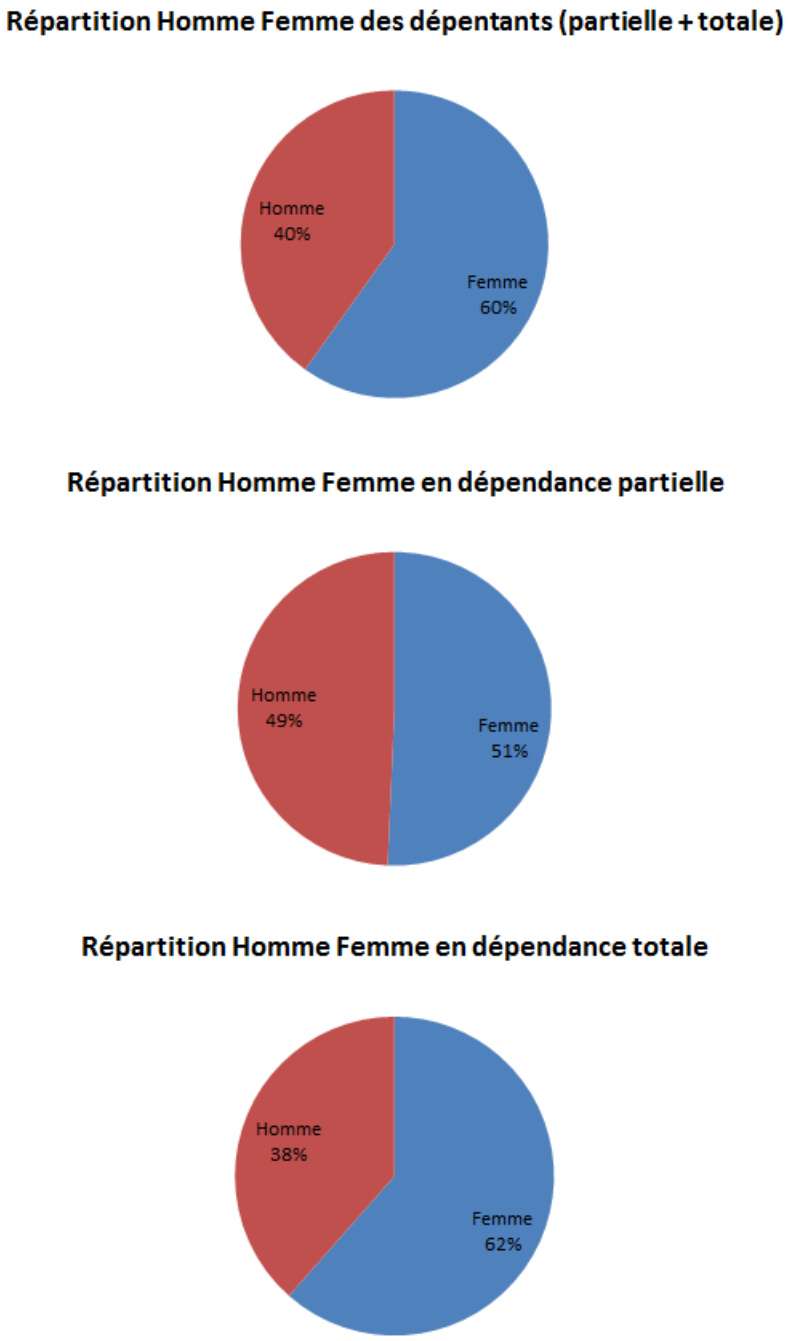

Illustration 4.16 : *Répartition des dépendants en fonction du sexe*

Durée passée en dépendance Les dépendants dont l'observation a été complète, c'est à dire dont la date d'entrée en dépendance et la date de décès sont connues, sont restés en moyenne 12.8 mois en dépendance partielle et totale, 7.7 en dépendance partielle et 13.3 mois en dépendance totale.

Hormis quelques exceptions, la durée passée en dépendance semble être fortement corrélée à l'âge à l'entrée en dépendance, en particulier pour la dépendance totale (cf. Illustration 4.17). Nous remarquons que plus la personne entre en dépendance jeune, plus rapidement elle décède. La variabilité que nous constatons pour les personnes âgées de moins de 60 ans est due à une exposition faible.

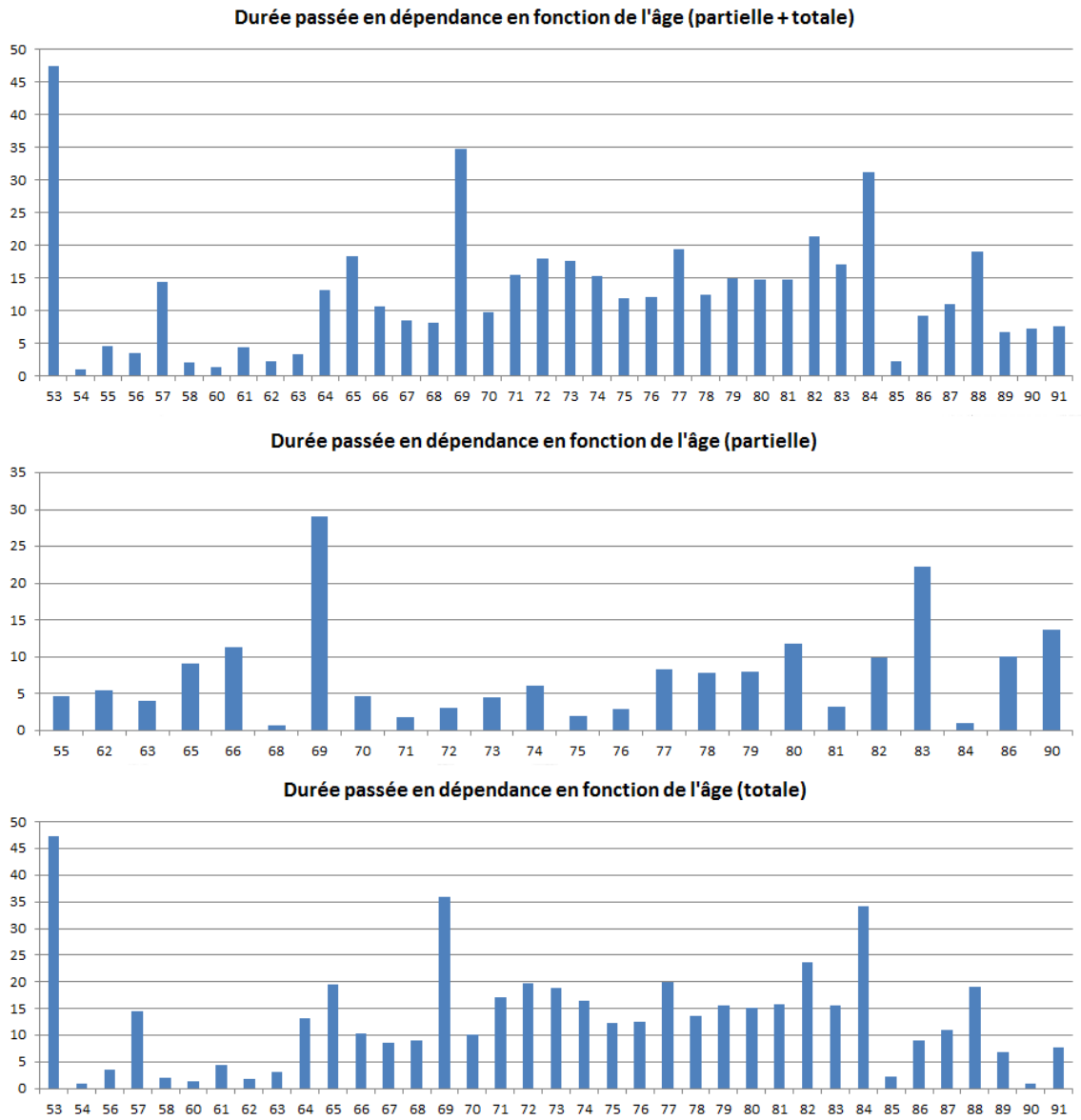

Illustration 4.17 : *Durée de dépendance en fonction de l'âge à l'entrée en dépendance*

Ce phénomène s'explique probablement par les différentes pathologies touchant les dépendants. En effet, nous constatons que les individus devenant dépendants jeunes le deviennent en grande majorité des suites d'un cancer.

En revanche, à mesure que les individus vieillissent, les maladies neurologiques expliquent de plus en plus l'entrée en dépendance. Sur le graphique 4.18, nous observons bien cette tendance : les chutes en dépendance des suites d'un cancer (en rouge) s'amenuisent avec l'âge pour laisser place aux chutes dues à des causes neurologiques (en violet) et autres maladies (en vert).

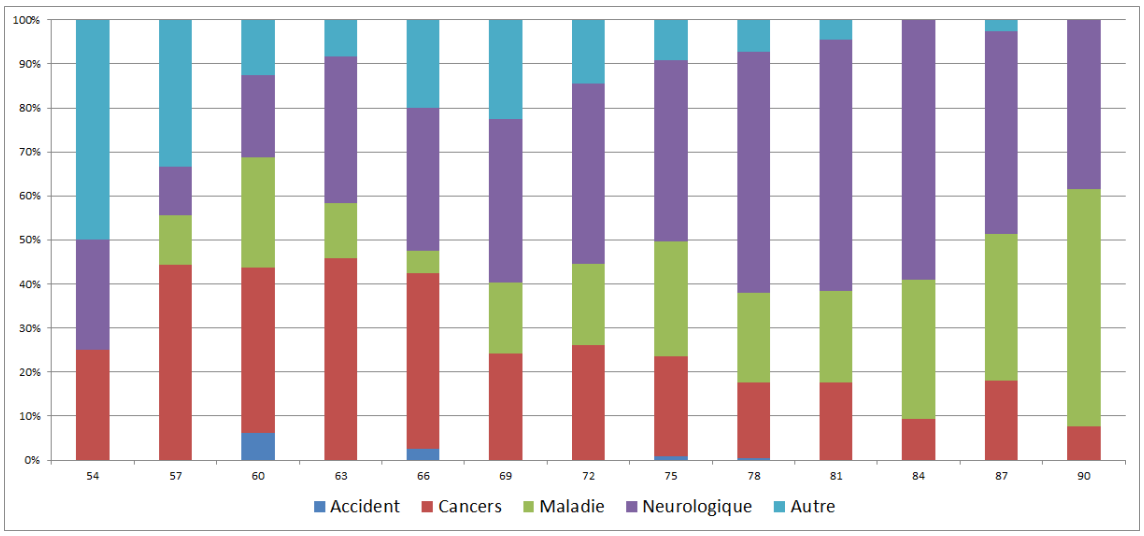

Illustration 4.18 : *Répartitions des pathologies en fonction de l'âge à l'entrée*

Partie 5

Estimation des probabilités

Dans la suite, nous estimerons les probabilités utiles aux projections en utilisant les méthodes présentées précédemment. Signalons que l'étude n'est pas effectuée par genre, même si ce critère est pertinent. Nous ne disposons pas de suffisamment de sinistres afin de réaliser cette segmentation et préférons obtenir des estimations robustes.

5.1 Estimation des probabilités de décès pour les autonomes

L'estimation de la probabilité de décès des autonomes est importante dans la mesure où ce sont ces derniers qui apportent des primes à la compagnie. Ainsi, plus ces personnes vont vivre longtemps, plus elles apporteront de primes. Toutefois, cet effet est à mitiger ; certes, le fait qu'elles apportent des primes est bénéfique pour la compagnie, mais le fait qu'elles soient encore en vie entraîne l'existence d'un risque pour l'assureur : plus ces personnes vieillissent, plus elles ont une probabilité élevée de tomber en dépendance. Il semblerait que ce dernier effet soit le plus "néfaste" pour la compagnie. En effet, si un assuré décède sans être tombé en dépendance, il n'a fait qu'apporter des cotisations et ainsi n'a rien coûté à la compagnie. Cette estimation se doit alors d'être la plus juste possible, afin de ne pas surestimer la durée de cotisation, mais aussi de ne pas sous-estimer l'exposition au risque.

Pour l'estimation des probabilités de décès des autonomes, nous positionnons leur mortalité par rapport à la mortalité nationale. Cela nous permet de nous baser sur des estimations fiables, et d'éviter la problématique du lissage. Nous ajustons donc la mortalité obtenue des tables réglementaires TH02 et TF02, afin de prendre en compte la plus faible mortalité d'une population d'assurés. En effet, cette population présente des particularités liées à la sélection médicale et à leurs conditions de vie probables.

Les assurés, qui ont au préalable suivi une sélection médicale, ont une probabilité de décès amoindrie, il s'agit principalement de personnes en bonne santé, contraire-

ment à la population nationale qui intègre aussi un certain nombre de personnes en mauvaise santé. Par ailleurs, du fait de la volonté de s'assurer, la personne témoigne d'une certaine prévoyance quant à l'avenir, elle se projette dans le futur, particulièrement pour les produits de dépendance. Nous pouvons en effet supposer qu'il s'agit de personnes souscrivant ces contrats car elles ont certes peur de leur vieillissement, mais toutefois envisagent la possibilité de vieillir.

Par ailleurs, puisque ces personnes s'assurent, nous en déduisons qu'elles disposent d'un niveau de vie suffisant pour leur assurer une hygiène de vie correcte. Toutefois, l'effet de cette sélection médicale s'estompe avec le temps (au bout de 4 ans). Nous tâcherons de visualiser cet effet sur les données étudiées.

La mortalité des tables TH02 et TF02 reflète une vue instantanée de la mortalité française.

Nous chercherons donc à abattre les tables TH02 et TF02 grâce à deux méthodes, la méthode SMR et la méthode d'ajustement des logits, puis nous sélectionnerons celle qui nous permet d'obtenir la meilleure estimation. Nous verrons ensuite comment nous pouvons fermer la table, c'est à dire extrapoler les taux de mortalité à des âges plus élevés que ceux observés.

5.1.1 Taux brut et lissage

Dans un premier temps nous présentons l'estimation des taux bruts de mortalité, lissés avec la méthode de Whittaker Henderson (cf. Illustration 5.1). Les taux bruts sont obtenus grâce à la fonction de survie de Kaplan Meier en remarquant que, avec t représentant le temps passé en vie d'un individu et x un instant donné :

$$S(x) = \mathbb{P}(t > x)$$

$$\begin{aligned} q_x^a &= \mathbb{P}(t < x+1 | t > x) \\ &= \frac{\mathbb{P}(t < x+1 \cap t > x)}{\mathbb{P}(t > x)} \\ &= \frac{\mathbb{P}(x < t < x+1)}{\mathbb{P}(t > x)} \\ &= \frac{S(x) - S(x+1)}{S(x)} \end{aligned}$$

Cet estimateur permet de ne pas faire d'hypothèse sur la répartition des décès à l'intérieur d'une année. De plus, il permet l'obtention d'intervalles de confiance.

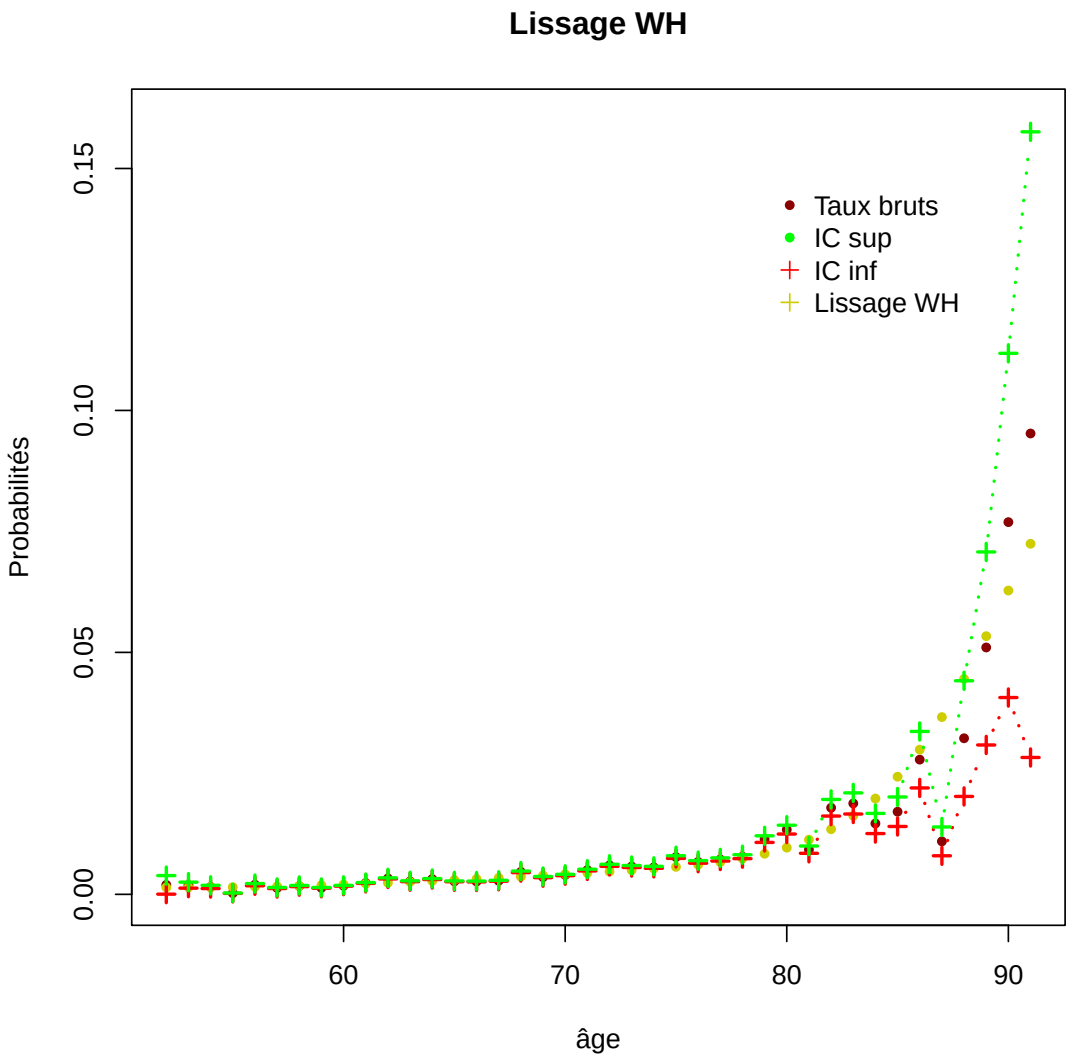

Illustration 5.1 : *Représentation des taux de mortalité bruts et lissage de Whittaker Henderson*

Nous constatons clairement l'augmentation progressive de la probabilité de décès à mesure que l'âge augmente. Toutefois, le manque de données pour les âges élevés entraîne une grande variabilité dans l'estimation des taux bruts à en juger par la largeur de l'intervalle de confiance, s'élargissant de plus en plus avec l'âge.

Dans la suite, nous abattons une table mixte. Nous avions constaté une présence féminine pour les autonomes d'environ 60%, nous prennons donc comme base :

$$q_x = 40\% q_x^{homme} + 60\% q_x^{femme}$$

Nous créons ainsi une nouvelle table, que nous nommons TH-TF.

5.1.2 Méthode SMR

Nous disposons d'observations de décès entre les âges 53 et 92. Le SMR pour cette plage est de 25% (cf. Illustration 5.2). Cela revient donc à dire que la population

assurée du portefeuille a une probabilité annuelle de décès environ 75% moins élevée que la population nationale. Pour rappel, le SMR est égal au rapport entre le nombre de décès observé et le nombre de décès attendu (tel que prédit par la table de référence, la table TH-TF en l'occurrence).

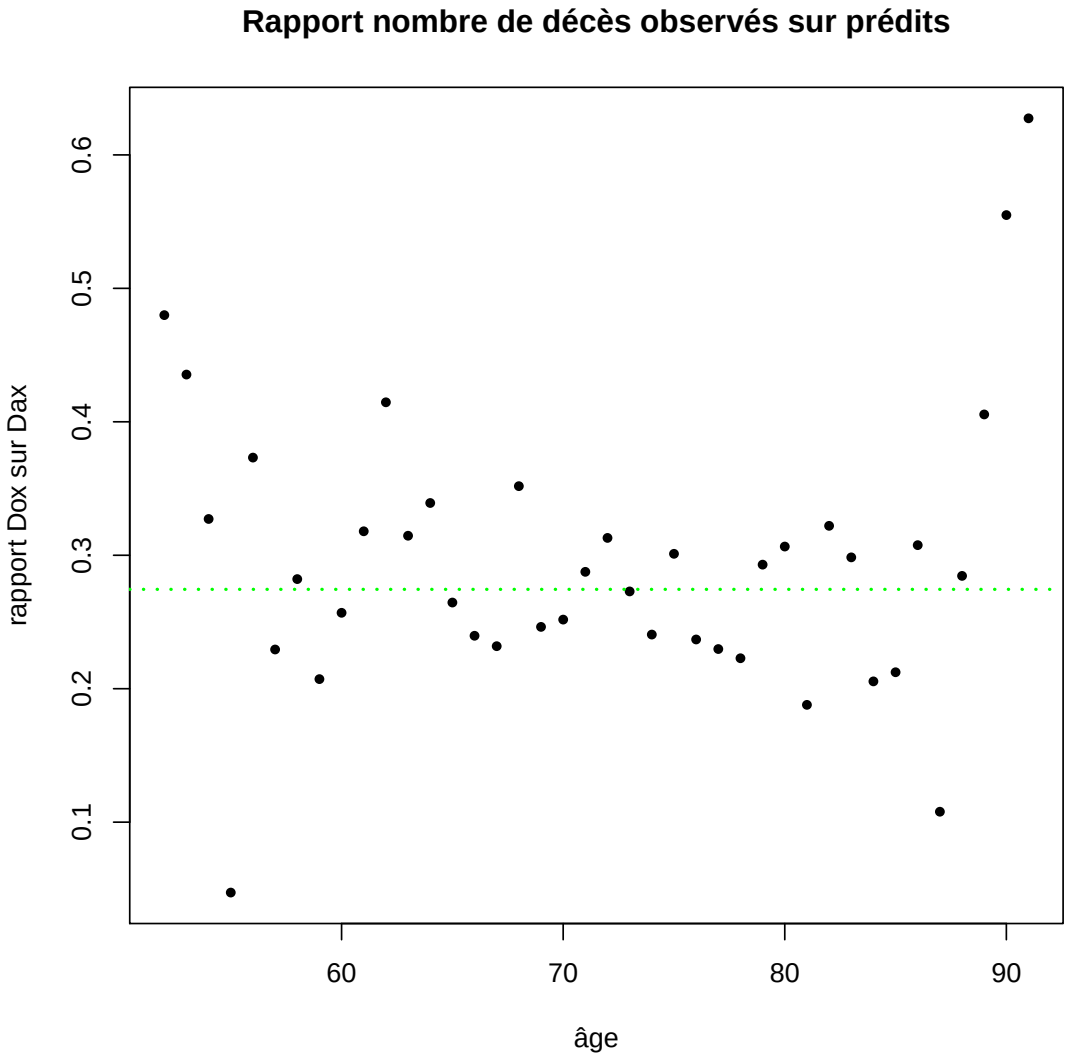

Illustration 5.2 : *Représentation du rapport nombre de décès observés sur nombre de décès attendus avec la table TH-TF*

La table TH-TF est donc abattue de 75% et nous obtenons :

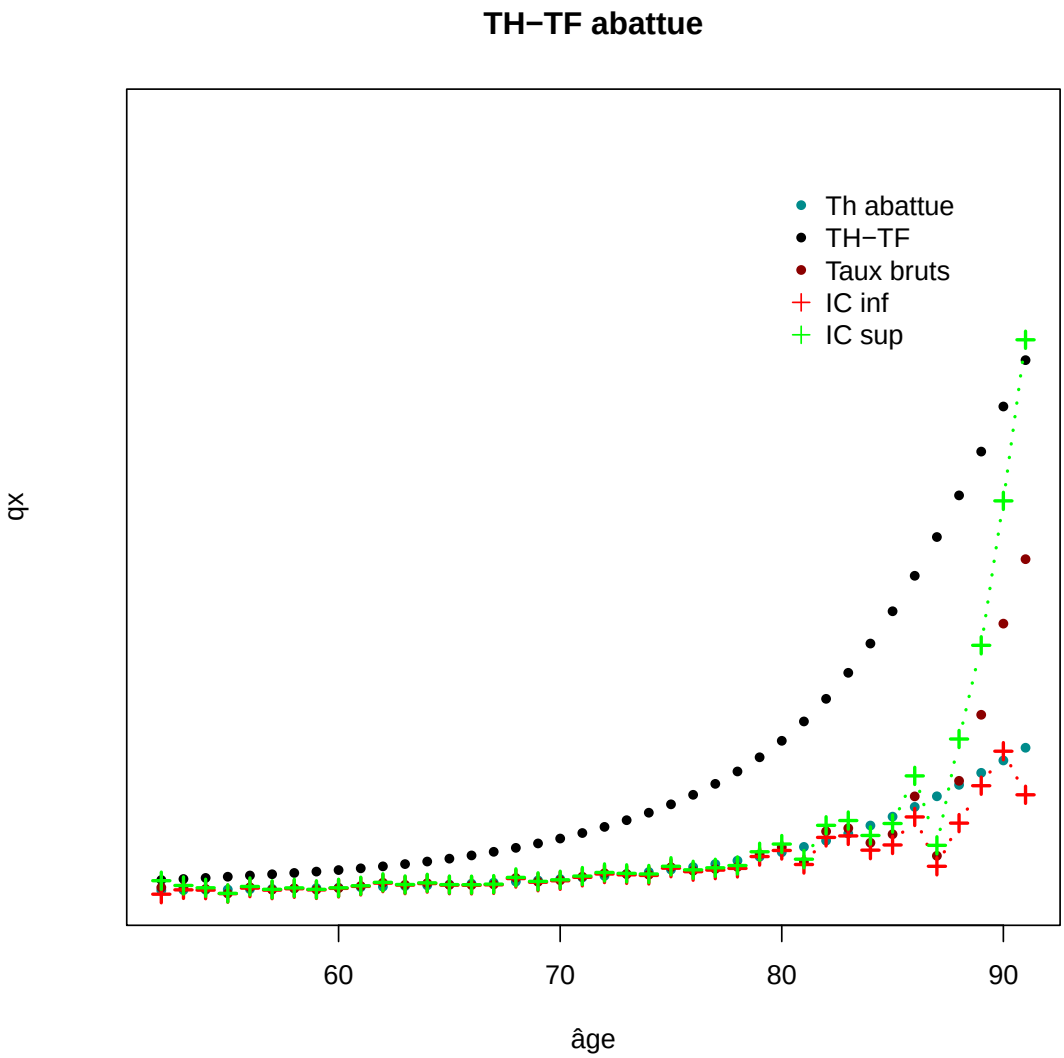

Illustration 5.3 : *Représentation des qx bruts et des qx abattus et des qx de la TH-TF*

Cette méthode pourrait permettre d'obtenir des taux de mortalité pour les âges supérieurs à 92 ans. Toutefois, et comme signalé dans la partie statistique sur les autonomes, il apparaît clairement que le manque d'observation après 87 ans conduit à un mauvais ajustement des taux.

Nous effectuerons une analyse de ces résultats afin de savoir si cet ajustement peut être validé ou non et si oui, pour quel intervalle. Nous mettons d'abord en place la seconde technique, l'ajustement des logits.

5.1.3 Modèle relationnel de Brass

Nous calculons les logits des taux de décès observés, ainsi que les logits des taux de décès de la table TH-TF (cf. Illustration 5.4). Pour rappel, le logit est le logarithme du rapport entre le taux et la différence entre 1 et ce taux. Nous effectuons le même ajustement sur l'intervalle de confiance à 95% des taux bruts afin d'obtenir un encadrement de l'ajustement (nous ne détaillons pas l'ajustement

sur les intervalles de confiance) Nous obtenons :

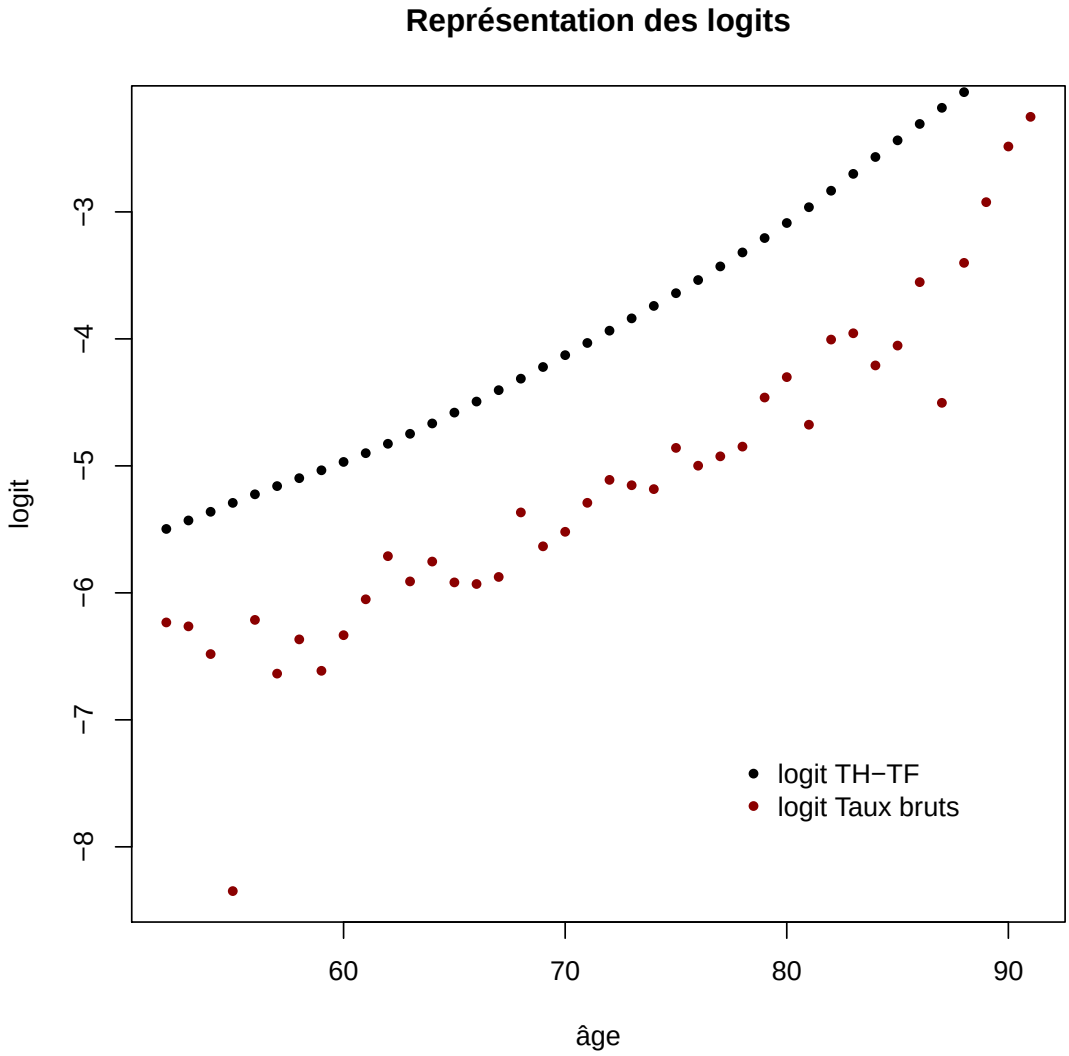

Illustration 5.4 : *Représentation des logits des taux de de mortalité de la table TH-TF et de ceux de la population autonome*

Nous calculons les paramètres $\hat{\alpha}$ et $\hat{\beta}$, estimés ici à -1.4 et 1 environ. Nous avons ainsi :

$$logit(\tilde{q}_{x^*}) = -1.4 + 1 \cdot logit(q_{x^*}^{ref})$$

Nous pouvons observer l'ajustement des logits :

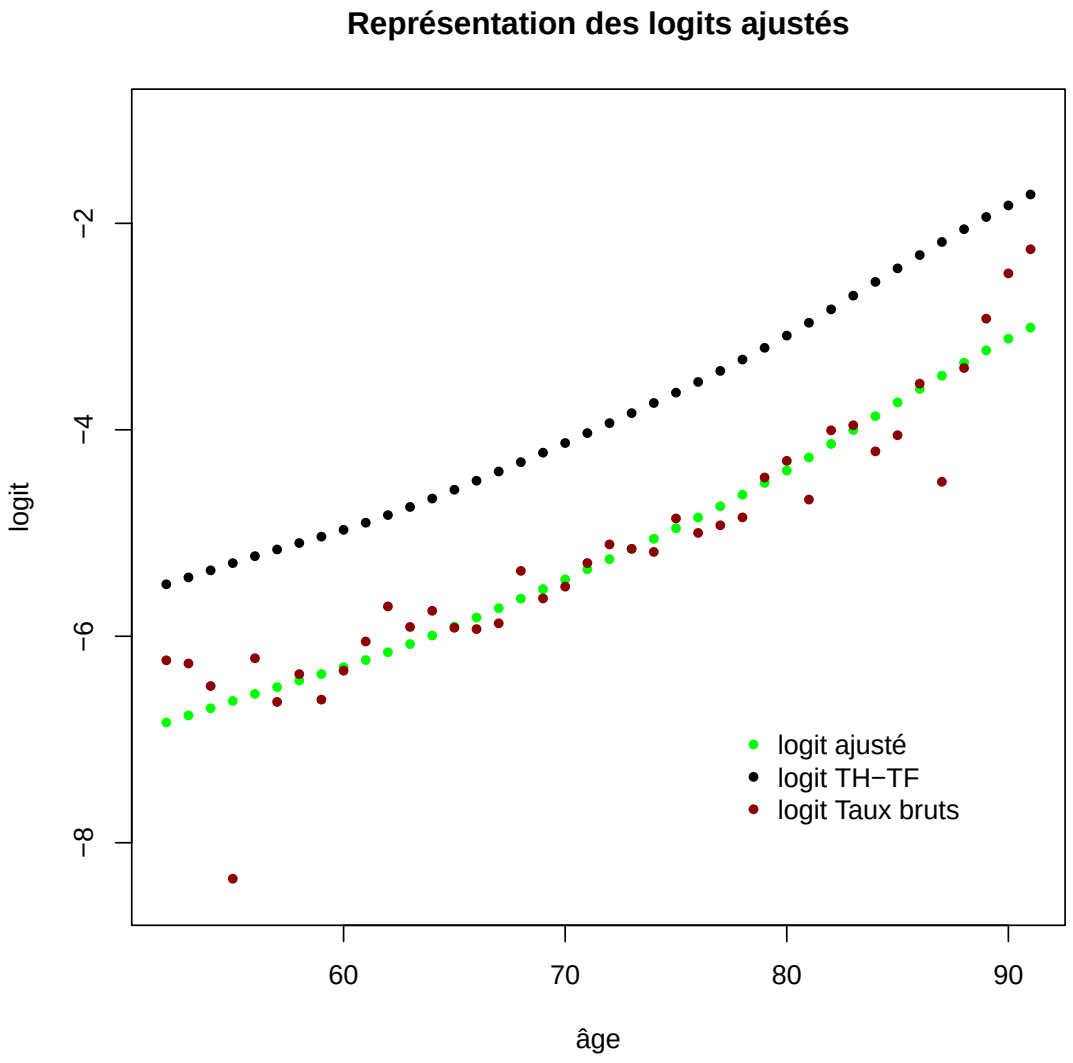

Illustration 5.5 : *Ajustement des logits des taux de mortalité de la TH-TF sur les logits des taux de mortalité observés*

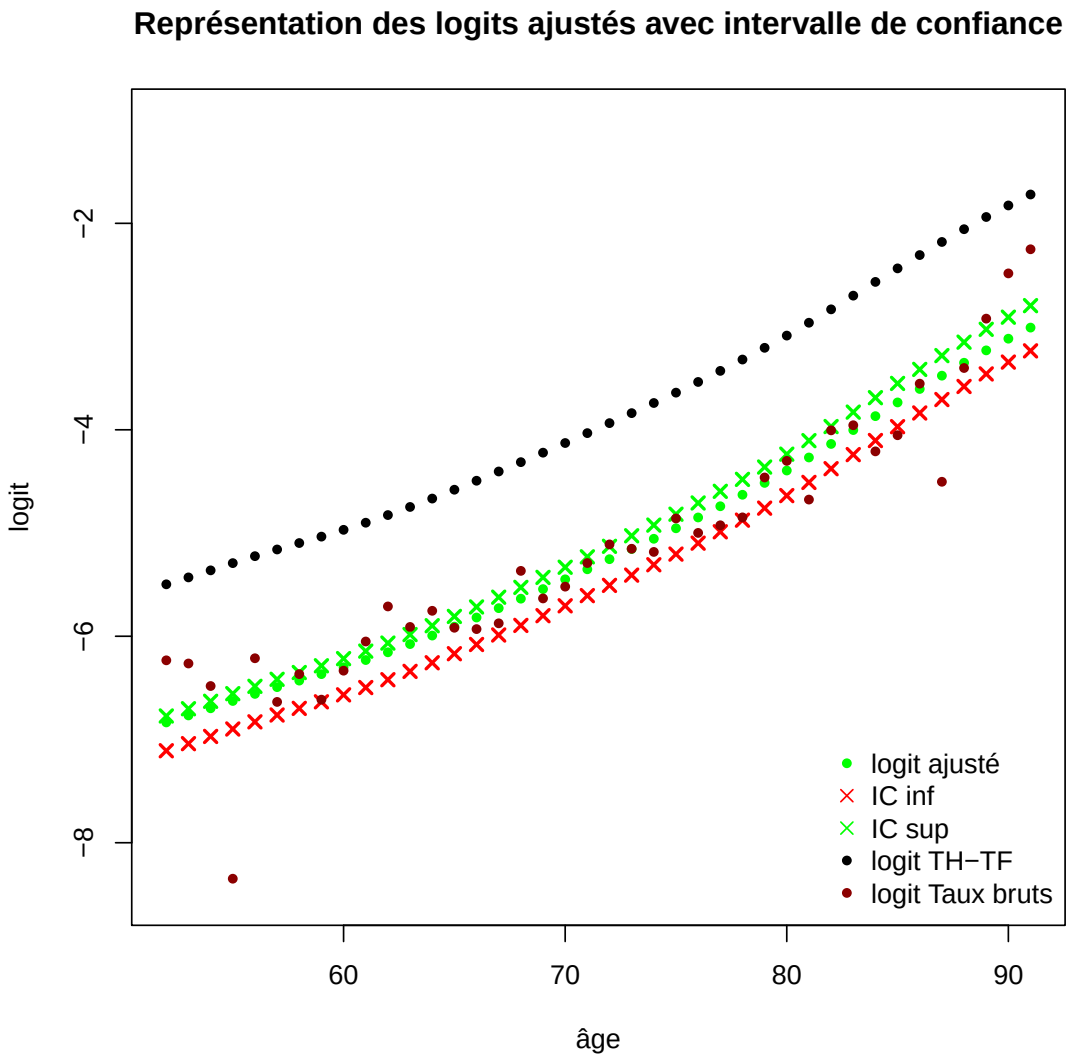

Illustration 5.6 : *Ajustement des logits des taux de mortalité de la TH-TF sur les logits des taux de mortalité observés*

L'ajustement est satisfaisant à en juger par le R^2 [1] de 0.88 suggérant que le modèle explique 88% de la variabilité du logit des taux de mortalités bruts. La p-value pour le test de Fisher [2] de nullité des paramètres est inférieure à 2.2e-16, indiquant que la relation entre le logit des taux bruts et le logit des taux de la TH-TF est statistiquement significative.

Nous déterminons alors les taux de décès ajustés :

$$\tilde{q}_{x^*} = \frac{1}{1 + \exp[-(-1.4 + 1 \cdot log(\frac{q_{x^*}^{ref}}{1-q_{x^*}^{ref}}))]}$$

1. Voir annexes
2. Voir annexes

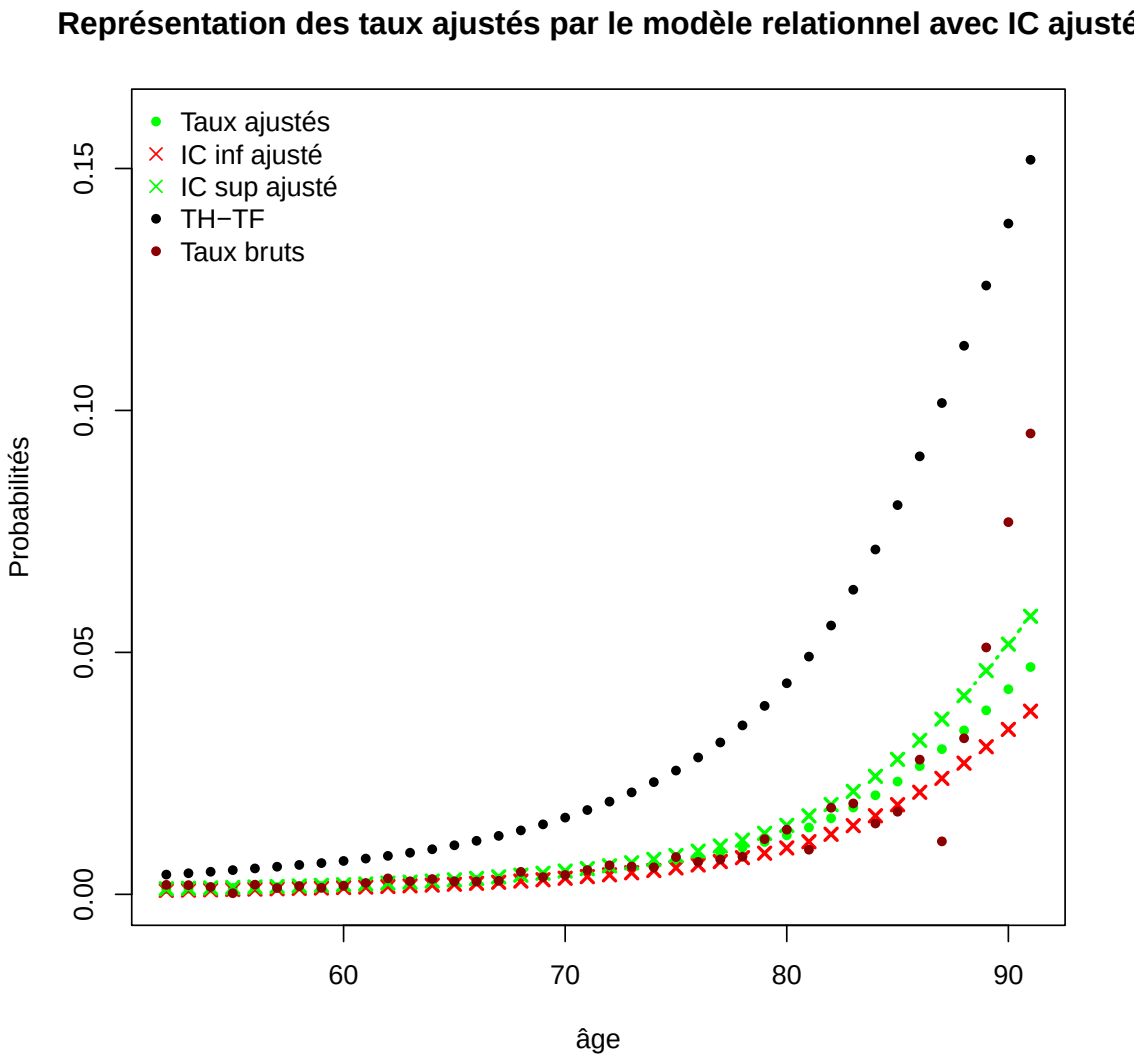

Illustration 5.7 : *Représentation taux de mortalité observés, des taux de mortalité de la TH-TF et des taux ajustés par le modèle relationnel*

Nous constatons à nouveau que la mortalité des assurés est largement inférieure à la mortalité nationale. Cette méthode présente aussi l'avantage d'une éventuelle extrapolation des taux de mortalité. Toutefois, l'ajustement pour les âges élevés ne semble pas correspondre à la réalité. En effet il semble y avoir une différence assez forte entre les taux bruts et les taux ajustés.
Pour valider si ces ajustements peuvent être acceptés ou non, nous effectuons des tests d'adéquation et une analyse des résidus.

5.1.4 Comparaison des deux méthodes

Analyse visuelle

Une comparaison visuelle des courbes obtenues à partir des deux méthodes semble indiquer que la méthode d'ajustement par le modèle relationnel est plus prudente, dans le sens où elle estime une mortalité plus élevée que la méthode

d'abattement par le SMR. Toutefois, les deux méthodes semblent assez proches l'une de l'autre. Et de même il ne semble pas évident que surestimer la mortalité des autonomes soit plus prudent que la sous-estimer dans notre cas de figure. Par conséquent, des tests supplémentaires s'imposent et seront réalisés par la suite. La deuxième méthode permettrait éventuellement d'effectuer une extrapolation plus prudente des taux de mortalité.

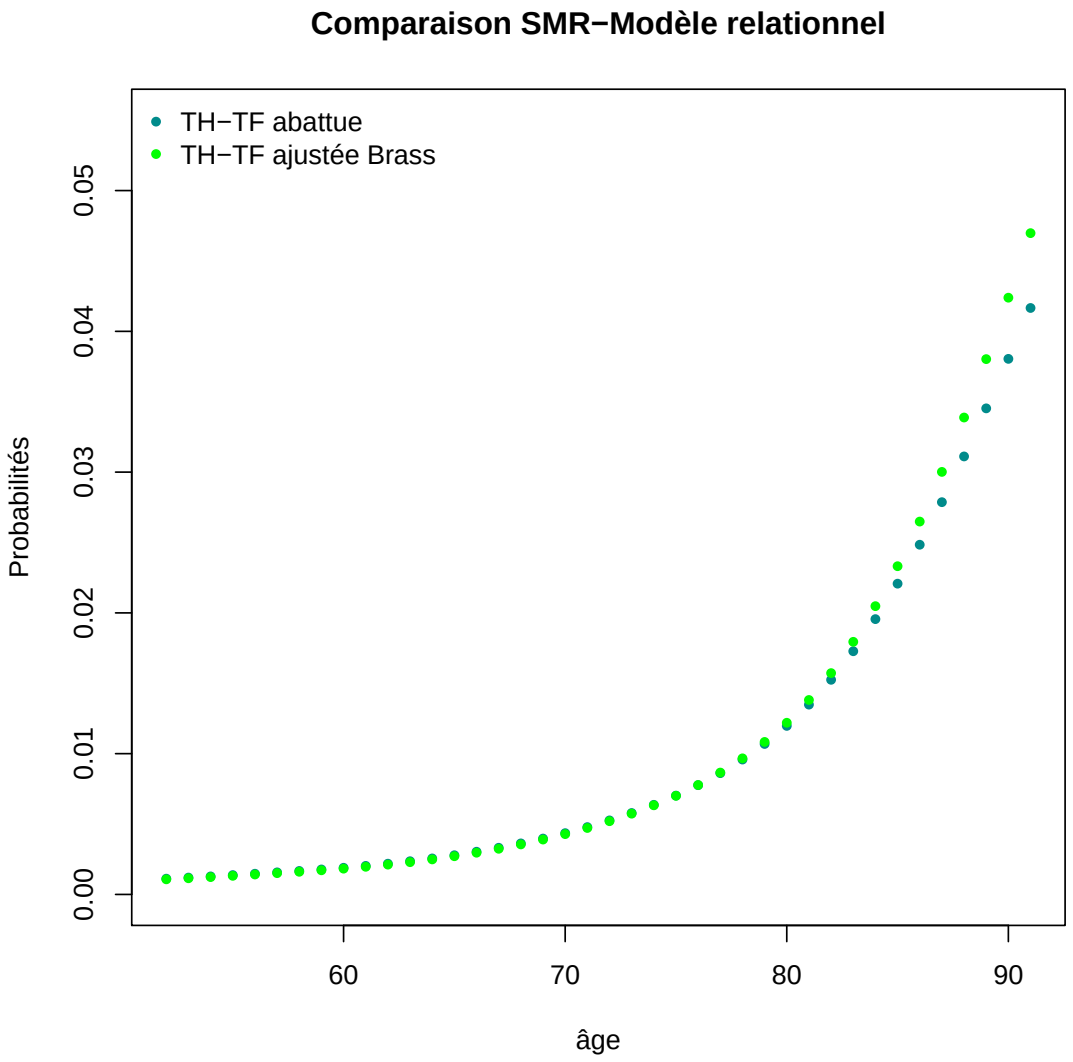

Illustration 5.8 : *Comparaison des taux de mortalité abattus par le SMR et des taux de mortalités ajustés par le modèle relationnel*

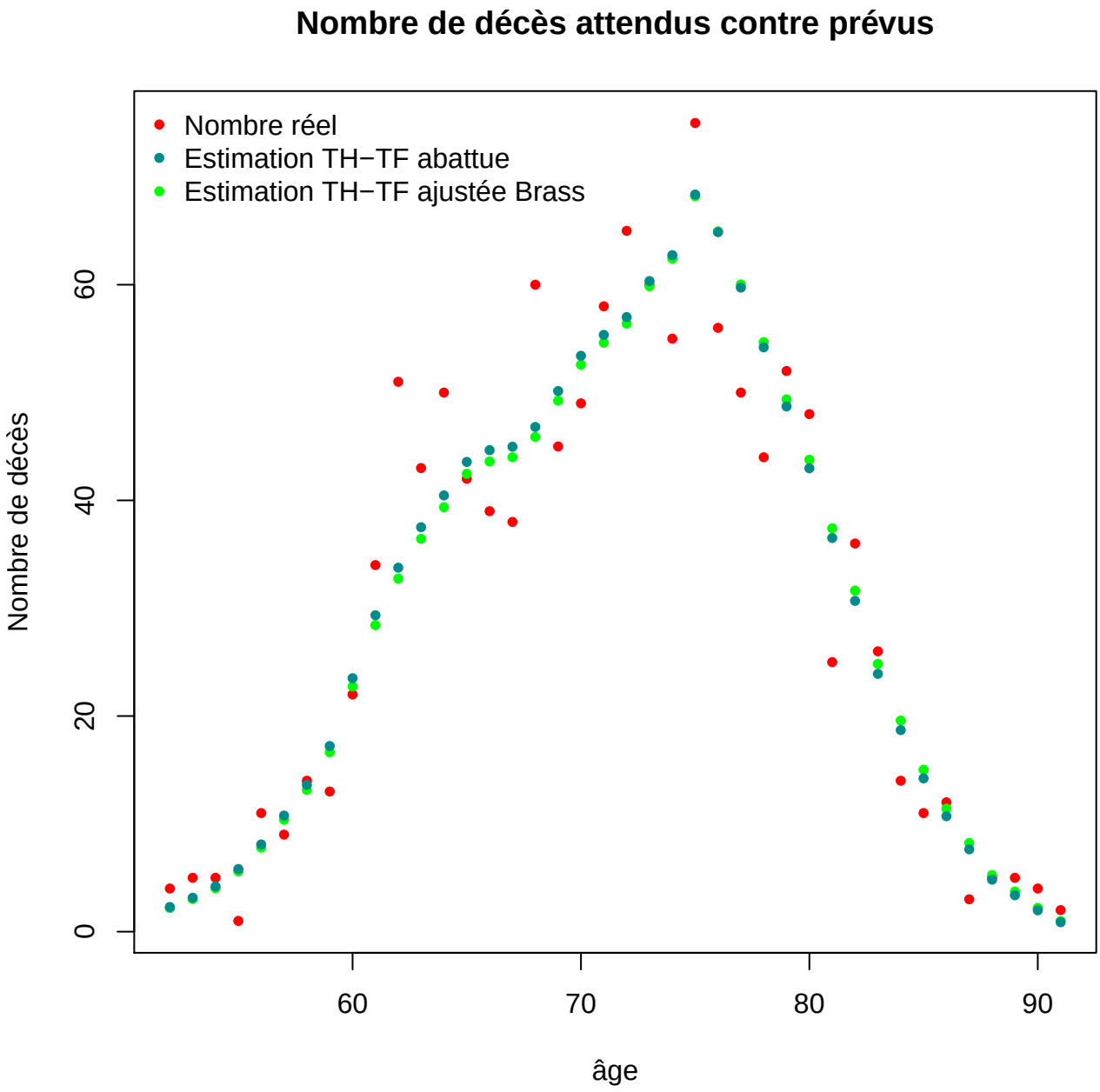

Illustration 5.9 : *Comparaison du nombre de décès attendus avec les taux de mortalité abattus par le SMR et ceux attendus avec les taux de mortalité ajustés par le modèle relationnel et le nombre de décès réels*

Pour valider si ces ajustements peuvent être acceptés ou non, nous effectuons des tests d'adéquation et une analyse des résidus.

Test	Valeurs des tests pour l'ajustement SMR	Valeurs des test pour l'ajustement relationnel de type Brass
Khi 2	0.011	0.011
MAPE	1.10	1.10
SMR	1	1.001

De manière générale, les tests d'adéquations obtenus nous feraient valider les ajustements effectués. En effet, les tests du Khi 2 sont très proches de 0, le coeffi-

cient MAPE est assez faible. Les SMR, très proches de 1, indiquent la proximité de l'ajustement avec les taux bruts. Aucun des deux modèles ne se distingue au vu de des tests d'adéquation.

Toutefois l'analyse des résidus est nécessaire car nous avions constaté un mauvais ajustement pour les âges élevés, qui pourrait ne pas se refléter dans les tests d'adéquation.

Analyse des résidus

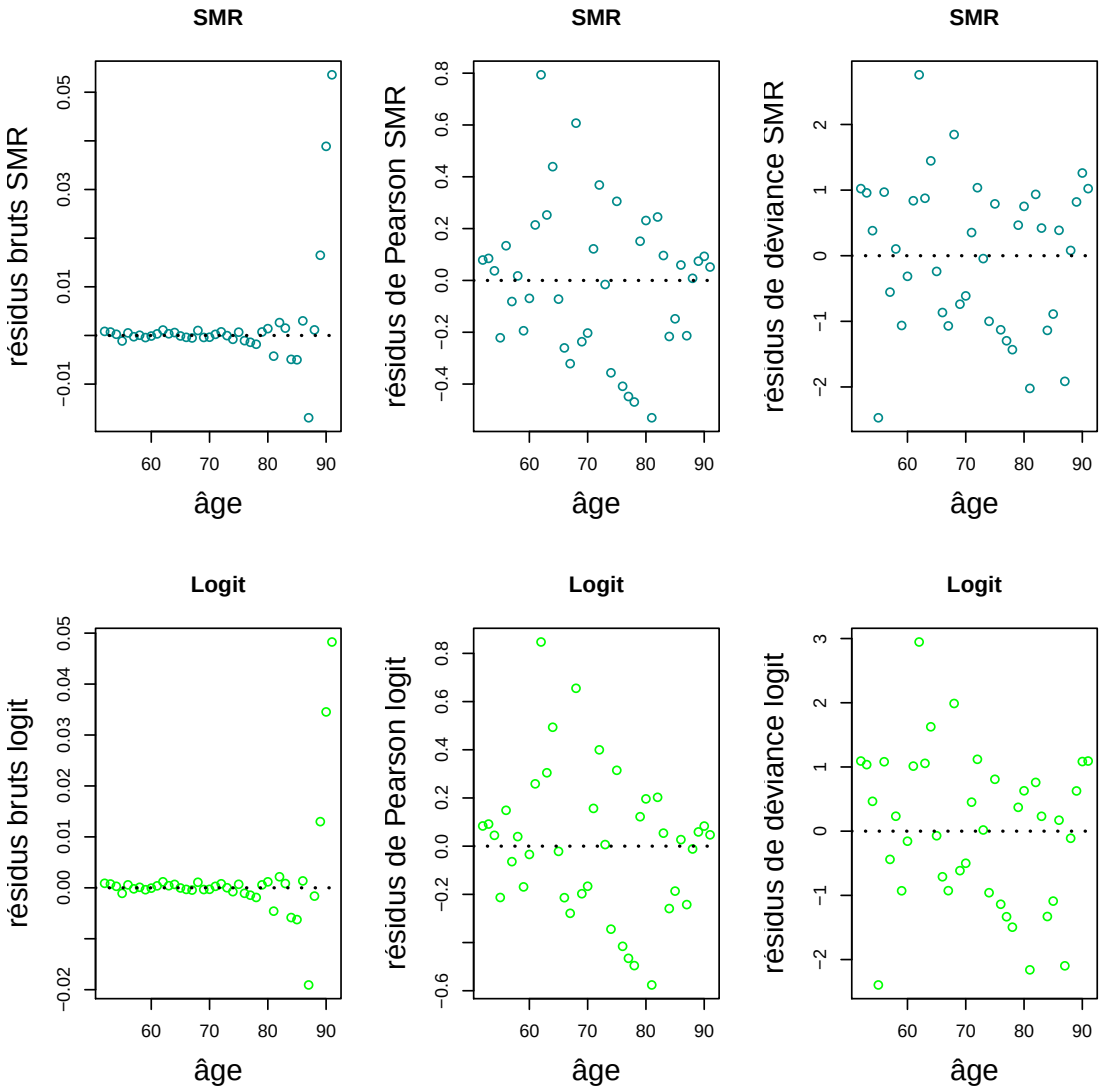

Illustration 5.10 : *Représentation des résidus pour les deux modèles*

Les résidus bruts sont bien proches de 0 jusqu'à 86, 87 ans. Toutefois après ces âges, les résidus s'en éloignent fortement.

Malgré tout, les résidus de Pearson ne décèlent pas de valeurs anormales, en effet, tous les résidus de Pearson pour les deux méthodes sont compris entre -1 et 1, ce qui indique que le modèle prend correctement en compte la variabilité des données.

Quant aux résidus de la déviance, ils semblent être bien distribués, mais ont toutefois à plusieurs endroits le même signe de manière consécutive, ce qui indique une surestimation de la mortalité lorsque ces derniers sont négatifs à plusieurs reprises et une sous-estimation lorsqu'ils sont positifs. Nous pouvons néanmoins conclure que l'ajustement est fiable, au moins pour les âges allant jusqu'à 86 ans.

Même si les tests statistiques d'adéquation valident les ajustements, ces derniers ne nous semblent pas satisfaisants au vu de l'analyse des résidus, en particulier pour les âges supérieurs à 87 ans. En effet, en effectuant une extrapolation des taux de mortalité par ces méthodes, nous prenons le risque de sous-estimer la mortalité des individus. Nous adaptons la méthode de Coale et Kisker à notre jeu de données. Pour rappel, dans la méthode de Coale et Kisker, nous calculons les $\mu_x = -ln(1 - qx)$ (hypothèse de répartition uniforme des décès) des taux bruts, puis les $k_x = ln(\frac{\mu_x}{\mu_{x-1}})$.

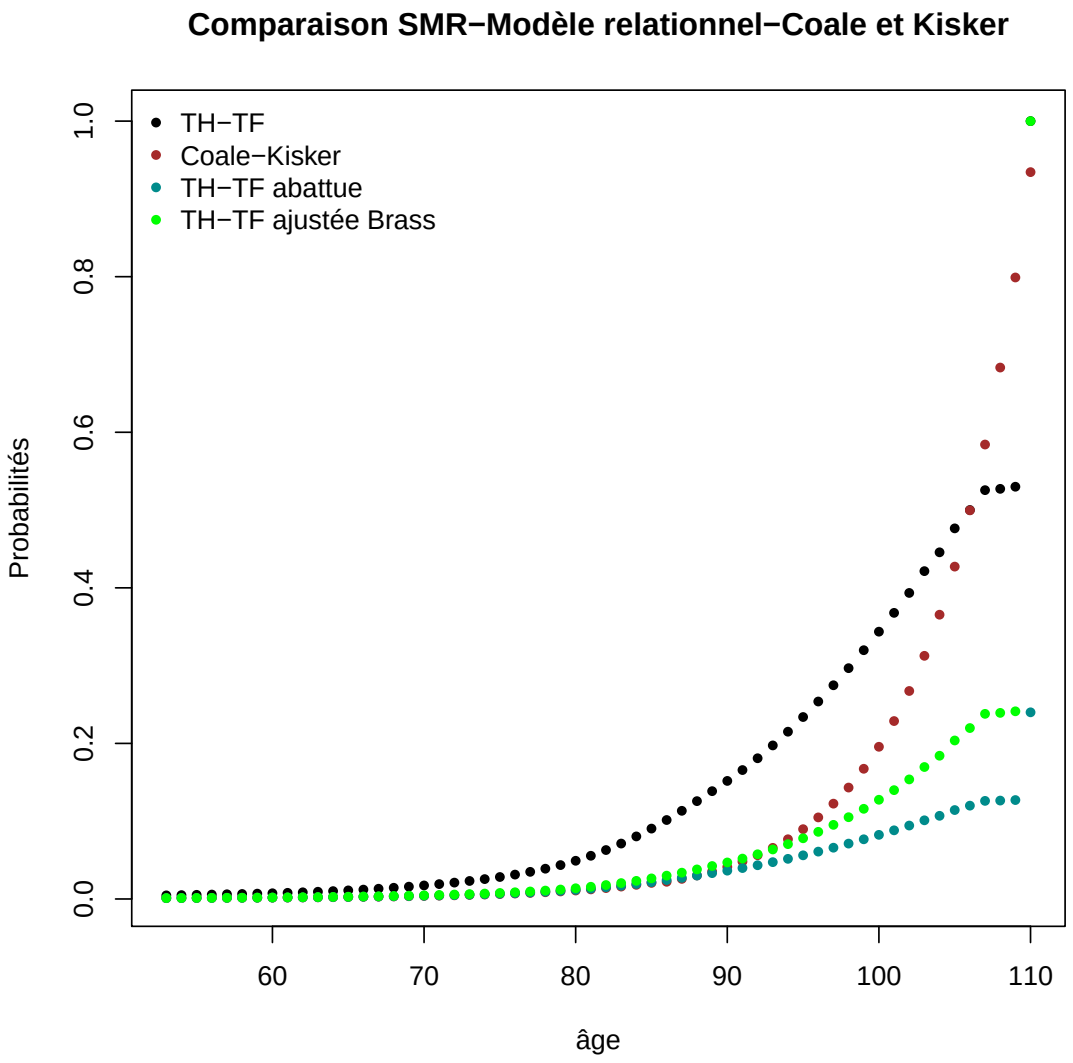

Illustration 5.11 : *Extrapolation par Coale et Kisker, les qx de la TH-TF, les qx de la TH-TF abattus par le SMR et les qx de la TH-TF abattus par le modèle relationnel*

Calcul des espérances de vie et des capitaux constitutifs de rente viagère

Afin de comparer les différents modèles entre eux, nous calculons les espérances de vie résiduelles à 60 ans et à 90 ans :

	A 60 ans	A 90 ans
SMR	36.41	11.97
Modèle relationnel	inf :32.64, centrale :35.43, sup :37.08	inf :8.5, centrale :10.72, sup :12.15
Coale et Kisker	34.58	8.91

De même nous comparons les montants des capitaux constitutifs de rente viagère d'âge initial 60 ans :

	a_{60}
SMR	27.26
Modèle relationnel	inf :25.47, centrale :26.85, sup :27.62
Coale et Kisker	26.55

Nous remarquons que les écarts, ne serait-ce que pour les espérances de vie résiduelles à 60 ans, semblent négligeables. Ils le sont encore plus pour les capitaux constitutifs de rente, gommés par l'effet d'actualisation (taux à 2%). Il sera pourtant intéressant de visualiser l'effet produit sur les simulations de compte de résultats par des variations a priori négligeables de la mortalité des autonomes.

5.2 Estimation de la probabilité de devenir dépendant

L'estimation des probabilités d'entrée en dépendance doit permettre d'obtenir une approximation raisonnable du nombre d'assurés qui tomberont en dépendance au cours d'une année.

L'étude statistique sur l'incidence nous a montré que l'entrée en dépendance semble fortement corrélée à l'âge de l'assuré : plus l'assuré est âgé, plus sa probabilité d'entrer en dépendance est grande. Ainsi, dans cette partie nous tâcherons d'estimer la probabilité annuelle de tomber en dépendance pour un assuré.

L'estimation des taux bruts présentant quelques irrégularités, nous effectuerons un lissage par la méthode de Whittaker Henderson. Nous tâcherons aussi d'extrapoler les taux d'incidences pour les âges élevés.

Dans un premier temps cette étude sera effectuée uniquement dans le cas de la dépendance totale et partielle. Nous verrons par la suite s'il est possible de différencier dépendance partielle et dépendance totale, voir d'estimer des probabilités de passage entre dépendance partielle et dépendance totale.

5.2.1 Distinction Homme Femme

La distinction Homme Femme dans l'étude des taux d'incidence semble significative.

A l'observation des fonctions de survie (cf. Illustration 5.12), nous constatons que la courbe de survie des hommes est, presque durant toute la période étudiée, inférieure à la courbe de survie des femmes. Cela indique que les femmes vivent plus longtemps sans dépendance que les hommes. Néanmoins, à en juger par la proximité des deux courbes, effectuer une segmentation homme femme ne ferait qu'augmenter la variance de notre estimation et la rendrait moins fiable.

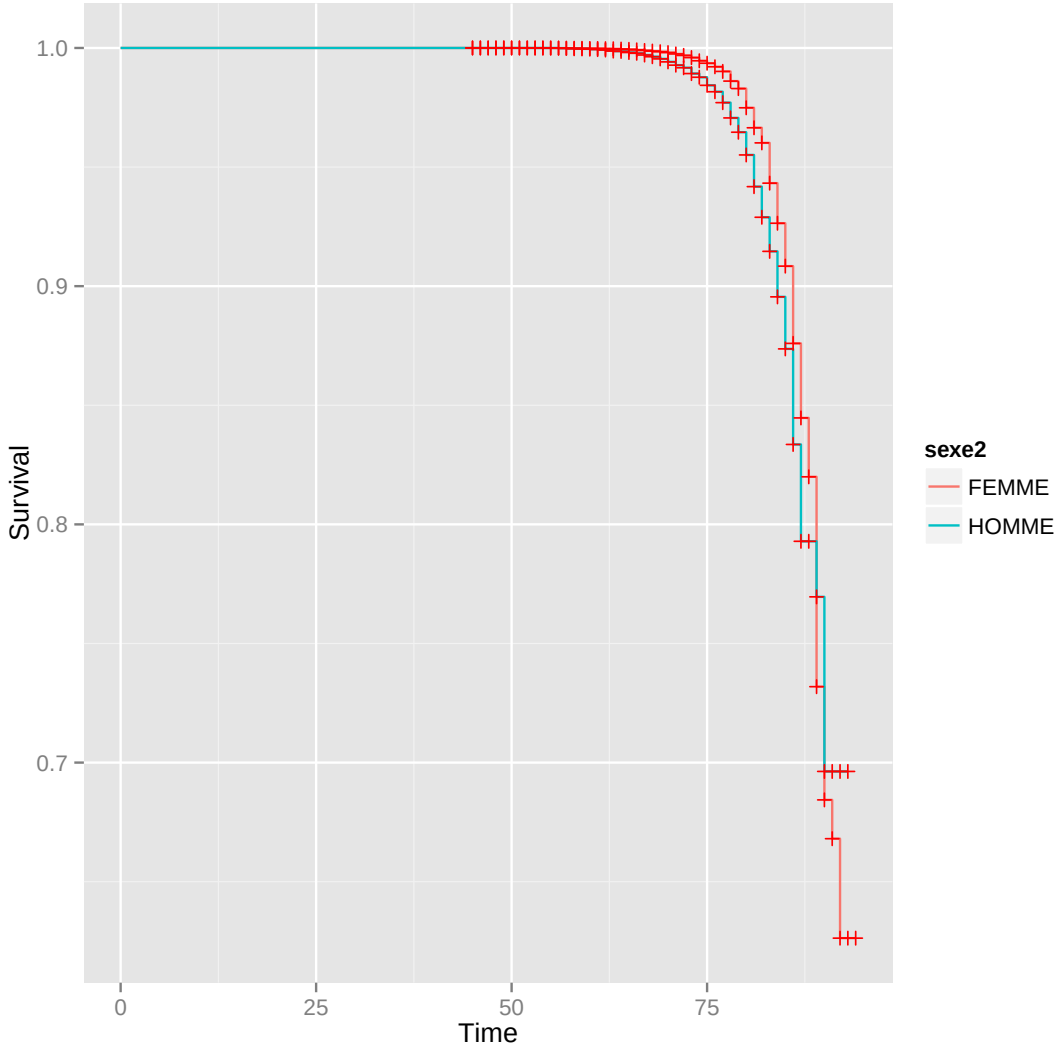

Illustration 5.12 : *Estimation de l'incidence par Kaplan Meier pour les Femmes et pour les Hommes*

Cette courbe de survie présente une convexité relativement forte, ce qui signifie que la durée de survie sans dépendance diminue plus que linéairement avec l'âge.

5.2.2 Estimation des taux d'incidence bruts

Nous déduisons de la fonction de survie les taux d'incidence :

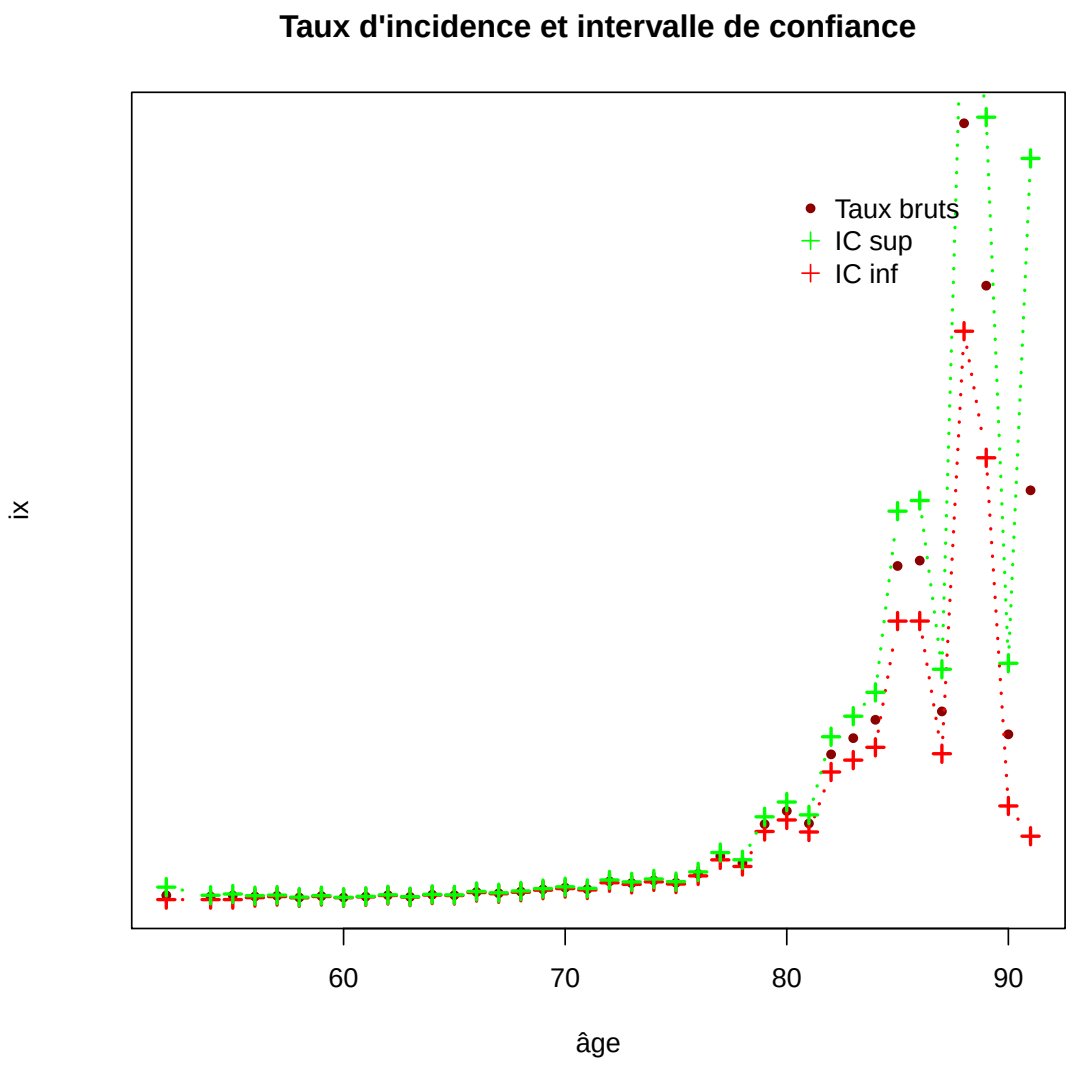

Illustration 5.13 : *Estimation de l'incidence par Kaplan Meier*

Nous remarquons clairement que comme précisé dans l'étude statistique sur les entrées en dépendance, il faut prendre beaucoup de précautions quant à l'interprétation de ces taux. En effet, nous avions supposé que l'estimation de ces taux en dehors de la plage 63 à 87 ans serait difficile. Nous constatons ici que d'une part les taux d'incidence avant l'âge de 63 ans semblent relativement faibles et que d'autre part les taux après 87 ans fluctuent fortement de même que l'intervalle de confiance qui s'élargit fortement pour ces derniers âges.

5.2.3 Lissage avec la méthode de Whittaker Henderson

Tout en gardant à l'esprit le problème posé par l'estimation des probabilités pour les âges situés en dehors de l'intervalle 63-87 ans, nous effectuerons dans un premier

temps un lissage sur les données avec la méthode de Whittaker Henderson. Le paramètre Lambda choisi est égal à 30 (pour rappel : $M = \sum_{i=1}^{n} w_i(h_i - g_i)^2 + \lambda \sum_{i=1}^{n-k} (\Delta^k g)^2$, le paramètre λ permettant d'effectuer l'arbitrage entre fidélité aux données et régularité. Ici ce paramètre est choisi volontairement bas, afin de conserver une importante fidélité aux données).

Nous obtenons ainsi une nouvelle courbe de taux lissés, que nous comparons avec les taux non lissés :

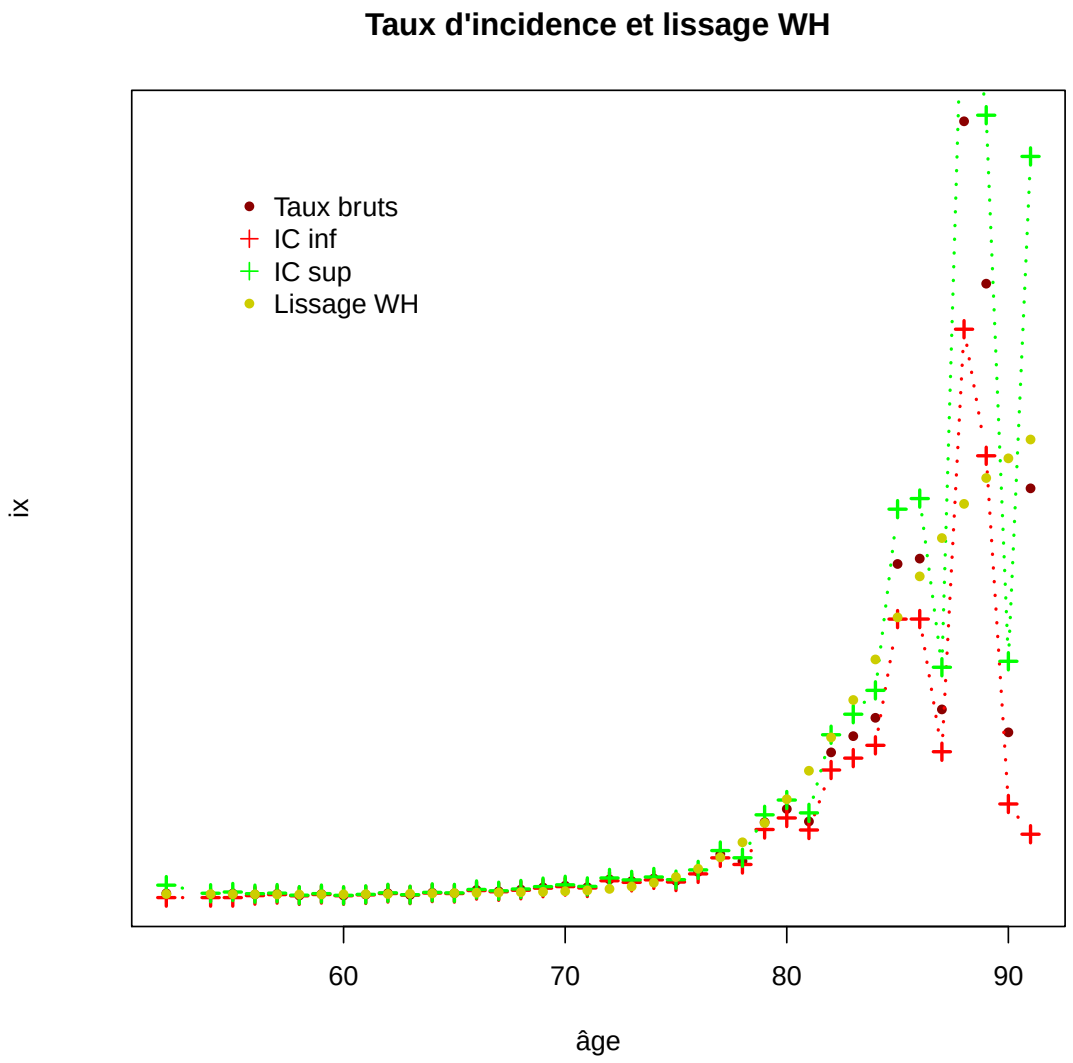

Illustration 5.14 : *Représentation des taux d'incidence bruts et lissés par la méthode Whittaker Henderson*

5.2.4 Validation de l'ajustement

Nous vérifions notamment la cohérence de la réestimation du nombre d'entrées en dépendance grâce aux taux d'incidence lissés :

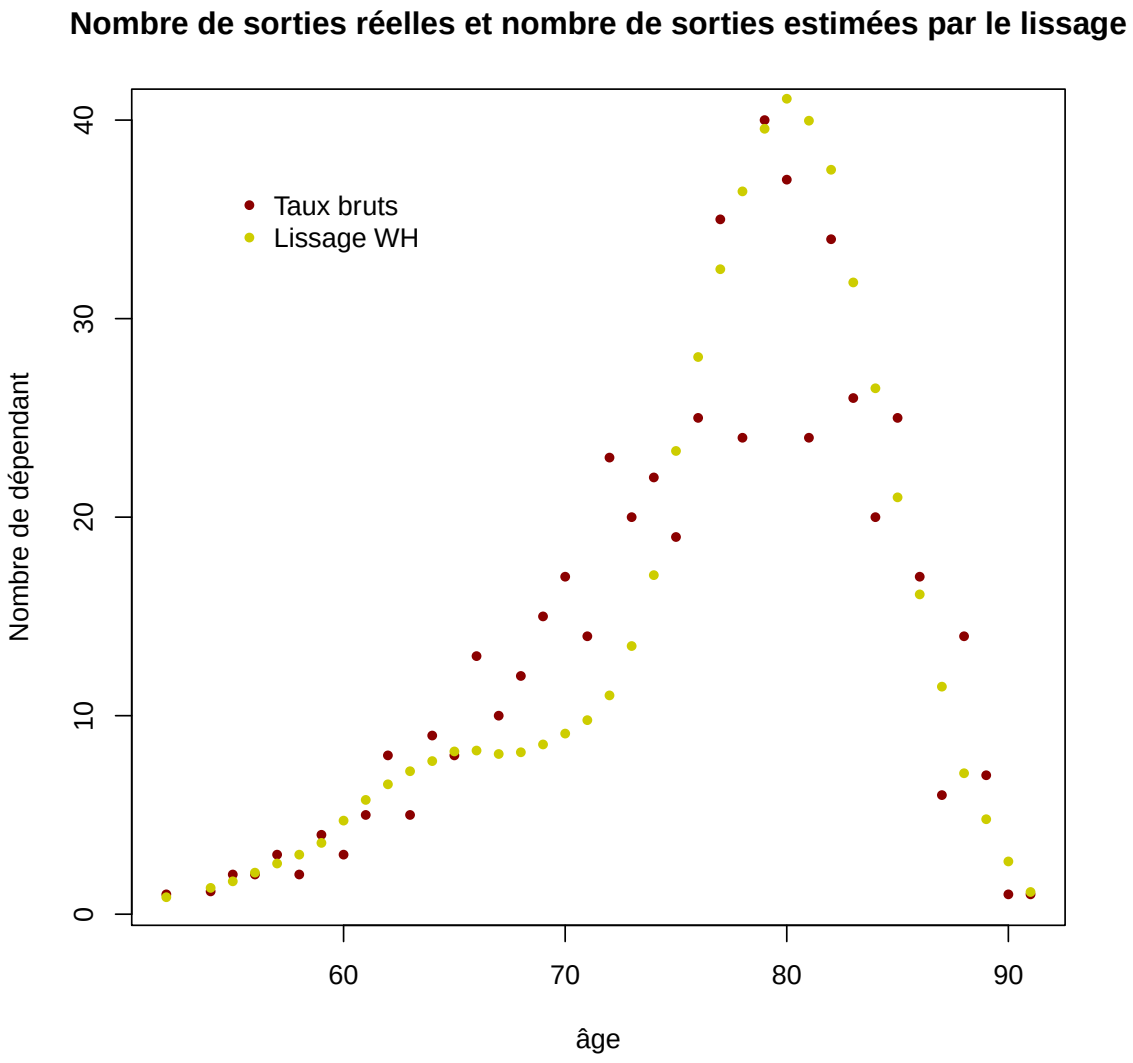

Illustration 5.15 : *Représentation du nombre de dépendants estimés par les taux bruts et lissés par la méthode de Whittaker Henderson*

Nous constatons que l'ajustement semble cohérent avec les observations.

5.2.5 Autre idée de modélisation

Une idée de modélisation, liée à l'abattement de table réalisé précédemment, serait de considérer que la mortalité et l'incidence sont expliquées par les mêmes causes, à savoir l'état de santé, qui se détériore avec l'âge, et d'éventuels accidents. L'intérêt d'une telle méthode est qu'elle permet de déterminer les taux d'incidence pour n'importe quel âge.

Cette méthode est équivalente à la régression logistique utilisée en épidémiologie afin de déterminer la survie d'individus malades comparée à celle d'individus sains, en fonction de certaines variables.

Elle a été utilisée par l'INSEE[3] pour réaliser des projections concernant le

3. Duée M., Rebillard C.

nombre de dépendants futurs et se base sur l'approche utilisée par Breuil et al (1998).

Si la mortalité est considérée comme une fonction d'une fonction dépendante de l'âge et reflétant la santé de l'individu, alors nous avons :

$$q_x = f_q(h(x))$$

où h(x) est la "fonction de santé" de l'individu dépendant de l'âge.
De même, si nous considérons que l'incidence est fonction de la fonction h, alors :

$$i_x = f_i(h(x))$$

Si nous supposons que $f_q^{-1} = logit()$ et $f_i^{-1} = logit()$ Nous avons donc le modèle :

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

et donc :

$$i_x = \frac{1}{1 + exp(-\alpha + \beta logit(q_x^{ref}))}$$

Il s'agit donc de trouver les paramètres α et β qui minimisent :

$$\sum_x |\hat{i}_x - \frac{1}{1 + exp(-\alpha + \beta logit(q_x^{ref}))}|$$

Nous obtenons :

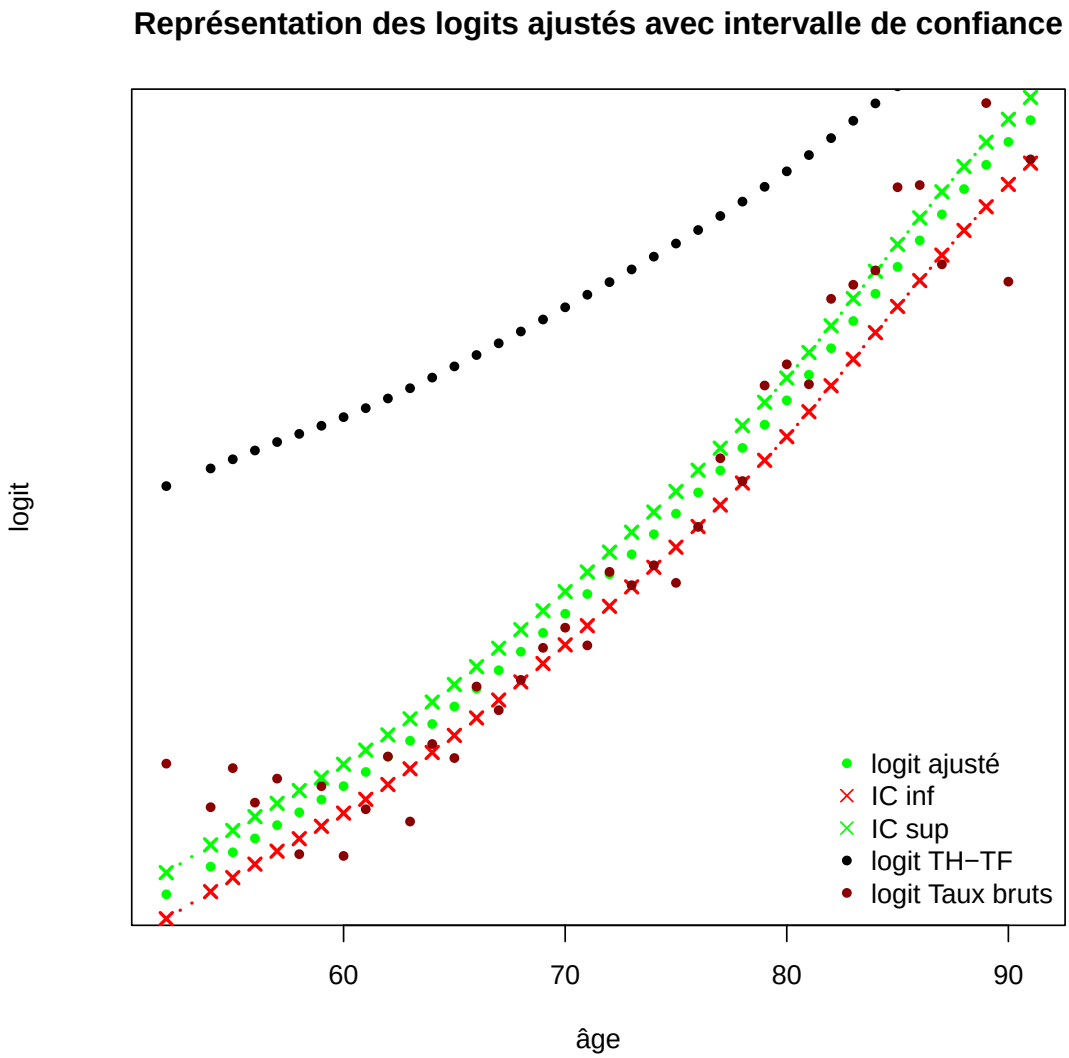

Illustration 5.16 : *Abattement des logits de la TH*

L'ajustement est satisfaisant, avec un R^2 égal à 0.94 et une p-value pour le test de Fisher inférieure à 2.2e-16.

De la même façon, nous calculons l'ajustement sur l'intervalle de confiance des taux d'incidence.

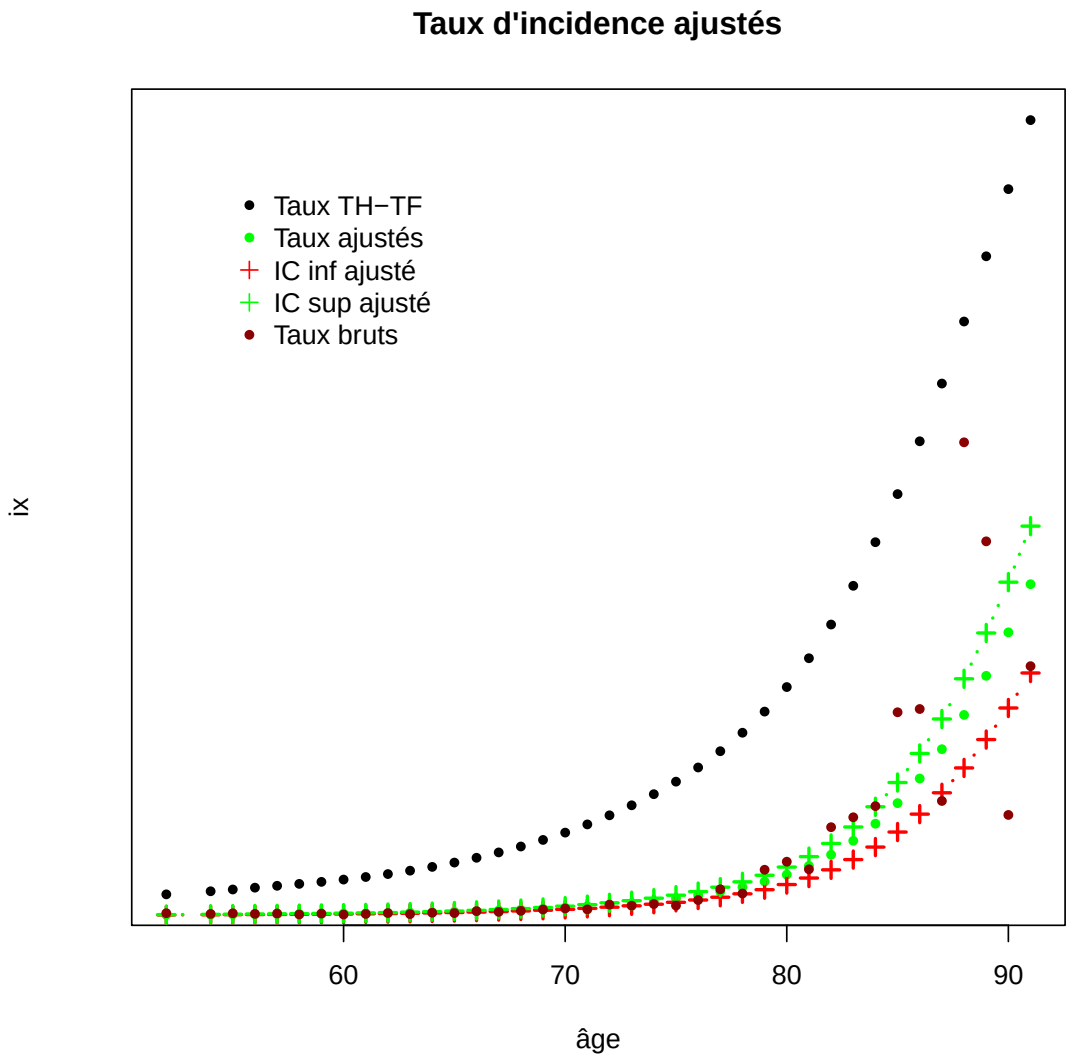

Illustration 5.17 : *Abattement de la TH, taux bruts d'incidence et taux de la TH*

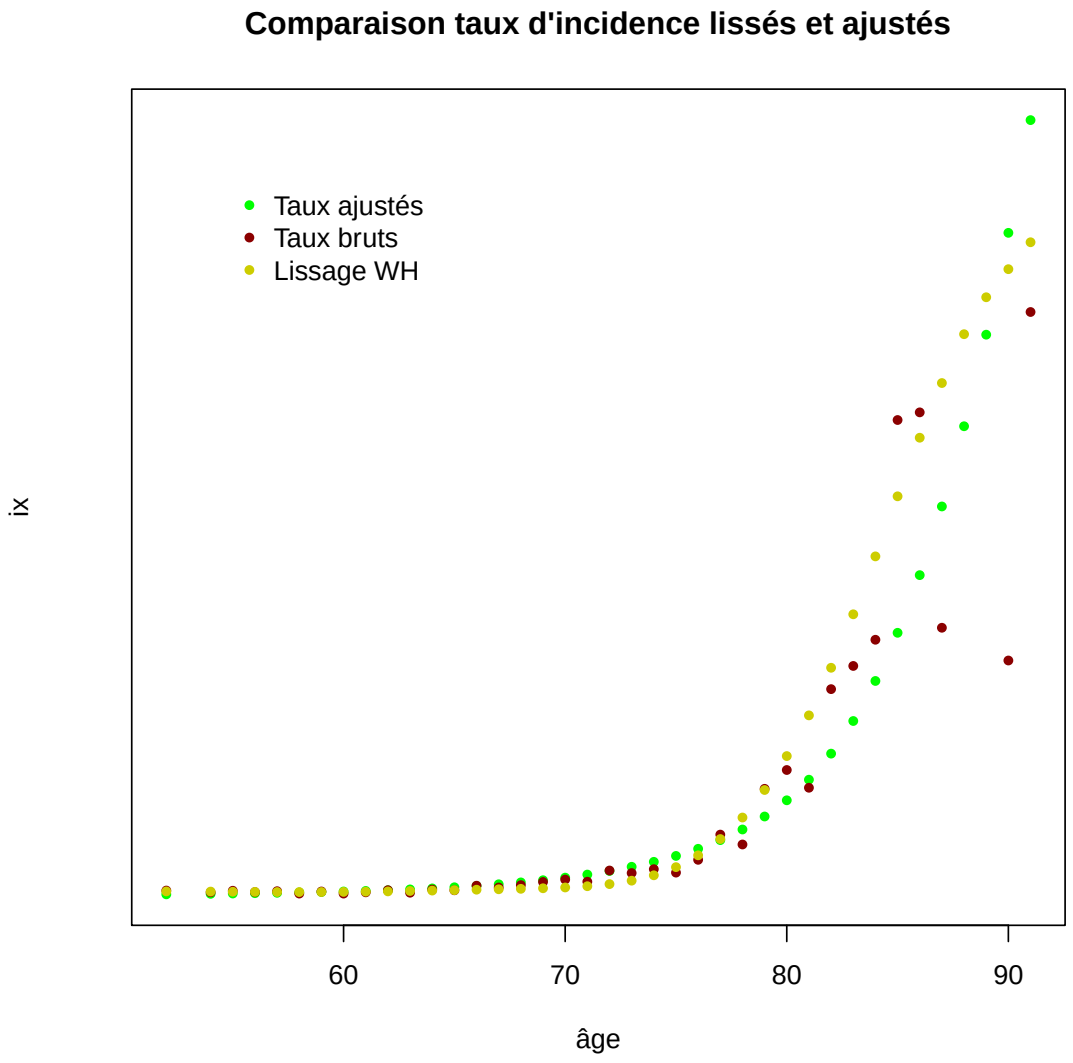

Illustration 5.18 : *Abattement de la TH et lissage par Whittaker, taux bruts, taux de la TH*

	Lissage par Whittaker	Ajustement logistique de la TH
Khi 2	0.03	0.03
MAPE	2.2	2.46
SMR	1.01	0.99

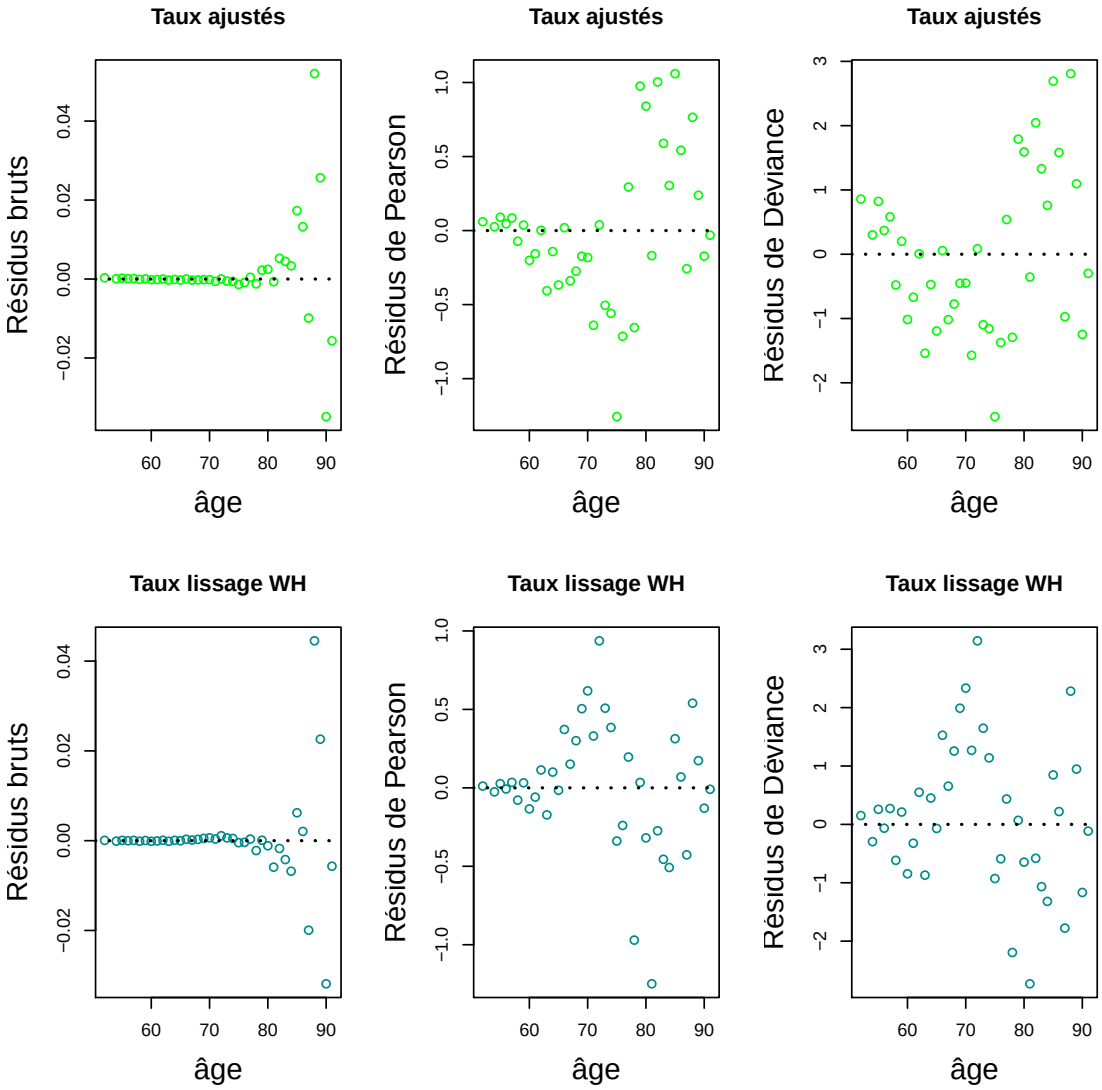

Illustration 5.19 : *Analyse des résidus*

Le modèle par ajustement logistique se distingue par sa prudence en comparaison du lissage par la méthode de Whittaker Henderson. Toutefois l'écart entre les deux modèles est très faible à en juger par les tests d'adéquation.

De manière générale pour les deux modèles, les tests d'adéquation permettent de valider la modélisation, avec un Khi2 proche de 0, un MAPE faible et un SMR proche de 1.

Néanmoins, le modèle avec lissage ne permet pas une extrapolation immédiate a contrario du modèle d'ajustement logistique qui offre l'avantage d'une extrapolation directe, et semble indiquer des taux d'incidence convexes avec l'âge, ce qui paraît plus cohérent avec la réalité.

L'analyse des résidus bruts confirme un mauvais ajustement pour les âges aux alentours de 90 ans, et ce pour les deux modèles.

Les résidus de Pearson ne décèlent aucune valeur anormale pour chacun des deux ajustements, puisqu'aucun résidu n'est plus grand que 2 ou plus petit que -2.

Les résidus de la déviance indiquent quant à eux que les deux ajustements se

caractérisent par des périodes de surestimation des taux puis des périodes de sous-estimation des taux. Cet effet est plus marqué pour l'ajustement logistique. Cet effet, qui tend à faire valider les ajustements en moyenne, peut être problématique car l'ajustement est bon en moyenne, mais pas ponctuellement. Toutefois, à en juger par l'observation du graphique, ce phénomène se produit principalement pour les estimations avant 80 ans. Or, à en juger par les résidus bruts, pour ces âges là, l'estimation est très proche de la réalité. La surestimation et sous-estimation ponctuelles observées sur les résidus de la déviance ne sont donc pas gênantes, puisqu'il s'agit de sur et sous-estimations faibles.

Comme signalé précédemment, le nombre d'observations de sorties étant faible en dehors de l'intervalle 63-87 ans, nous nous interrogeons sur la pertinence d'un ajustement du type

$$logit(i_x) = \alpha + \beta logit(q_x^{ref}) \ \forall \ x \ \in \ [52, 91]$$

Nous allons comparer cette modélisation avec un ajustement où $x \ \in \ [63, 87]$ et nous ferons une comparaison des deux modèles.

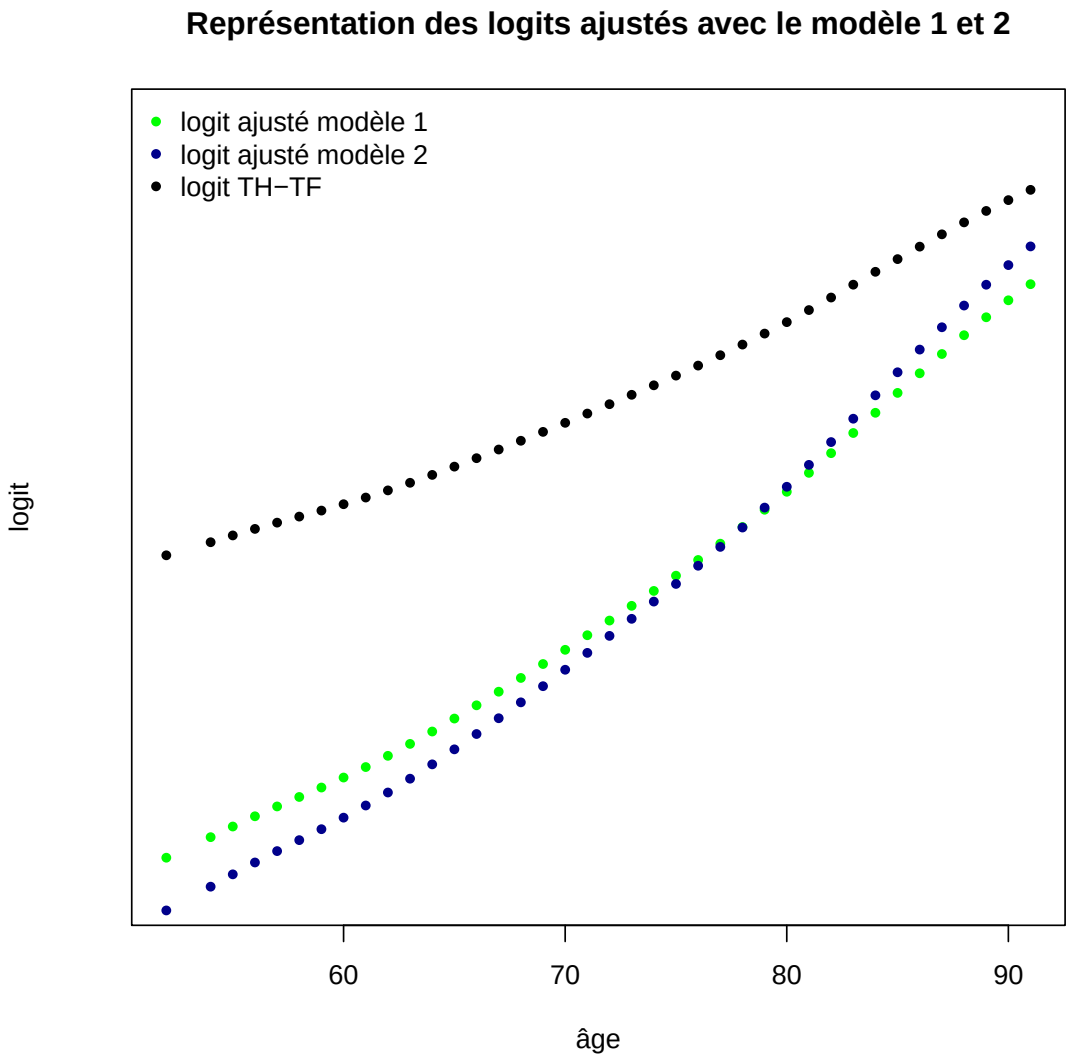

Illustration 5.20 : *Comparaison des deux ajustements, avec le logit des taux bruts, le logit des taux obtenus avec le premier ajustement et le logit des taux du second ajustement*

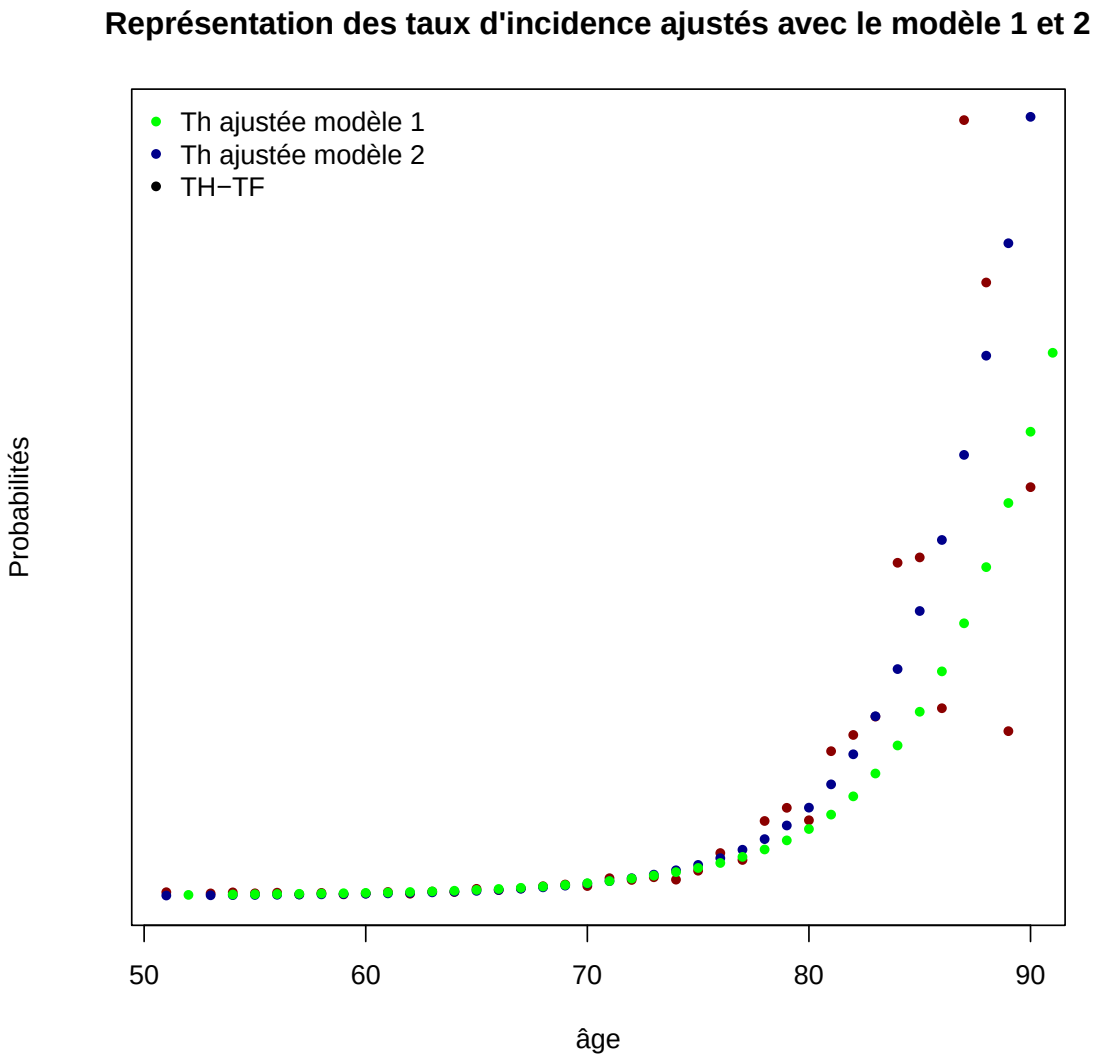

Illustration 5.21 : *Comparaison des deux ajustements, avec les taux bruts, les taux obtenus avec le premier ajustement et les taux du second ajustement*

Nous obtenons pour ce modèle les tests d'adéquation suivant :

khi2	MAPE	SMR
0.04	2.56	1

Le Khi2 et le MAPE restent relativement faibles, le SMR de ce modèle est plus élevé que le SMR du modèle précédent où les taux étaient ajustés pour tous les âges disponibles. Cela indique une moins grande prudence, mais comme il se rapproche de 1, cela indique un meilleur ajustement global. Le SMR de ce modèle est plus élevé en raison de la sous-estimation de l'incidence pour les âges avant 80, or, c'est là où l'effectif est le plus important. Néanmoins ce modèle présente une convexité plus forte, indiquant une croissance de la probabilité d'entrée en dépendance plus forte que pour le premier modèle.

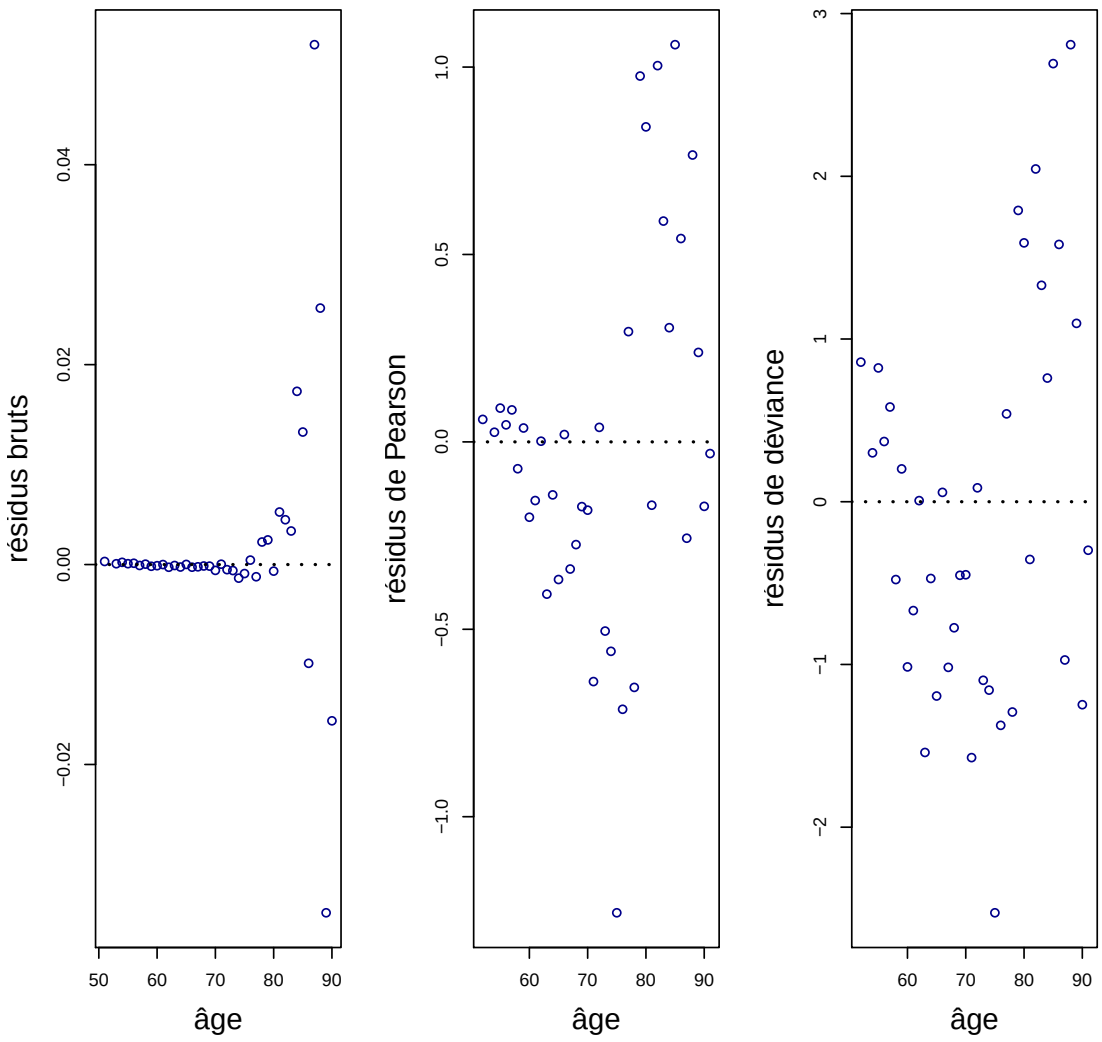

Illustration 5.22 : *Analyse des résidus de la deuxième modélisation*

Mesure du risque d'estimation

Principe Nous procédons maintenant à une mesure du risque d'estimation effectuée dans le cadre de l'ajustement des logits des taux d'incidence sur les logits de la TH-TF. Pour ce faire, nous utilisons la méthode proposée dans Kamega et Planchet (2010) [4] que nous appliquons à l'estimation de l'incidence et non pas de la mortalité.

Plus précisément, nous nous intéressons au risque systématique lié aux fluctuations d'échantillonage, c'est à dire le risque présenté par les données uniquement. Le risque de modèle n'est pas considéré ici, au contraire l'hypothèse est que le modèle spécifié est juste (i.e. nous avons bien $logit(i_x) = \alpha + \beta logit(q_x^{ref}) + \epsilon_x$ où ϵ_x est un bruit blanc). La détermination des paramètres se fait par moindres carrés ordinaires, ce qui représente la meilleure méthode d'obtenir les paramètres α et β, car non biaisée. Pour mesurer le risque d'estimation sur les taux de décès ajustés, nous allons générer de manière aléatoire des taux bruts pour observer l'impact sur

4. Kamega et Planchet (2010) Mesure du risque d'estimation associé à une table d'expérience

l'estimation des paramètres α et β. Cette méthode s'appuie sur la simulation des résidus, à savoir :

$$e_x = logit(i_x) - (\alpha + \beta logit(q_x^{ref}))$$

Dans un premier temps, et puisque nous faisons l'hypothèse que le modèle est bien spécifié, nous supposons que ces résidus suivent une loi Normale, de paramètre inconnu a priori. Puis nous vérifions cette hypothèse de normalité grâce à un test de Shapiro Wilk. Si ce test nous permet d'accepter l'hypothèse de normalité des résidus, nous calculons ensuite la moyenne et l'écart type sans biais de ces résidus, que nous notons μ_e et σ_e. Ces paramètres sont ensuite utilisés pour générer différents scénarios, que nous noterons k, avec $k \in [1, K]$. Nous avons :

$$logit(i_x)^k = \alpha + \beta logit(q_x^{ref}) + \epsilon_x^k$$

où

$$\epsilon_x^k \sim N(\mu_e, \sigma_e)$$

Ensuite, pour chaque scénario k, les paramètres α et β sont réestimés tels que :

$$logit(i_x)^k = \alpha^k + \beta^k logit(q_x^{ref}) + \epsilon_x^k$$

Puis nous en déduisons :

$$i_x^k = \frac{1}{1 + \exp[-(\alpha^k + \beta^k logit(q_x^{ref}))]}$$

Application aux deux modèles ajustés

Comme dans l'article, nous utilisons 15 000 simulations, ie K=15 000.

Pour rappel, nous obtenons des coefficient α et β dans le premier modèle, que nous appelons modèle 1, pour les âges allant de 53 à 92 ans, avec une p value pour le test de Fisher inférieure à $< 2e^-16$. De plus, le R^2 est de 94%. De même nous obtenons des paramètres α et β pour le deuxième modèle, que nous appelons modèle 2, avec une p value pour le test de Fisher de nullité des paramètres inférieure à $< 2e^-16$ et un R^2 égal à 97.01%.

La p-value du test de Shapiro Wilk pour le test de normalité des résidus pour le modèle 1 est de 0.66 et pour le modèle 2 de 0.34, indiquant que nous ne pouvons pas rejeter l'hypothèse de normalité des résidus.

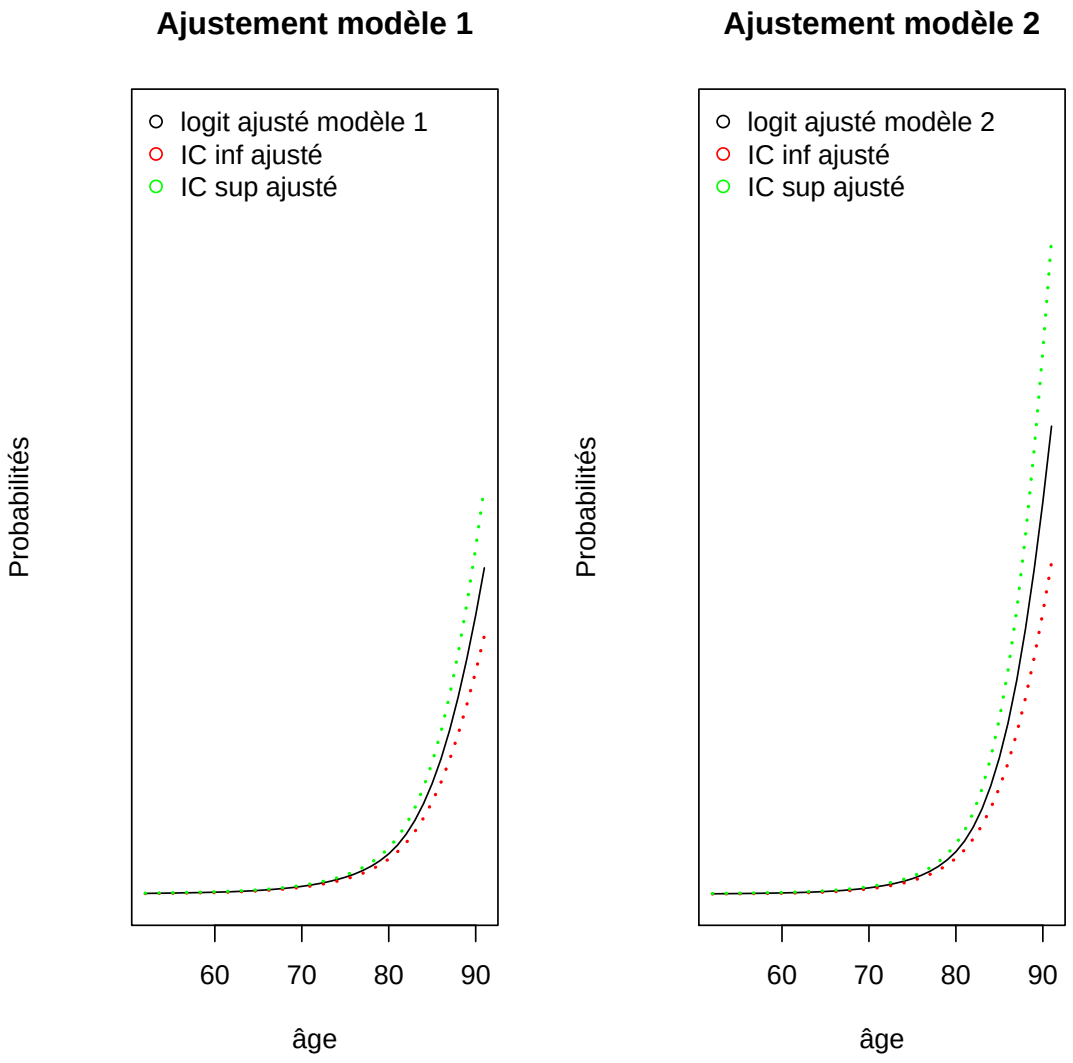

Illustration 5.23 : *Taux d'incidence ajustés et taux d'incidence simulés (avec intervalle de confiance à 95% et 5%)*

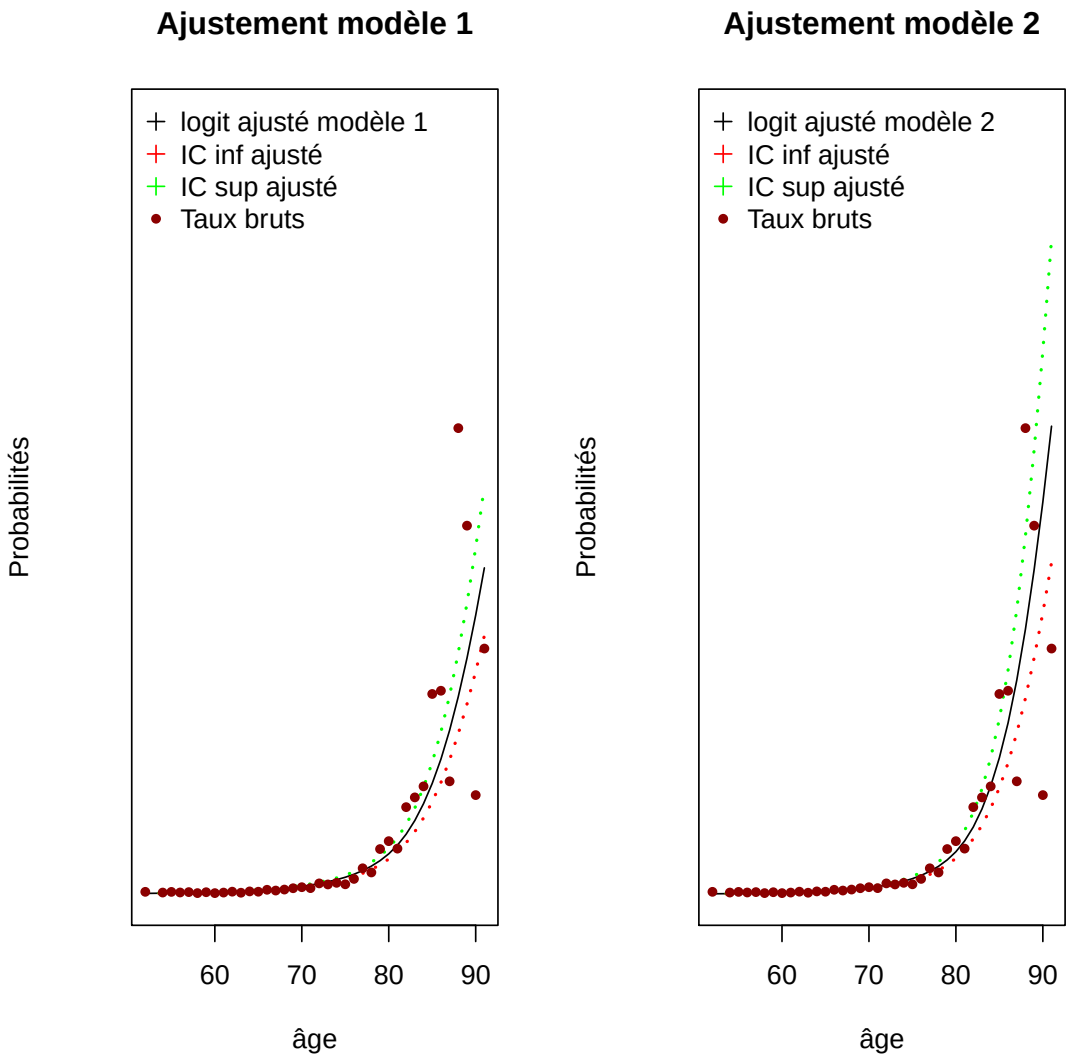

Illustration 5.24 : *Taux d'incidence ajustés et taux d'incidence simulés (avec intervalle de confiance à 95% et 5%) et taux bruts*

Nous remarquons alors que le modèle 2 entraîne des taux d'incidence sensiblement plus élevés en comparaison du modèle 1 (cf. Illustration 5.23).

L'espérance de vie résiduelle sans dépendance à l'âge 53 est comparée pour les deux modèles :

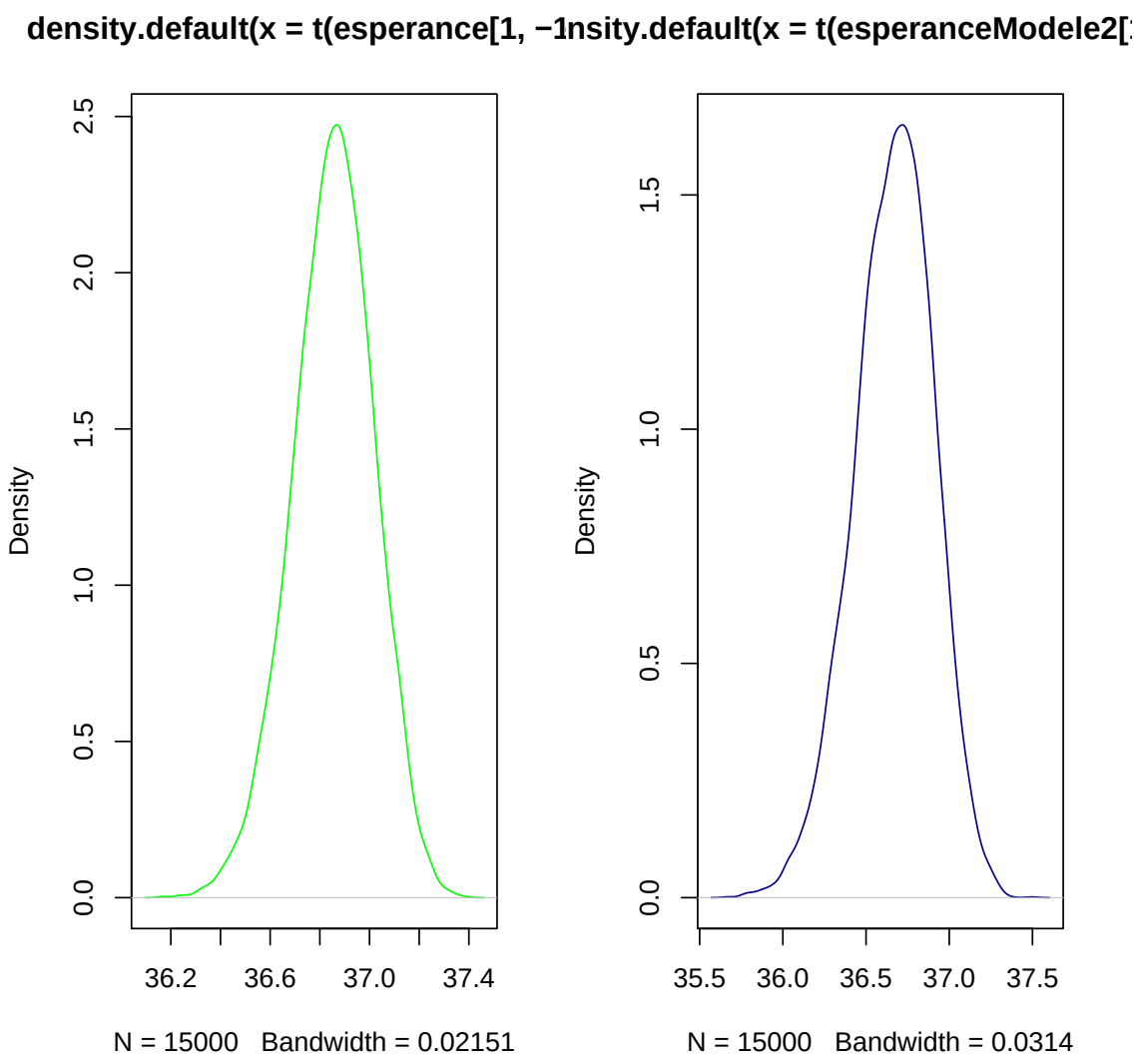

Illustration 5.25 : *Espérance de survie partielle sans dépendance dans le modèle 1 et dans le modèle 2*

Nous constatons que la distribution de l'espérance de vie résiduelle sans dépendance pour les deux modèles est relativement proche. Néanmoins, la densité du deuxième modèle est plus "large" : elle prend des valeurs plus extrêmes que celle du deuxième modèle, avec une probabilité plus élevée.

Afin d'extrapoler la table d'entrée en incidence aux âges non observés, nous prolongeons l'estimation obtenue grâce aux modèles relationnels et à la connaissance des probabilités de décès de la TH-TF pour les âges élevés, nous permettant ainsi d'obtenir par ajustement les taux d'incidence pour ces âges. Nous utiliserons principalement le premier modèle, les différences entre les deux modèles n'étant pas flagrante, mais le premier modèle estime des taux d'incidence plus élevés que le deuxième modèle pour les âges avant 80 ans. Or lors des projections, nous simulerons la durée de vie d'individus ayant en moyenne 62 ans. Il nous faut donc des taux prudents pour les âges aux alentours de 62 ans.

Cependant, signalons que cette extrapolation pourrait ne pas correspondre à la

réalité : il s'agit d'une hypothèse prise et qui suppose que le logit des taux ajustés va rester linéairement proportionnel au logit de la table TH-TF et selon le même ajustement que celui effectué pour les âges entre 52 et 91 ans. Cette hypothèse implique aussi, comme nous pouvons le constater sur le graphique suivant qu'a un âge très élevé, le logit du taux d'incidence va être égal au logit du taux de mortalité de la TH-TF. Ainsi la probabilité d'entrée en dépendance à cet âge sera quasiment égale à la probabilité de décéder. Une autre hypothèse possible est que le taux d'incidence sera bien inférieure. Mais il s'agit d'une hypothèse peu prudente, nous obligeant à déroger à l'ajustement effectué, sans raisons valables. L'autre hypothèse extrême serait de penser que cette probabilité serait bien supérieure.

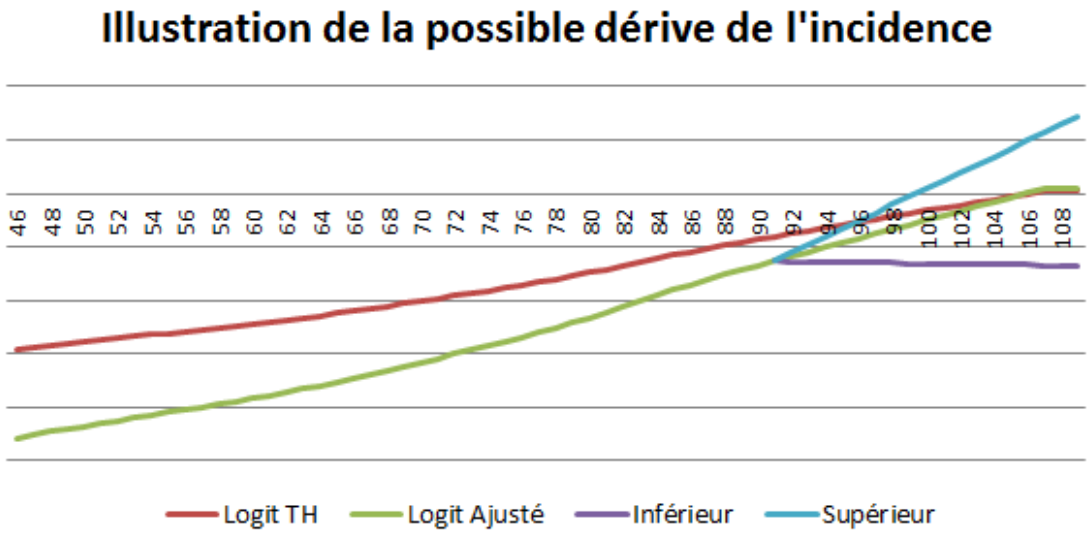

Illustration 5.26 : *Illustration de la possible dérive de l'incidence*

Néanmoins, nous relativisons ce choix. Le point d'incertitude étant principalement après 95 ans, nous pouvons supposer que lorsqu'un individu entrera en dépendance à cette âge, il n'y demeurera pas longtemps, ainsi, l'impact de son entrée ou pas ne sera pas important dans nos résultats. C'est ce que nous vérifierons avec l'étude de la durée de vie des dépendants par la suite.

5.2.6 Passage de dépendance partielle à dépendance totale

Nous faisons l'hypothèse communément admise, qu'aucune rémission n'est possible. Ainsi, la probabilité de passage de dépendance totale vers dépendance partielle est nulle, de même que celle de passage de dépendance totale ou partielle vers l'autonomie.

La probabilité de passage de dépendance partielle vers la dépendance totale, quant à elle ne peut pas être supposée nulle. En effet, nous disposons de multiples observations de ce phénomène. Néanmoins, nous ne disposons pas de suffisamment de données pour effectuer une loi de passage de dépendance partielle vers dépendance totale pour chaque âge.

Nous proposons ainsi un taux de passage de dépendance partielle vers dépendance totale, que nous noterons dans la suite $i_x^{DP->DT}$, égal à :

$$i_x^{DP->DT} = \frac{\textit{Nombre de passage DP->DT}}{\textit{Effectif en DP}}$$

5.2.7 Entrée en dépendance partielle ou dépendance totale

Dans ce qui précède, nous avons déterminé la probabilité d'entrée en dépendance en fonction de l'âge. Toutefois, il est important de faire la distinction entre entrée en dépendance partielle et entrée en dépendance totale.

Plutôt que d'estimer deux lois d'entrée en dépendance pour la dépendance partielle et totale, nous n'en estimons qu'une seule, pour la dépendance partielle et totale. Mais, par une observation empirique sur les données, nous constatons que la proportion d'individus en dépendance totale par rapport aux individus en dépendance partielle reste stable à travers les différentes années d'observation. Notons ratioDT cette proportion d'individus en dépendance totale au cours des différents années. Ainsi, un individu d'âge x a une probabilité égale à i_x d'entrer en dépendance, et une probabilité égale à ratioDT d'entrer en dépendance totale.

Nous validons ce ratio grâce à un test du Khi 2 (avec satisfaction des conditions de Cochran [5]) avec :

- H0 : la proportion d'individus en dépendance totale est égale à ratioDT chaque année ;
- H1 : la proportion d'individus en dépendance totale est distinctement différente de ratioDT chaque année.

Le résultat de ce test est l'absence de rejet de H0, au seuil de significativité 5%.

5.2.8 Entrée en dépendance pour cause de maladie, démence ou accident

Afin de savoir si un assuré entre en dépendance pour cause de maladie, démence ou accident, nous calculons le rapport entre nombre d'assurés entrés en dépendance pour ces différentes causes et le nombre d'assurés entrés en dépendance. Ces probabilités vont nous servir dans la simulation des assurés, plus précisément dans la prise en compte des délais de carence.

En effet, puisque nous simulons des assurés ayant souscrit juste avant le début de la simulation, selon la cause de leur dépendance, il est possible que ces assurés ne puissent pas bénéficier de la prestation en raison des délais de carence. Il est ainsi important de connaître la cause d'entrée en dépendance.

5. Cochran (1954), toutes les classes doivent avoir une valeur théorique non nulle et 80 % des classes doivent avoir une valeur théorique supérieure ou égale à 5

Notons :

$$ratioMaladie = \frac{Nombre\ d'assurés\ entrés\ en\ dépendance\ pour\ cause\ de\ maladie}{Nombre\ d'assurés\ dépendant}$$

,

$$ratioDemence = \frac{Nombre\ d'assurés\ entrés\ en\ dépendance\ pour\ cause\ de\ démence}{Nombre\ d'assurés\ dépendant}$$

et

$$ratioAccident = 1 - ratioMaladie - ratioDemence$$

5.3 Probabilités de survie en dépendance

L'estimation des probabilités de survie en dépendance doit permettre d'avoir une approximation de la durée de survie en dépendance d'un individu, et donc de savoir, puisque nous sommes dans le cas de rente, pendant combien de temps la compagnie va devoir verser une rente à l'assuré. Il faut que cette estimation soit prudente, c'est à dire qu'il est préférable de sous-estimer les probabilités de décès en dépendance, afin de surestimer la durée de survie en dépendance plutôt que de la sous estimer. Nous tenterons d'effectuer la meilleure estimation possible.

Cette estimation va être effectuée grâce à l'estimateur de Kaplan Meier. Par ailleurs, nous avions constaté dans l'étude statistique qu'il était pertinent d'étudier la durée de survie en dépendance en fonction de l'âge à l'entrée en dépendance. Comme nous l'avions vu, cette segmentation s'explique par le fait que les assurés entrant en dépendance précocement y entrent pour une maladie dont la durée de survie est faible, par exemple un cancer. D'un autre côté les personnes plus âgées entrent en dépendance pour des maladies principalement neurologiques, affectant peu leur durée de survie. Nous vérifierons par des tests la pertinence d'une étude en fonction de l'âge à l'entrée en dépendance.

Signalons de plus que nous ne faisons pas une étude de la survie par sexe, même si ce critère est certainement pertinent. Nous ne disposons pas de suffisamment de données afin de pouvoir effectuer cette segmentation et garder une robustesse de l'estimation suffisante.

Par ailleurs, puisque nous faisons l'hypothèse que la durée de survie en dépendance dépend de la maladie causant la dépendance, il aurait pu être judicieux d'évaluer directement la durée de survie en fonction de la maladie, et non de l'âge à l'entrée. Toutefois cette approche a ses limites. En effet, il faut pouvoir estimer pour chaque âge la probabilité qu'un individu tombe pour tel ou tel type de dépendance. Nous utilisons dans la partie précédente une approximation, car l'enjeu est simplement de vérifier qu'il n'y a pas de carence, mais cette approximation ne serait pas suffisamment précise pour effectuer ce genre de segmentation ayant un impact sur la

durée de survie de l'individu. De plus, il n'est pas aisé de classifier les maladies. Par exemple, s'il est vérifié qu'un cancer augmente plus les probabilités de décès qu'une démence, tous les cancers ne sont pas identiques, certains ayant une durée de survie plus longue que d'autres.

L'interpolation entre les dates où aucune observation n'est disponible sera effectuée par spline cubique sur les taux estimés. Le lissage quant à lui sera effectué par la méthode de Whittaker Henderson, avec un paramètre de lissage de 50.

5.3.1 Justification de la segmentation par âge

L'idée va être de segmenter l'estimation de la fonction de survie pour différentes classes d'âge. En effet, puisque nous avons remarqué que la durée de survie en dépendance semble dépendre de l'âge à l'entrée en dépendance, il semble justifié de vérifier cette hypothèse. Il serait préférable d'obtenir une loi pour chaque âge à l'entrée, mais cela est difficilement réalisable avec une base sinistres assureur, qui ne contient pas suffisamment de sinistres. C'est pourquoi nous tentons d'effectuer cette analyse en regroupant les observations en classe d'âge.

Pour ce faire nous effectuons dans un premier temps un test du log rank pour vérifier si l'hypothèse d'égalité de la distribution de survie en fonction de différents regroupements d'âges à l'entrée en dépendance doit être rejetée, dans un premier temps sans faire de distinction entre dépendance totale et dépendance partielle. Nous segmentons d'abord les observations en 3 classes, de telle façon que chaque classe ait le même nombre d'effectifs.

Sous l'hypothèse H0 d'égalité des fonctions de distributions dans les 3 groupes, au temps de décès i, la proportion attendue de décès parmi les individus est identique dans les 3 groupes.

Ainsi, nous calculons la statistique de test approchée du log rank :

$$\Phi = \frac{(D_1^{théorique} - D_1^{observé})^2}{D_1^{théorique}} + \frac{(D_2^{théorique} - D_2^{observé})^2}{D_2^{théorique}} + \frac{(D_3^{théorique} - D_3^{observé})^2}{D_3^{théorique}}$$

qui suit sous H0 une loi du Khi^2 à 2 degrés de liberté et avec le nombre de décès théorique dans le groupe j égal à :

$$D_j^{théorique} = \sum_i \frac{n_i^j * d_i}{n_i}$$

avec :

- n_i^j l'effectif sous risque du groupe j à l'instant i
- d_i le nombre de sorties observées dans tous les groupes à l'instant i
- n_i l'effectif sous risque total à l'instant i

Ensuite, nous calculons la p-value.

Elle correspond à la probabilité sous H0 que la statistique observée du Khi^2 soit dépassée. Autrement dit, c'est la probabilité sous H0 d'obtenir une une statistique aussi extrême que la valeur observée de l'échantillon. La p valeur se détermine par lecture sur un tableau de distribution du Khi^2 ou à l'aide d'un logiciel statistique.

Afin de rejeter ou accepter H0, un seuil de significativité α est fixé. Si $p \leq \alpha$ alors H0 est rejetée.

Dans notre cas, nous calculons la statistique à l'aide du logiciel R et du package "survival".

Nous obtenons alors une p-value égale à 1.53e-07. Nous pouvons donc rejeter l'hypothèse que les fonctions de distributions des 3 classes d'âge soient les mêmes (cf. Illustration 5.27). Toutefois, un examen visuel et un principe de prudence du fait d'un nombre d'observations restreint entraînant une variance assez forte dans l'estimation nous incite à diminuer le nombre de classes.

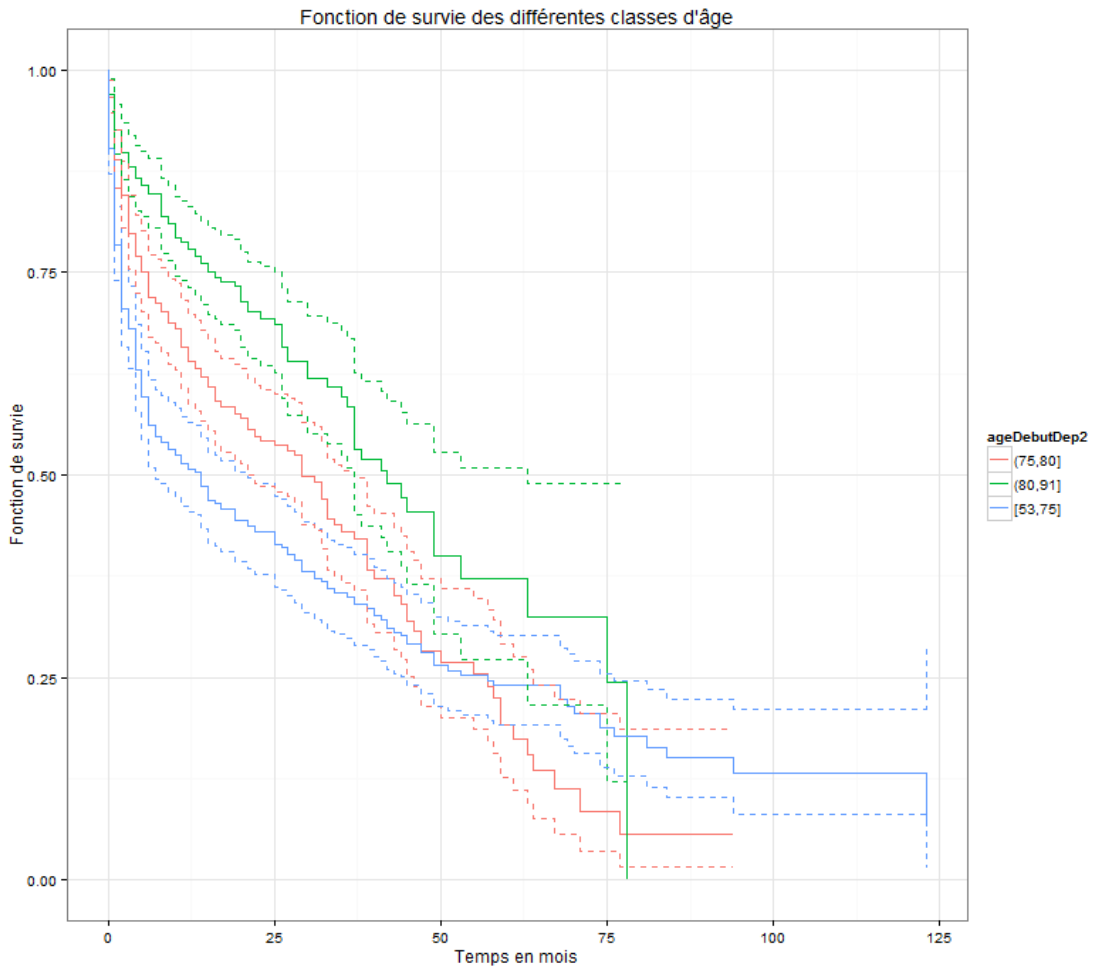

Illustration 5.27 : *Estimation de la fonction de survie pour les différentes classes d'âge*

En réduisant à deux le nombre de classes, nous obtenons une p valeur encore plus faible. De plus la variance de l'estimation a été grandement diminuée à en juger par l'intervalle de confiance (cf. Illustration 5.28).

Un autre essai a été effectué en créant des classes d'étendue égale. Toutefois

cette solution ne sur performe pas la première technique, c'est-à-dire la création de classes d'effectifs égaux, en matière de maximisation de la p valeur et donc de prise en compte maximale de l'hétérogénéité inter classe.

Notons aussi que les fonctions de survie se croisent, ce qui peut fausser le test du log rank. Nous décidons alors de retenir deux classes d'âges, afin d'une part d'avoir une variance d'estimation minimale et d'autre part d'éviter le croisement précoce des fonctions de survie.

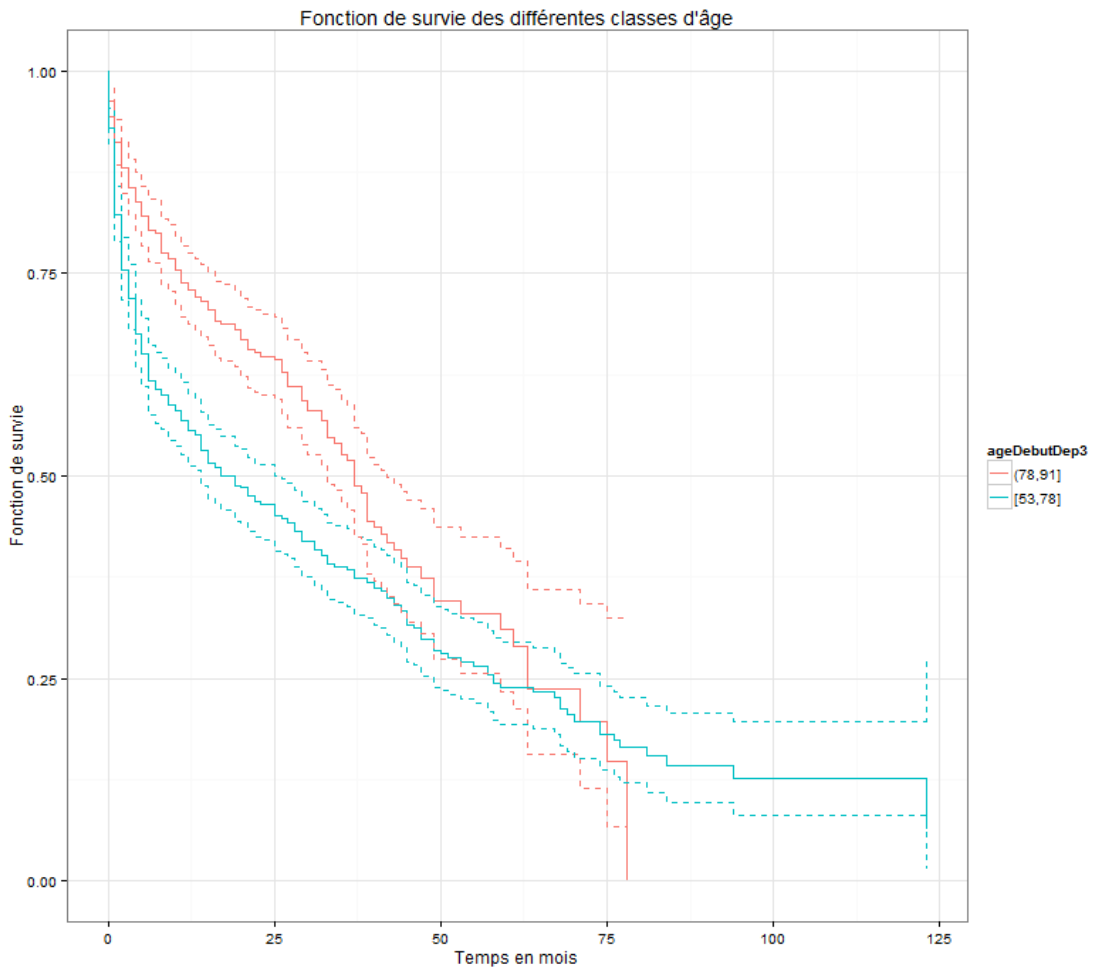

Illustration 5.28 : *Estimation de la fonction de survie pour les différentes classes d'âge*

Toutefois, cette segmentation a été effectuée manuellement. Même si les tests statistiques indiquent qu'il est pertinent d'étudier la fonction de survie en fonction de ces classes d'âges, rien n'indique qu'il s'agit là des meilleurs regroupements possibles afin de capter au mieux les différences de survie de la population étudiée.

L'utilisation d'arbres de régression, qui ont pour objectif de produire des groupes les plus homogènes possible du point de vue d'une variable à prédire, pourrait être pertinent dans ce type d'étude. Ici cette variable est la fonction de survie estimée par la méthode de Kaplan Meier. Nous présenterons en fin de partie les résultats obtenus par cette méthode. Nous déterminerons d'abord la loi de maintien en dépendance totale et partielle des individus par âges à l'entrée selon trois des classes utilisées

Classe d'âge :	[53,78]	(78,91]
Espérance de vie médiane :	16	35
Intervalle de confiance inférieur à 95% :	12	30
Intervalle de confiance supérieur à 95% :	25	42

pour le test du log rank précédent. Pour la dépendance partielle, en raison d'un nombre de données peu important, nous décidons de ne créer qu'une seule classe afin de limiter la variance de l'estimation.

5.3.2 Justification de la concaténation des deux bases

Puisque nous étudions conjointement la mortalité de deux bases sinistres, nous effectuons un test du log rank afin d'évaluer si la mortalité des dépendants des deux bases diffère ou non.

L'hypothèse est que la fonction de survie par âge à l'entrée est égale entre les dépendants du premier et du deuxième portefeuilles. Ici, nous ne prenons en considération que le cas de la dépendance totale, puisque la deuxième base ne contient que des observations pour ce type de dépendance. Nous effectuons ce test pour les deux classes d'âge proposées précédemment.

La p-valeur du test est largement supérieure à 0.05 pour les deux tests indiquant que nous ne pouvons pas rejeter l'hypothèse d'égalité des fonctions de survie par classe d'âge à l'entrée des deux différentes bases. Nous pouvons donc conclure que l'étude conjointe de la fonction de survie des deux populations est possible. L'ajout de cette seconde base va nous permettre d'avoir plus d'observations de sorties pour les différents âges. Nous aurons ainsi moins d'interpolations à effectuer, et probablement moins de lissage. La variance de l'estimation va en outre diminuer par l'ajout de ces données.

5.3.3 Dépendance totale

Étude des fonctions de survie par classe d'âge

La médiane de l'espérance de vie est inférieure dans la première classe d'âge. Cette médiane confirme bien le fait que les individus tombant en dépendance jeune décèdent rapidement.

Grâce aux fonctions de survie précédentes, nous calculons des taux de décès annuels (cf. Illustration 5.29).

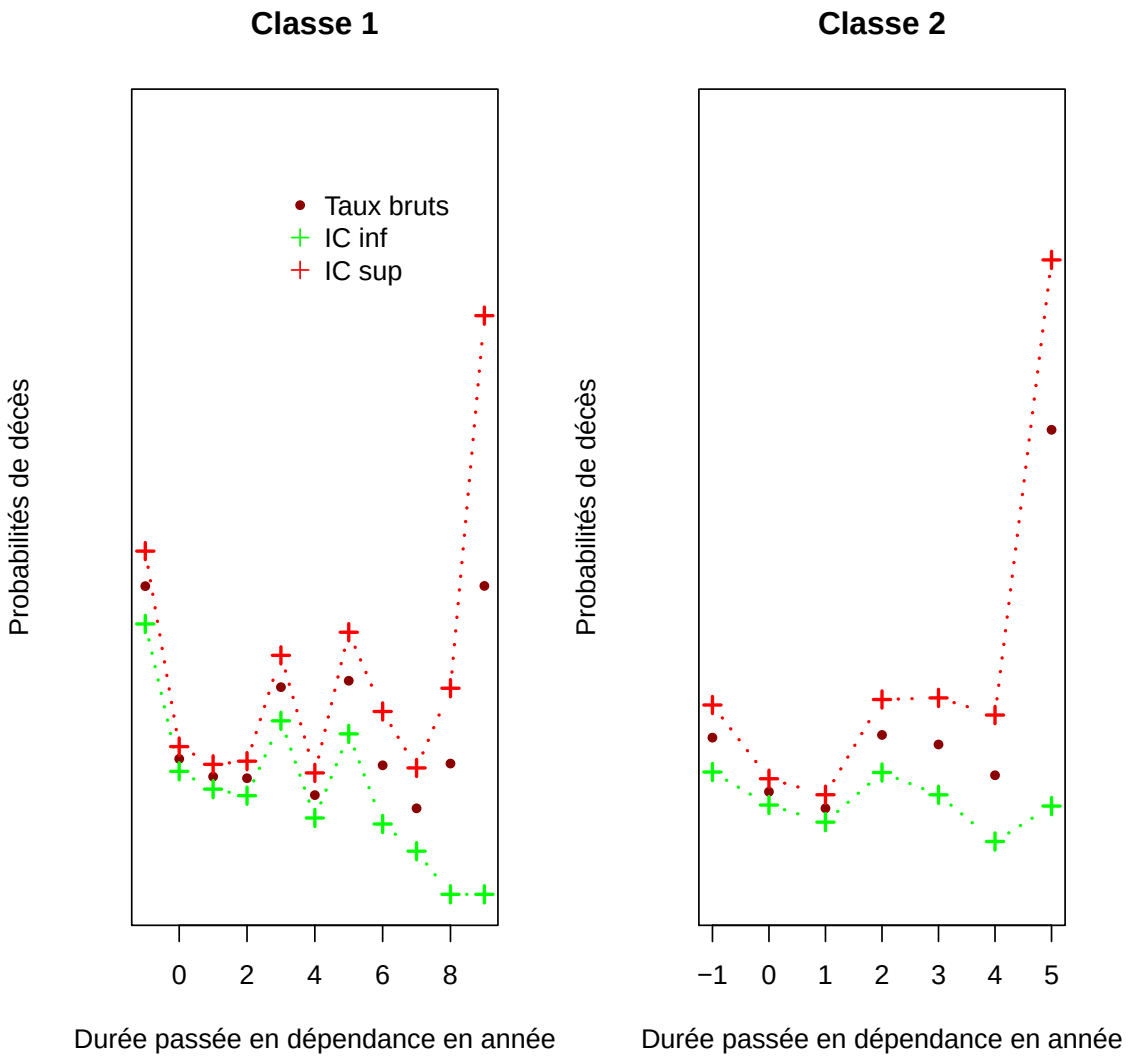

Illustration 5.29 : *Estimation des taux bruts*

Rappelons que, avec t représentant le temps passé en dépendance d'un individu et x un instant donné :

$$S(x) = \mathbb{P}(t > x)$$

$$\begin{aligned} q_x^i &= \mathbb{P}(t < x+1 | t > x) \\ &= \frac{\mathbb{P}(t < x+1 \cap t > x)}{\mathbb{P}(t > x)} \\ &= \frac{\mathbb{P}(x < t < x+1)}{\mathbb{P}(t > x)} \\ &= \frac{S(x) - S(x+1)}{S(x)} \end{aligned}$$

Nous nous apercevons qu'il est pertinent d'étudier ces fonctions de survie séparément. En effet, les individus entrant en dépendance avant 78 ans ont une probabilité de décéder beaucoup plus importante que les individus entrant en dépendance à

des âges plus élevés. Cette tendance s'explique par les pathologies touchant ces personnes, comme démontré dans la partie statistique descriptive. Par ailleurs, il semble que la mortalité des individus entrants en dépendance après 78 ans soit plus faible.

Interpolation par spline des taux bruts, puis lissage avec la méthode de Whittaker Henderson

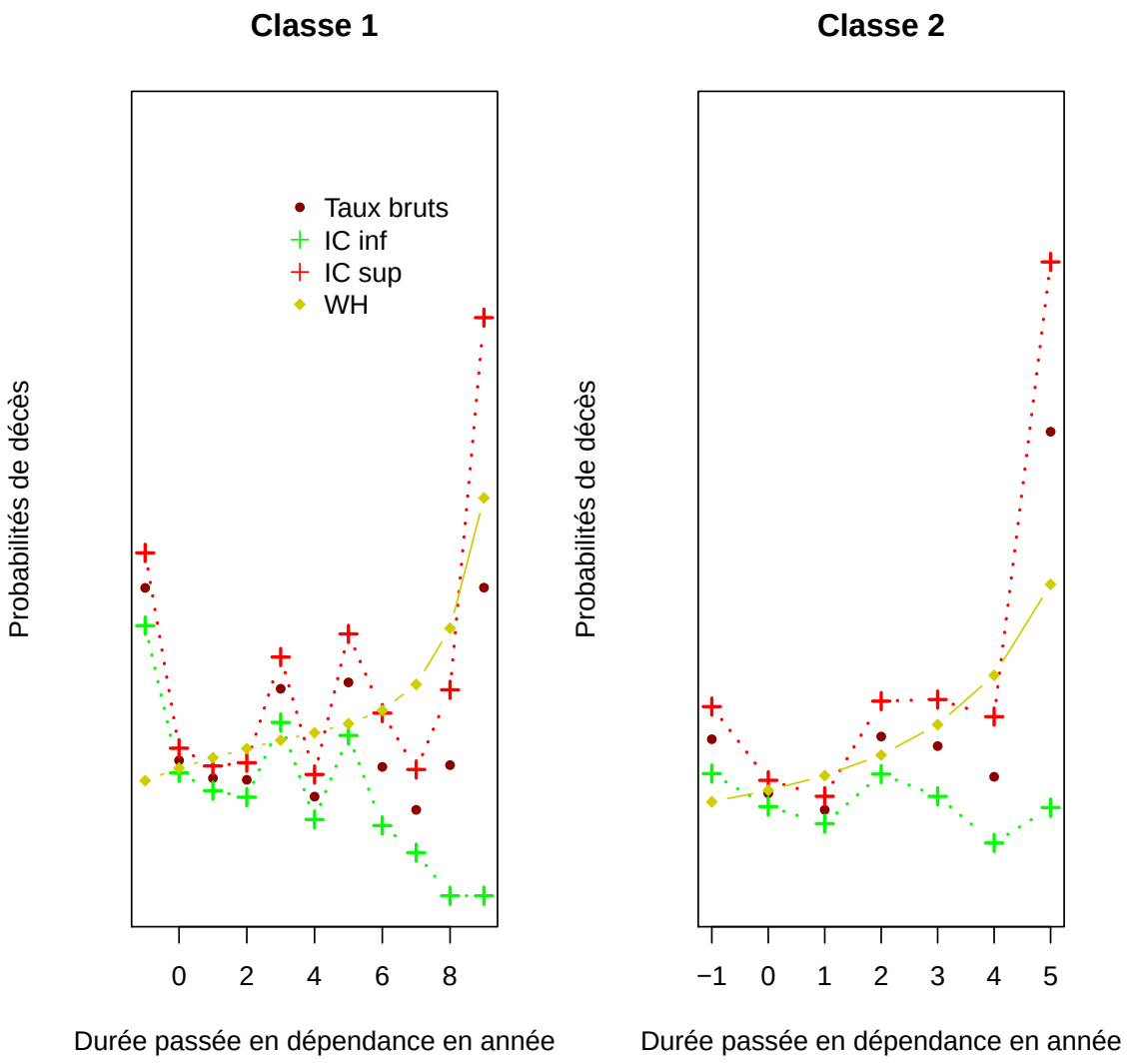

Illustration 5.30 : *Interpolation et lissage des taux bruts*

L'interpolation des taux bruts avec la méthode de Whittaker Henderson permet d'obtenir des courbes à l'aspect régulier (cf. Illustration 5.30). Un test du signe va nous permettre de voir s'il y a sur lissage ou non.

Obtenant une p-value de 0.55 pour la première classe et de presque 1 pour la seconde, nous pouvons conclure à un lissage raisonnable, puisque nous ne pouvons pas rejeter l'hypothèse selon laquelle la médiane entre le nombre de signes positifs et négatifs des résidus est nulle.

	Khi 2	MAPE	SMR
Classe 1	3.7	2.17	1.61
Classe 2	0.38	1.64	1.17

Test d'adéquation

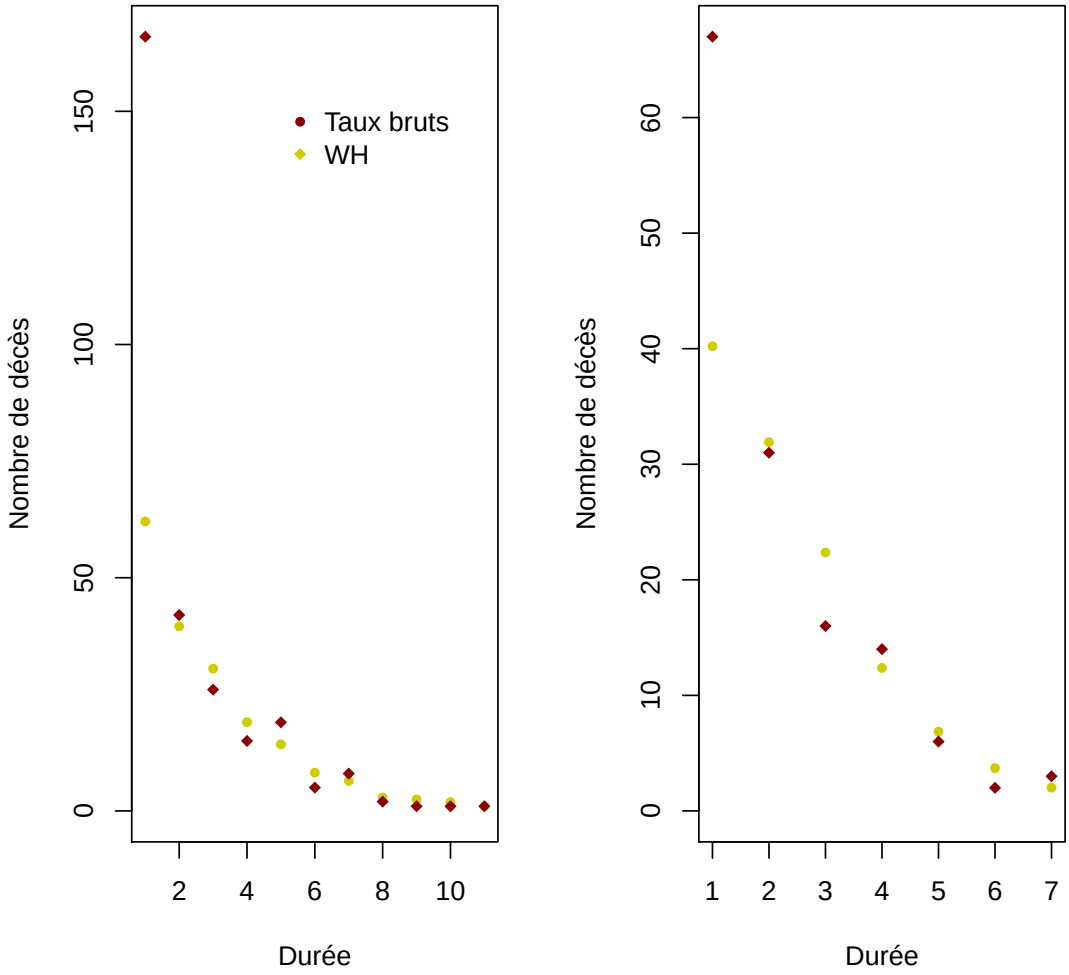

Illustration 5.31 : *Adéquation du nombre de sorties estimé par les taux lissés par rapport à celui estimé avec les taux bruts*

Les tests d'adéquations indiquent un ajustement par le lissage peu satisfaisant. Particulièrement pour le lissage des probabilités de décès pour la classe 1, le Khi 2 et le MAPE sont assez élevés.

Le SMR des deux ajustements, supérieur à 1, indique une sous évaluation globale des taux de décès. Cette sous-évaluation est due à l'estimation des taux de décès entre les instants 0 et 1. En effet, le nombre de sorties pour cette période est très important et mal reflété dans le lissage.

Analyse des résidus

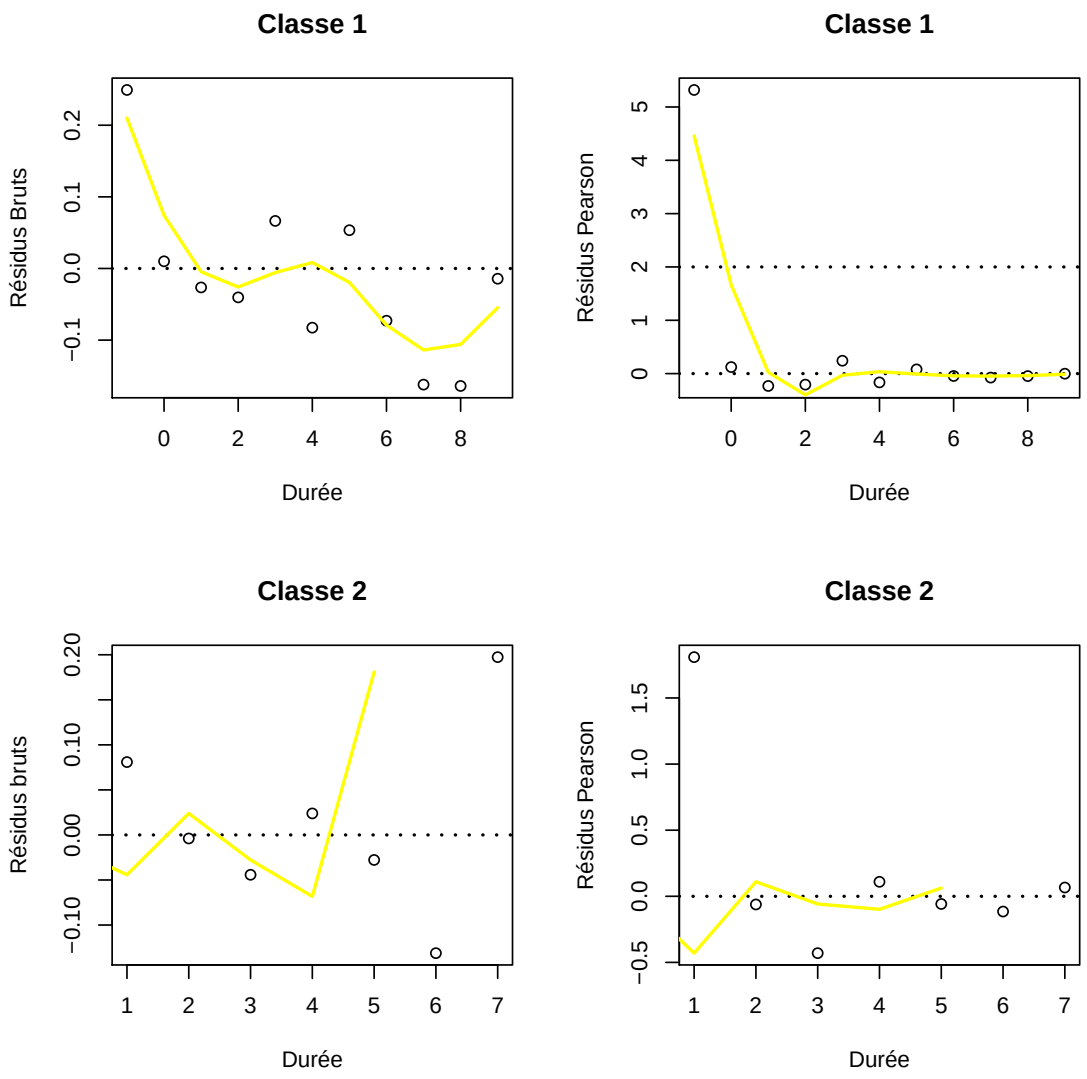

Illustration 5.32 : *Résidus entre l'estimation lissée et l'estimation brute pour les 2 classes d'âge*

Pour la classe 1, l'analyse des résidus bruts indique la mauvaise estimation de la première valeur, ainsi que des dernières valeurs, entre 7 et 10 ans. Les résidus de Pearson décèlent le mauvais ajustement de la première valeur, le résidu étant bien supérieur à 2.

Pour la classe 2, l'analyse des résidus est quasiment identique, toutefois l'ajustement est quelque peu meilleur : le résidu de Pearson pour la première valeur est plus proche de 0, mais reste supérieur à 2 en ce point.

Comparaison des espérances de survie

L'espérance de survie estimée par les taux bruts est inférieure pour les deux classes à l'espérance de survie obtenue par les taux lissés. Cette observation est

Espérance de vie résiduelle :	Taux bruts	Lissage WH
Classe 1	2.96	4.10
Classe 2	3.45	3.68

en accord avec les SMR obtenus précédemment : les taux lissés sous-estiment la mortalité, estimant ainsi une durée de vie plus longue qu'en réalité. Le lissage nous permet donc d'ajouter une certaine prudence à l'estimation de nos taux, puisqu'il est préférable de sous-estimer la mortalité des dépendants, afin de ne pas sous-estimer les engagements de l'assureur envers les individus dépendants.

Fermeture de la table de maintien en dépendance

Il demeure le problème de la fermeture des tables de survie pour les différentes classes d'âge de dépendants. Puisque l'hypothèse communément faite est qu'aucune rémission n'est possible, l'assuré décédera donc en dépendance.

Puisque l'observation de nos données ne nous permet pas d'émettre une conclusion à propos de l'évolution future de la durée de vie des dépendants, nous effectuons deux hypothèses que nous comparerons lors des projections du compte de résultat afin de visualiser leurs effets respectifs.

Premièrement, pour les durées pour lesquelles nous ne disposons plus d'observations, nous nous alignerons sur la table TH-TF. Cette table minore les probabilités de décès des dépendants, nous prenons ainsi une option prudente. Ainsi, pour les âges pour lesquels nous n'avons pas d'information quant à la mortalité des individus dépendants, nous faisons l'hypothèse que leur mortalité est égale à la mortalité de la table TH-TF. Cette approche nous permet de prendre en compte la mortalité particulièrement sévère à l'entrée en dépendance, la mortalité se normalisant au fur et à mesure de la survie de la personne dépendante, jusqu'à une période où nous pouvons émettre l'hypothèse que la personne est quasi saine, ou que du moins elle ne présente pas une surmortalité par rapport à la population nationale.

Deuxièmement, nous émettons l'hypothèse que nos estimations s'alignent également sur la TH-TF, mais cette fois ci cette dernière sera majorée par un coefficient de type SMR. Celui-ci est calculé en effectuant le rapport entre le nombre de décès observés et le nombre de décès prédits par la table TH-TF pour chaque âge et pour chaque période.

Il s'agit là de deux hypothèses extrêmes. Toutes deux doivent avoir un impact limité sur les résultats des simulations que nous présenterons dans la suite. En effet, la durée de vie d'un individu en dépendance est courte, en raison des probabilités de décès élevés lors des premières années. Ainsi, la plupart des individus décéderont avant d'avoir atteint l'âge de clôture de la table. Il faudra tout de même vérifier que ce petit nombre d'individus n'entraînera pas des résultats trop différents. Si tel n'était pas le cas, nous estimons qu'il ne nous sera pas possible de simuler sur un horizon de temps supérieur à la plus petite durée de clôture de nos différentes tables

de décès des dépendants.

Première méthode : alignement sur la TH-TF Nous obtenons ainsi pour la première méthode :

age/durée	0	1	2	3	4	5	6	7	8	...
45	$q_{45}^{i,estimé}$	$q_{46}^{i,estimé}$	$q_{47}^{i,estimé}$	$q_{48}^{i,estimé}$	$q_{49}^{i,estimé}$	$q_{50}^{a,TH-TF}$	$q_{51}^{a,TH-TF}$	$q_{52}^{a,TH-TF}$	$q_{53}^{a,TH-TF}$	...
46	$q_{46}^{i,estimé}$	$q_{47}^{i,estimé}$	$q_{48}^{i,estimé}$	$q_{49}^{i,estimé}$	$q_{50}^{i,estimé}$	$q_{51}^{a,TH-TF}$	$q_{52}^{a,TH-TF}$	$q_{53}^{a,TH-TF}$	$q_{54}^{a,TH-TF}$	...
...	...	...	...	...	...	...	...	...	...	...

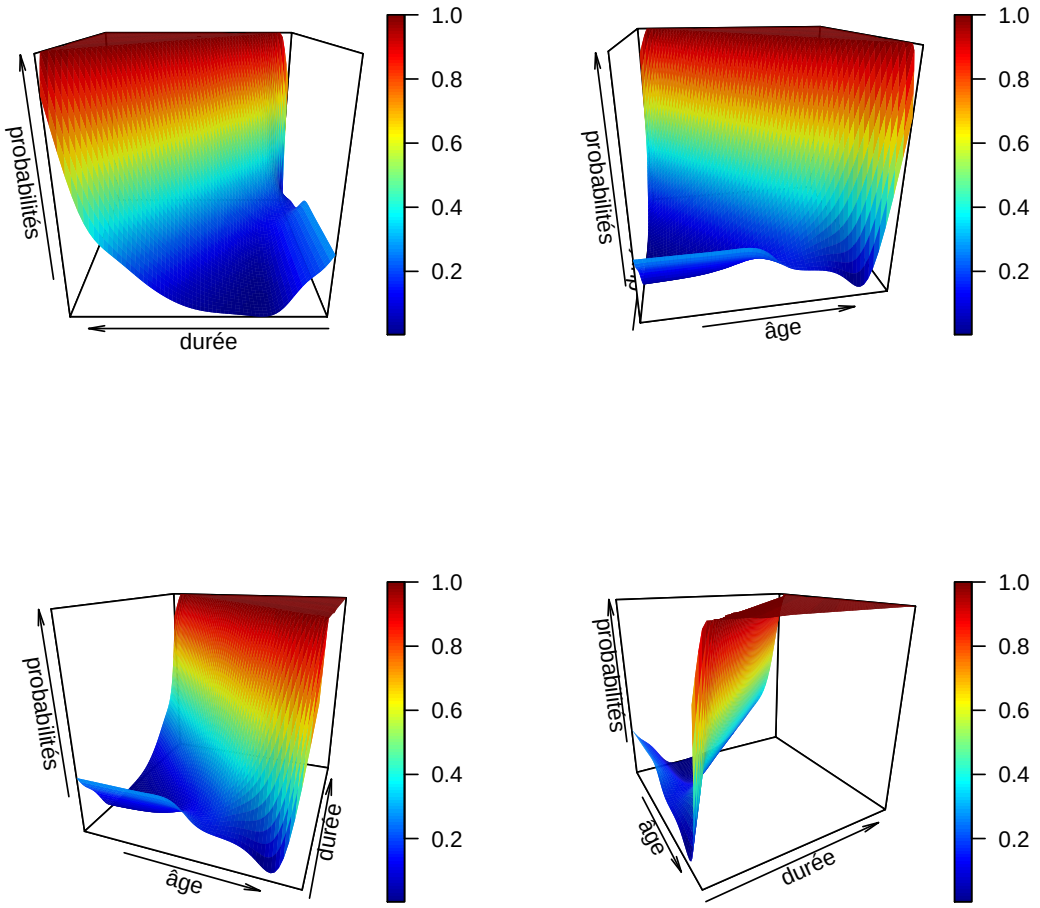

Illustration 5.33 : *Loi de décès des dépendants en fonction de l'âge à l'entrée obtenue avec la première méthode*

Nous constatons le creux dans cette surface. Celui ci provient de l'hypothèse prudente prise dans la prolongation de la table de mortalité des dépendants. La mortalité est dans un premier temps assez forte, puis diminue pour atteindre la mortalité d'une personne non dépendante. Enfin cette dernière augmente à nouveau, en raison du vieillissement de la personne.

Première méthode : alignement sur la TH-TF Et pour la deuxième méthode :

age/durée	0	1	2	3	4	5	6	7	8	...
45	$q_{45}^{i,estimé}$	$q_{46}^{i,estimé}$	$q_{47}^{i,estimé}$	$q_{48}^{i,estimé}$	$q_{49}^{i,estimé}$	$q_{50}^{a,TH-TF} \cdot SMR^{45}$	$q_{51}^{a,TH-TF} \cdot SMR^{45}$	$q_{52}^{a,TH-TF} \cdot SMR^{45}$	$q_{53}^{a,TH-TF} \cdot SMR^{45}$	...
46	$q_{46}^{i,estimé}$	$q_{47}^{i,estimé}$	$q_{48}^{i,estimé}$	$q_{49}^{i,estimé}$	$q_{50}^{i,estimé}$	$q_{51}^{a,TH-TF} \cdot SMR^{46}$	$q_{52}^{a,TH-TF} \cdot SMR^{46}$	$q_{53}^{a,TH-TF} \cdot SMR^{46}$	$q_{54}^{a,TH-TF} \cdot SMR^{46}$	...
...	...	...	...	...	...	...	...	...	...	...

où SMR^x est définit comme :

$$SMR^x = \frac{\sum_{k=x+1}^{\omega} Do_{k-x}}{\sum_{k=x+1}^{\omega} E_{k-x} \cdot q_k^{i,estimé}}$$

avec ω l'âge de clôture de la table, Do_k le nombre de sorties à la période k et E_k l'effectif sous risque à la période k.

NB : dans l'estimation du SMR, nous ne prenons pas en compte la forte mortalité survenant lors de la première période. Celle ci est liée à la survenance de la maladie et n'a aucune raison de se refléter continuellement dans la mortalité de l'individu.

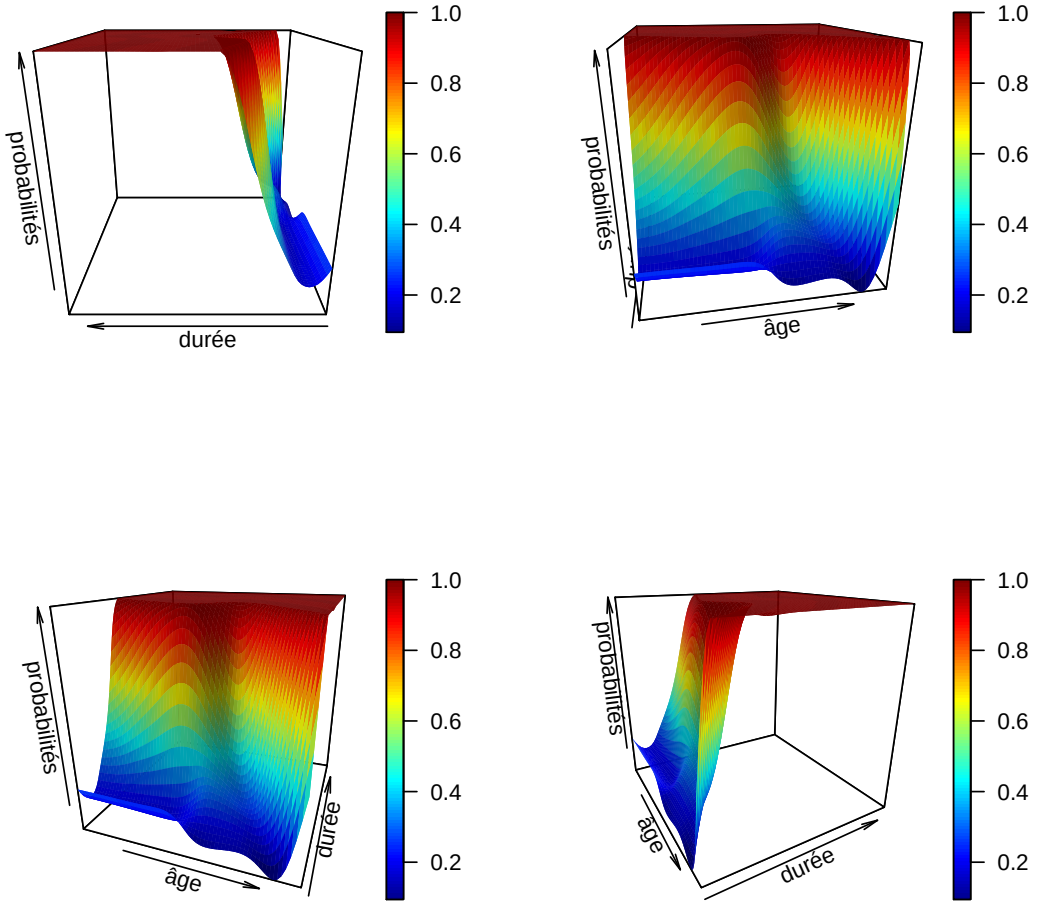

Illustration 5.34 : *Loi de décès des dépendants en fonction de l'âge à l'entrée obtenue avec la deuxième méthode*

A nouveau, le creux présent dans la surface est expliqué par la forte mortalité survenant après la dépendance. Néanmoins, ce dernier est beaucoup moins profond que précédemment. Ce phénomène s'explique par le fait que la mortalité passe par une phase de stabilisation, puis contrairement à la première méthode, ré-augmente fortement, puisque la mortalité est celle de la table TH-TF, mais majorée afin de prendre en compte la surmortalité d'une personne dépendante.

5.3.4 Dépendance partielle

Pour la dépendance partielle, nous appliquons la même méthodologie que pour la dépendance totale, à savoir que nous déterminons la fonction de survie et ses intervalles de confiance grâce à la méthode de Kaplan Meier, nous interpolons et lissons lorsque nécessaire grâce aux splines et au lissage de Whittaker Henderson. Afin de clôturer la table, nous procédons également par alignement sur la TH-TF.

Toutefois, nous ne distinguons pas de classes : la durée de survie en dépendance partielle est supposée être la même pour tous les individus, peu importe leur âge à l'entrée.

Grâce à l'estimateur de Kaplan Meier, nous obtenons la fonction de survie suivante :

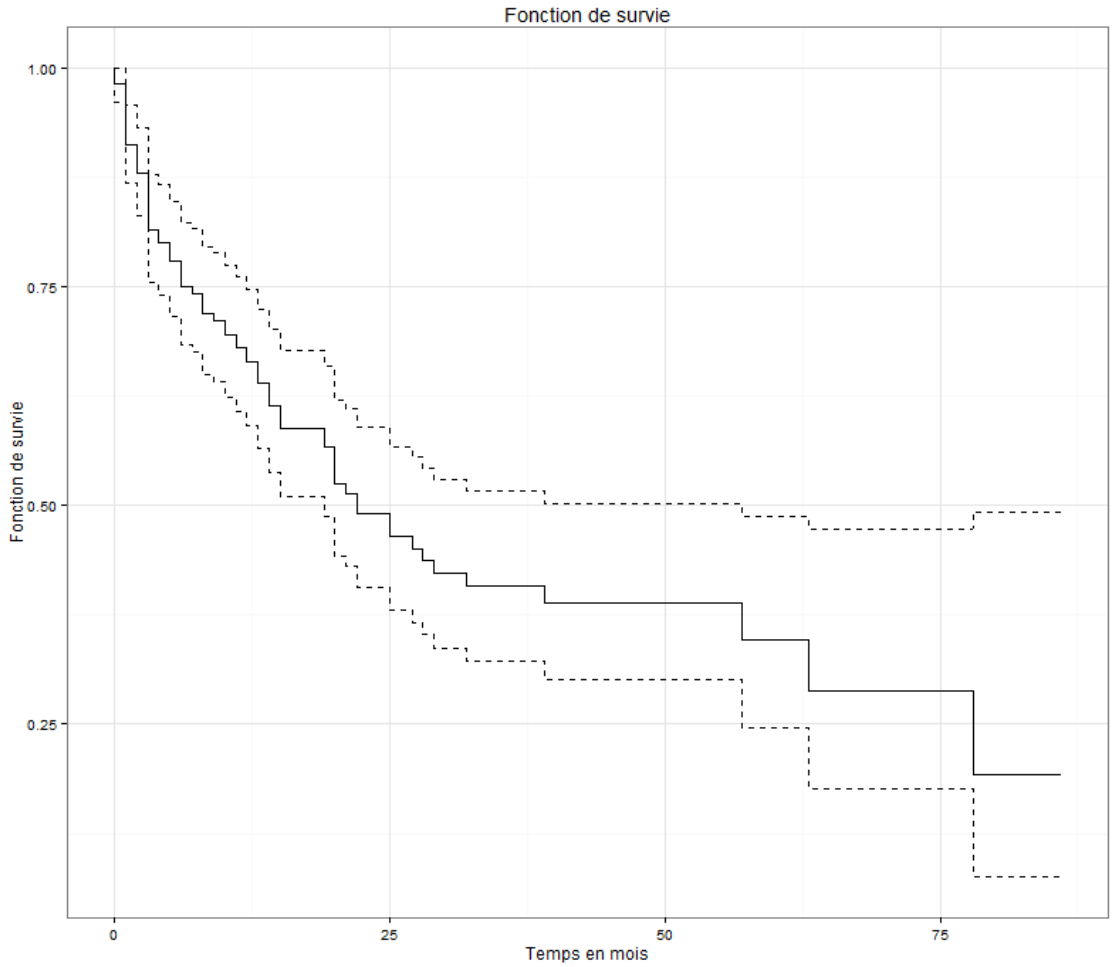

Illustration 5.35 : *Estimation de la fonction de survie*

Nous remarquons que cette fonction de survie décroît très rapidement pour les premiers mois, puis marque des paliers. Ce phénomène s'explique par le manque de données pour les durées élevées en dépendance partielle. D'une part, une grande majorité des individus entrant en dépendance est immédiatement en dépendance totale, et d'autre part un certain nombre d'individus en dépendance partielle passe en dépendance totale au bout d'un certain temps. Nous les avons ainsi traités comme des censures.

Néanmoins, l'aspect convexe de cette fonction de survie nous indique la forte mortalité des individus en dépendance partielle.

L'espérance de vie médiane des individus en dépendance partielle indique que la distinction dépendance partielle et dépendance totale est justifiée, puisque sensiblement différente des espérances de vie médiane calculées pour les deux classes de

dépendance totale.

Classe d'âge :	[53,90]
Espérance de vie médiane :	22
Intervalle de confiance inférieur à 95% :	19
Intervalle de confiance supérieur à 95% :	57

Toutefois, l'intervalle de confiance pour l'estimation de cette médiane est très large, du fait du manque de données pour ce type de dépendance. Cependant, puisqu'il apparaît pertinent d'étudier séparément la survie pour ce type de dépendance, nous acceptons le risque d'erreur quant à cette modélisation, et le prendrons en compte via la projection que nous effectuerons par la suite en utilisant l'intervalle de confiance supérieur et inférieur dans la simulation.

Grâce à la fonction de survie de Kaplan Meier, nous obtenons les taux de mortalité suivant :

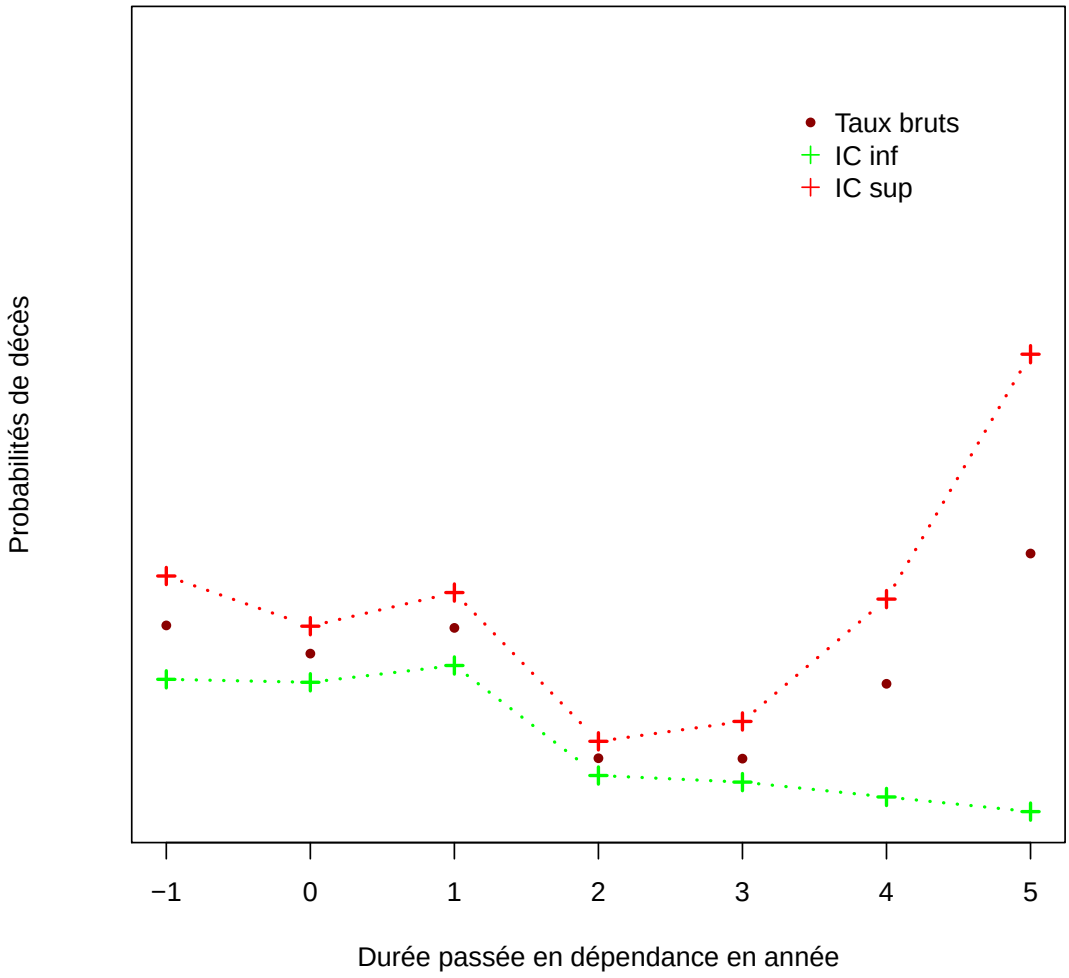

Illustration 5.36 : *Estimation des taux bruts*

	Khi 2	MAPE	SMR
Classe 1	0.43	4.74	1.5

Ces taux de mortalité témoignent bien de la sévérité de la mortalité pour les individus lors des premières périodes en dépendance partielle.

Nous lissons ces taux par la méthode de Whittaker Henderson, afin de visualiser la tendance sous-jacente de ces taux.

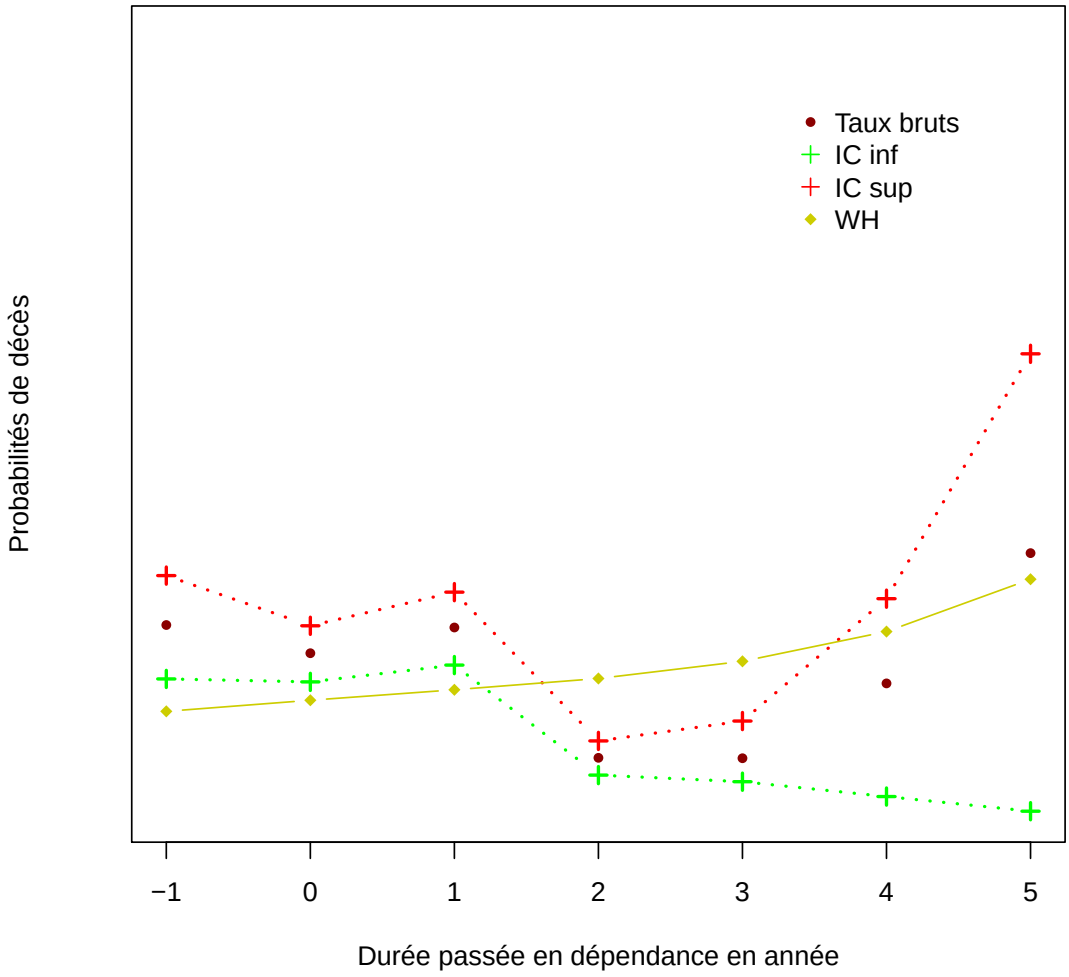

Illustration 5.37 : *Interpolation et lissage des taux bruts*

Analyse de l'ajustement L'ajustement semble peu satisfaisant au vu des tests d'adéquation : le Khi 2 s'éloigne de 0, le MAPE est élevé et le SMR s'éloigne de 1. Néanmoins, le SMR supérieur à 1 indique une sous-estimation des taux de mortalité. Si c'est cet effet qui est la cause des mauvais résultats aux tests d'adéquation, cela n'est pas gênant pour notre estimation. En effet, il est plus prudent de sous-estimer la mortalité des dépendants afin de réaliser des projections.

A présent, nous observons les résidus bruts et de Pearson de cet ajustement afin

de déceler le mauvais ajustement local.

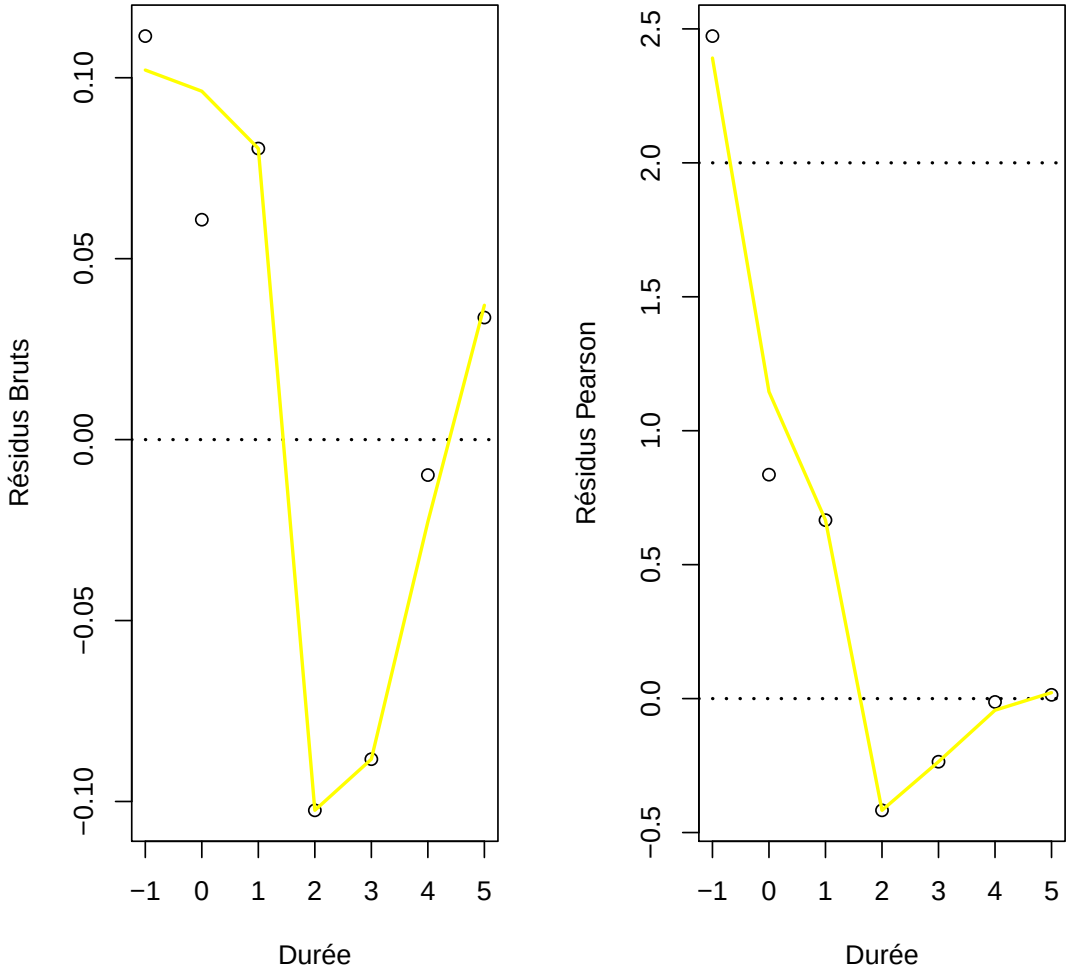

Illustration 5.38 : *Résidus entre l'estimation lissée et l'estimation brute*

L'analyse des résidus bruts révèle un mauvais ajustement : l'ajustement sous-estime les taux de mortalité des dépendants pendant une longue période, puis les surestime. Les résidus de Pearson décèlent comme dans le cas de la dépendance totale la première valeur de l'ajustement particulièrement éloignée des taux bruts. Nous effectuons un test du signe afin de mettre en lumière un éventuel sur lissage, et nous obtenons une p-value égale proche de 1, ce qui indique que nous ne pouvons pas rejeter l'hypothèse selon laquelle la médiane entre le nombre de signes positifs et négatifs des résidus est nulle.

Cohérence de l'estimation L'espérance de survie résiduelle obtenue par les taux bruts est de 2.96 ans. En lissant ces taux, l'espérance de vie est de 4.10 ans. Nous constatons ainsi la sous-estimation des taux de mortalité effectuée par le lissage.

5.3.5 Application des arbres de régression

Pour l'estimation de la loi de mortalité des dépendants, nous avions segmenté les dépendants en catégories d'âge, en choisissant des classes d'âge avec des effectifs égaux pour chacune des classes. Mais rien n'indique que le choix de ces différentes classes d'âge est justifié. Cette catégorisation aurait pu être effectuée par la technique des arbres de régression. Cette technique étudie les différentes classes d'âge possibles par un algorithme permettant de créer des groupes d'homogénéité maximale. L'algorithme se base sur un test du log rank afin de déterminer quels facteurs (sexe, âge à l'entrée par exemple) permettent de différencier la fonction de survie, et à l'intérieur de ces facteurs, quelles variables (50 ans et plus par exemple) sont importantes dans l'explication de la survie. Il faut alors spécifier un seuil de significativité : ce seuil de significativité correspond à la valeur de la p-value maximale que chaque test du log rank doit avoir afin que le facteur ou la variable soit considéré comme particulièrement explicative de la survie. Nous nous servons du package "party" du logiciel R afin de mettre en oeuvre cette technique.

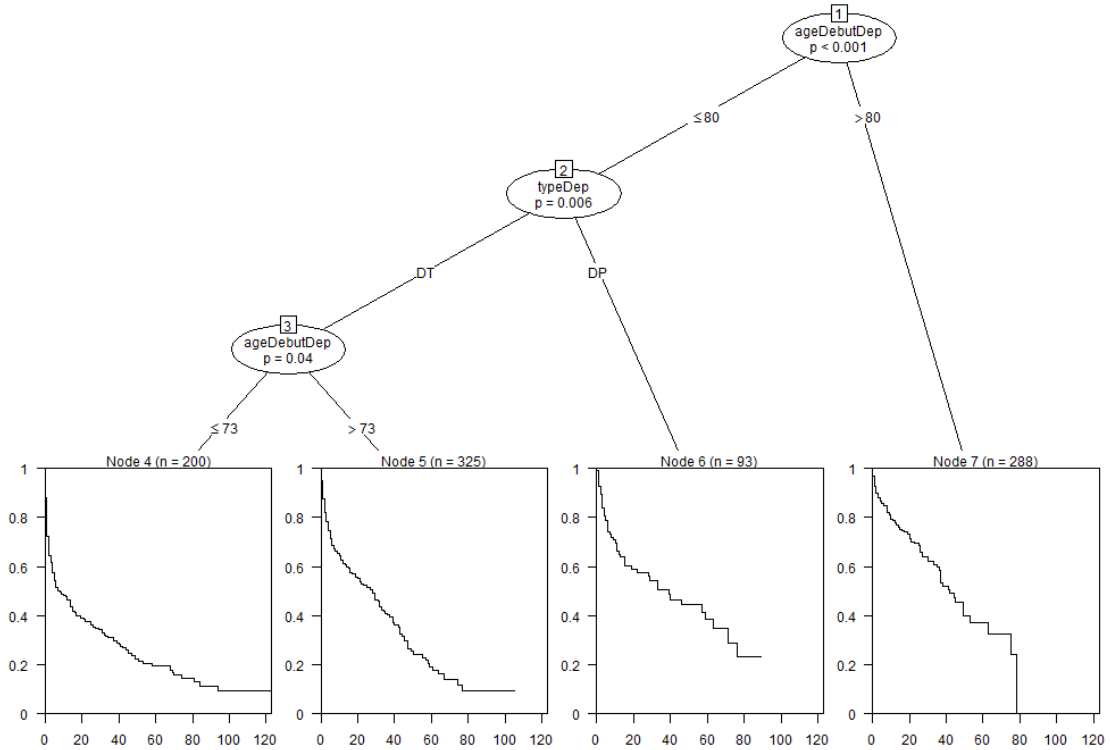

Illustration 5.39 : *Arbre de régression*

Le seuil de significativité est de 0.05. Le résultat de l'arbre de régression confirme nos premières observations : l'âge à l'entrée est un facteur important dans la détermination de la durée de survie des dépendants. En particulier, il semblerait peu significatif du point de vue de la survie que l'assuré soit en dépendance partielle ou totale lorsque ce dernier entre en dépendance après 80 ans. Toutefois avant 80 ans, l'arbre indique une différence significative entre la dépendance partielle et totale.

Ainsi, il semblerait pertinent de distinguer la survie des individus en dépendance totale en fonction de 3 classes d'âge : (-,73],(73,80),[80,-)

Néanmoins, nous ne nous servirons pas de ces résultats par la suite. En effet, une étude plus approfondie de l'algorithme utilisé serait nécessaire.

Section IV

Applications : Simulation du compte de résultat

L'objectif de ce mémoire est de mesurer la rentabilité d'un produit d'assurance dépendance. Par le terme rentabilité, nous entendons déterminer si ce produit va générer des pertes car le devenir de ce type de produit est souvent incertain. Ayant été tarifié dans les années 1990, la question se pose s'il y a eu une sur-estimation ou une sous-estimation du nombre d'entrées en dépendance, de la mortalité des autonomes et de celle des dépendants.

Afin de répondre à ces questions, et afin de mesurer la rentabilité future de ce produit, nous réaliserons des projections du compte de résultat.

Ces projections vont être réalisées grâce à des simulations de Monte Carlo et se veulent être les plus proches de la réalité en raison de l'estimation des lois propres aux observations. En effet, dans les parties qui ont précédé, nous nous sommes attachés à déterminer les lois de mortalité des autonomes, des dépendants, du passage en dépendance, partielle et totale, ce qui va par la suite nous permettre de projeter le nombre futur d'autonomes ainsi que le nombre futur de dépendants. Une estimation du montant des cotisations et des prestations futures va ainsi pouvoir être déterminée.

Nous prendrons également en compte l'impact des intérêts financiers générés par le placement des provisions pour risques croissants et pour rente.

Ainsi, nous présentons maintenant la construction du compte de résultat, avec les différents éléments qui seront simulés, puis nous analyserons le fonctionnement des simulations et enfin nous calculerons les résultats de ces simulations en fonction des différentes lois estimées auparavant.

Partie 6

Aspects comptables

6.1 Préliminaires : fonctionnement de la réassurance

Avant de décrire le compte de résultat à projeter, nous allons présenter le fonctionnement de la réassurance pour le produit considéré afin de comprendre les éléments composant le compte de résultat.

6.1.1 Principes généraux et histoire

De manière conceptuelle, la réassurance est l'assurance des assurances. "C'est le fait, pour un assureur direct, de répercuter une partie des périls ou des risques qu'il assume pour le preneur d'assurance au titre de contrats d'assurance ou de dispositions juridiques, sur un deuxième assureur qui n'est pas directement en contact avec le preneur d'assurance, c'est-à-dire le réassureur." [1] La réassurance fit son apparition au 14 ème siècle. À cette époque, le transport de marchandises s'effectuait principalement par voie maritime. Ce mode de transport présentait des risques importants liés à un faible niveau technologique et à la présence de pirates. C'est ainsi qu'en 1370 un assureur Vénitien assurant le transport de marchandises entre l'Italie et les Pays-Bas décida de céder le risque à un autre assureur sur la partie la plus risquée du voyage, autour du détroit de Gibraltar et lors de l'escale à Cadix en Espagne.

La réassurance telle que nous la connaissons aujourd'hui apparaît au XIXème siècle avec la révolution industrielle allemande. Notamment après l'incendie de Hambourg en 1842 qui dura 3 jours et entraîna la destruction de 4000 logements. Les assureurs décidèrent alors de créer la Koelnische Rueck afin de les aider à prendre en charge des sinistres ayant des conséquences financières de plus en plus graves.

1. Grossmann Marcel, 1983, Rückversicherung, Paris : l'Argus.

6.1.2 Fonctionnement d'une compagnie de réassurance

La société d'assurance, qui est ici l'assuré, est désignée par le terme de cédante. Elle soumet une partie de son risque à une autre société, le réassureur, en échange d'une prime. Ce transfert de risque est appelé la cession.

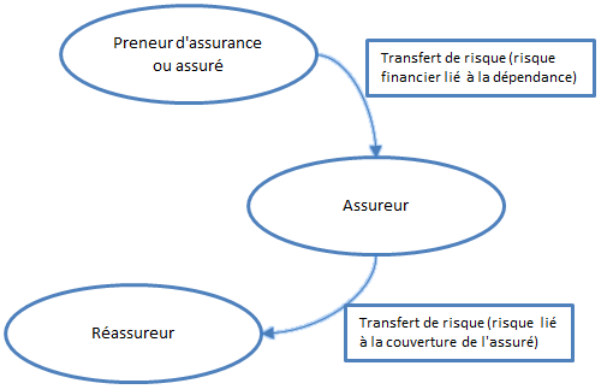

Illustration 6.1 : *Illustration de la cession du risque d'une compagnie d'assurance à une compagnie de réassurance*

Rétrocession Le réassureur peut lui-même céder une partie de son risque à un autre réassureur, c'est ce qui est appelé la rétrocession.

Pool de réassurance Un pool de réassurance est le partage entre plusieurs réassureurs du risque d'un même assureur. Le réassureur ayant la plus grande part de cession est l'apériteur ou le leader, les autres sont des followers. L'apériteur décide des conditions du traité de réassurance et peut les modifier.

6.1.3 Utilités de la réassurance

Transfert de risques Le transfert de risques permet à l'assureur de bénéficier d'une plus grande capacité financière, ce qui lui permet de souscrire plus de contrats. La réassurance permet aussi à la cédante de lisser les résultats financiers d'une année à l'autre. En effet, sans la réassurance, l'assureur pourrait être gravement affecté par un pic de sinistralité au cours d'une année (exemple du World Trade Center en 2001). Nous verrons dans le chapitre suivant comment est effectué le transfert de risques selon le contrat de réassurance.

Amélioration de la solvabilité Elle permet aux sociétés d'augmenter leurs marges de solvabilité. En effet, la législateur autorise à ne retenir dans le calcul des exigences en matière de solvabilité que les risques conservés par l'assureur. Ces exigences de solvabilité correspondent à un capital supplémentaire à constituer par l'assureur afin d'assurer la continuité de son activité, même en cas de choc de sinistralité important.

Financement du coût d'acquisition L'assureur a besoin d'agents commerciaux pour la distribution de ses produits. Ces derniers sont rémunérés sous forme de commissions importantes lors des premières années du contrat. Ce coût représente une charge importante pour l'assureur. Le réassureur vient en aide à l'assureur en participant au financement des coûts d'acquisition.

Expérience du réassureur De par son expérience à l'international ainsi que son accès à de nombreuses bases de données, le réassureur est en mesure d'épauler l'assureur dans sa tarification ou création de nouveaux produits.

Croissance économique De par son soutien aux activités industrielles et agricoles des pays émergents, le réassureur participe à la croissance économique. En effet, l'offre d'assurance dans ces pays étant faible, le réassureur vient en appui des courtiers et assureurs présents dans ces pays, car ces derniers disposent de peu de capitaux.

Cette présentation de la réassurance nous a permis de comprendre pourquoi la réassurance a été créée, comment fonctionne une compagnie de réassurance et quels sont les avantages de cette activité pour un assureur. Nous allons maintenant essayer de comprendre plus en détail comment fonctionne un traité de réassurance, c'est à dire les différentes modalités que peut prendre un contrat entre un assureur et un réassureur.

6.2 Caractéristiques d'un traité de réassurance

De manière générale la réassurance vie est distinguée de la réassurance non vie. De même ces deux branches se distinguent en deux modes de réassurances, (réassurance facultatives et réassurances obligatoires) et en plusieurs types de contrats (proportionnel et non proportionnel) que nous détaillons dans la présente partie.

6.2.1 La réassurance proportionnelle

La réassurance est proportionnelle lorsque la prise en charge du risque est partagée entre l'assureur et le réassureur en fonction d'une quote part identique sur le montant des primes et sur le montant des sinistres. Ce type de réassurance limite l'aléa moral, puisque le sort de l'assureur est directement lié à celui du réassureur.

Quote-part Un traité en quote-part est un traité dans lequel l'assureur cède au réassureur une proportion fixe de chaque sinistre et une même proportion de primes. Il en garde la proportion complémentaire.

Ce type de réassurance présente l'avantage d'être simple à gérer. Il fait bénéficier la cédante de commissions de réassurance. En effet la quote part est appliquée sur le montant de prime global, comprenant la couverture du risque mais aussi la couverture des frais de fonctionnement de l'assureur. Or le réassureur n'ayant pas à s'occuper de la distribution des contrats notamment, a beaucoup moins de frais que l'assureur. Ainsi, il reverse une partie de la portion de prime afférant à la couverture des frais.

L'inconvénient (pour la cédante) est qu'un certain pourcentage d'affaires qui pourrait être profitable est cédé.

Excédent de pleins L'assureur direct cède l'ensemble des polices présentant une somme assurée supérieure à un certain niveau d'engagement appelé plein de rétention. Il limite son engagement en fixant une capacité maximale (souvent exprimée en nombre de rétention).
Le taux de cession t par police est égal à, avec R la rétention, C la capacité et K le capital assuré :

Si K-R >= C :

$$t = max(\frac{K-R}{K}; 0)$$

Sinon :

$$t = \frac{C}{K}$$

L'avantage de ce type de réassurance est le fort pouvoir de lissage du portefeuille de l'assureur. Il fait aussi bénéficier la cédante de commission de réassurance.
Son inconvénient est sa complexité qui entraîne un coût plus important.

6.3 La réassurance non proportionnelle

La réassurance est non proportionnelle lorsque l'assureur détermine la limite de sinistre ou de perte qu'il est prêt à assumer, la partie excédentaire étant prise en charge par le réassureur. Elle dépend d'une priorité (qui correspond à une franchise, montant en dessous duquel le réassureur n'intervient pas) et d'une portée (montant maximal que le réassureur s'engage à rembourser).
Dans ce mode de réassurance, la cédante et le réassureur sont beaucoup moins liés.
Ce mode de réassurance permet d'écrêter le risque à charge de l'assureur en cas de sinistre de forte intensité. Il sert par exemple à couvrir les risques de pandémie, de catastrophes naturelles, terrorisme, nucléaire, biologique ou chimique.

Excédent de sinistre Dans les traités en excédent de sinistres, le réassureur s'engage à prendre en charge le montant des sinistres dépassant une certaine somme :

la priorité. La cédante conserve le montant du sinistre inférieur à la priorité et cède la partie supérieure au réassureur, dans les limites de la capacité fournie par le réassureur définie de manière contractuelle.
Toutefois, lorsque la capacité a été atteinte, une reconstitution est possible, moyennant une surprime (prime de reconstitution).

Excédent de perte Appelé Stop loss en anglais, ce type de réassurance permet à l'assureur de se protéger contre les fluctuations de la sinistralité. En effet, le réassureur s'engage à verser la part de la charge de sinistre totale annuelle dépassant un montant (priorité). Le réassureur n'intervient que si l'assureur direct est en perte, c'est-à-dire si les primes perçues sont inférieures à la somme sinistre et charge.
Ce mode de réassurance permet une couverture complète de l'assureur mais demeure toutefois limité car risqué pour le réassureur. C'est ce type de traité qui réassure souvent les risques de tempêtes et de grêle.

6.3.1 Traités obligatoires et traités facultatifs

Facultatifs La réassurance facultative est basée sur l'analyse individuelle des polices soumises par l'assureur direct, notamment via un processus de sélection médicale. Cela permet au réassureur de réduire le risque de perte pour certaines polices jugées plus à risque et apporte une expertise supplémentaire quant à la sélection des polices effectuée par l'assureur.
Cette sélection est effectuée par des services spéciaux de sélection médicale présents chez le réassureur.

La réassurance proportionnelle en Quote Part est le seul type de couverture utilisée par les assureurs dans la réassurance de la dépendance. Tarifier une couverture non proportionnelle pour la dépendance serait difficile pour le réassureur par manque d'expérience à long terme.

De plus, la dépendance étant un risque dont les conséquences financières les plus importantes risquent de survenir dans un horizon de temps lointain, une couverture non proportionnelle de type excédent de sinistre serait inadaptée car ce type de couverture est généralement annuelle.

Par ailleurs, la quote part permet comme nous l'avons vu de limiter l'aléa moral. Puisqu'il y a partage du risque, l'assureur a tout intérêt à ne pas réaliser de pertes et de procéder à d'éventuelles augmentations de tarifs si nécessaire. Cela ne serait pas forcément le cas avec une couverture de type non proportionnelle.

6.4 Description du compte de résultat du réassureur

Le compte de résultat est "... le document comptable qui regroupe les différents produits et charges pour une période déterminée : il fournit l'explication et permet donc le contrôle de l'enrichissement ou de l'appauvrissement de l'entreprise." [2]

Il permet ainsi d'obtenir le résultat net d'un exercice, c'est à dire la variation de patrimoine positive ou négative pour l'entreprise résultant de l'exercice passé.

Le compte de résultat que nous considérons ici est celui concernant le réassurance, et se compose des éléments suivants :

Compte annuel de résultat :
+ Cotisation cédées - Commissions de réassurance - Frais de gestion du réassureur
+ PRC à l'ouverture - PRC à la clôture
- Prestations réglées cédées + PMR à l'ouverture - PMR à la clôture
+ Produits financiers techniques
= Résultat technique de l'exercice N

Cotisations cédées Les cotisations cédées sont égales aux cotisations originales que l'assureur cède au réassureur (dans le cas présent, un certain pourcentage du montant total de prime).

Commissions de réassurance Les commissions de réassurance ont pour but, comme nous l'avons vu précédemment, de dédommager la cédante des frais d'acquisition, de gestion des contrats qu'elle a, alors que le réassureur ne les a pas et lui sont tout de même versés puisque intégrés aux cotisations cédées.

Frais de gestion du réassureur Les frais de gestion du réassureur représentent les coûts de fonctionnement du réassureur.

Provisions pour Risques Croissants et Provisions Mathématiques de Rente Les PRC sont les provisions pour risques croissants, auxquelles le réassureur participe en fonction de sa quote part. Il en est de même pour les provisions mathématiques de rente.

2. GILLET, J.-P., GINSBURGH, V., MICHEL, P. et KHROUZ, F., Comptabilité, Office International de Librairie, Bruxelles, 1985

Prestations Les prestations sont les prestations payées aux individus en dépendance, c'est à dire les montants de rente annuelle.

Produits financiers Les produits financiers techniques proviennent du placement des PRC et des PMR détenus par le réassureur dans le cas du nantissement de titre. Sinon ils proviennent de la rémunération des espèces si les provisions sont détenues par la cédante.

En effet, les provisions techniques des contrats réassurés font généralement l'objet d'un dépôt du réassureur qui peut être réalisé soit sous forme d'un nantissement de titres, soit sous la forme d'espèces.

Dans le cas d'un nantissement, l'assureur verse au réassureur une somme correspondante aux provisions à représenter qui sera investie par le réassureur sur des titres nantis au profit de l'assureur. Une partie de la rémunération revient au réassureur.

Dans le cas d'un dépôt espèces, c'est l'assureur qui conserve les provisions et les place. Le réassureur exige ainsi une rémunération annuelle sur ces actifs.

6.5 Description des éléments simulés

Les éléments simulés vont être :

- les cotisations, qui seront payées tant qu'un individu ne sera ni décédé, ni dépendant, ni réduit (toutefois nous ne simulons pas de nouvelles réductions, les réductions simulées concernent les contrats déja réduits au moment de la simulation),
- les PRC, sous les mêmes conditions que les cotisations,
- les prestations qui seront payées aux personnes dépendantes et à celles qui le deviendront,
- les provisions de rente, sous les mêmes conditions que les prestations.

Par ailleurs, nous utiliserons des hypothèses quant aux taux techniques et taux de rendement des produits financiers.

Ainsi, l'objectif va être de simuler le résultat futur sur un horizon suffisamment long afin de visualiser d'éventuelles pertes futures.

Compte annuel de résultat :	Compte de résultat simulé N+1	Compte de résultat simulé N+2	...
+ Cotisations cédées	?	?	...
- Indemnités de gestion de la cédante	?	?	...
- Frais de gestion du réassureur	?	?	...
+ PRC à l'ouverture	?	?	...
- PRC à la clôture	?	?	...
- Prestations réglées cédées	?	?	...
+ PMR à l'ouverture	?	?	...
- PMR à la clôture	?	?	...
+ Produits financiers techniques	?	?	...
= Résultat de l'exercice N	=Résultat de l'exercice N+1	=Résultat de l'exercice N+2	...

Nous avons vu dans cette partie quels étaient les éléments à simuler, nous présenterons par la suite une façon de les simuler, consistant en la simulation de la vie de chaque assuré du portefeuille.

Partie 7

Utilisation des simulations

7.1 Théorie des simulations de Monte Carlo

La méthode de Monte Carlo, du nom de la ville célèbre pour ses casinos, est apparu sous une forme rudimentaire en 1777, décrite alors pour le calcul du nombre Π, avec la réalisation d'expériences répétées. Les méthodes utilisées actuellement ont été inventées en 1947 par N. Metropolis [1]. Il s'agit de méthodes d'approximation. Dans notre cas, nous chercherons à approximer la distribution de probabilité des éléments à simuler.

Soit X une variable aléatoire. Nous cherchons à calculer $\mathbb{E}(X)$ par le biais de simulation de variable aléatoire suivant la loi de X. Soit $(X_i)_{1\leq i\leq N}$ les N réalisations de la variable aléatoire X.
Nous obtenons alors :

$$\mathbb{E}(X) \simeq \frac{1}{N}(X_1 + ... + X_N)$$

La convergence vers $\mathbb{E}(X)$ est assurée par la loi forte des grands nombres (à condition que X soit intégrable)

L'erreur commise θ_N est égale à :

$$\epsilon_N = \mathbb{E}(X) - \frac{1}{N}(X_1 + ... + X_N)$$

Le théorème centrale limite permet d'affirmer que l'erreur converge en loi vers une gaussienne centrée :

$$\frac{\sqrt{N}}{\sigma}\epsilon_N \simeq N(0,1)$$

où σ est l'écart type de X.

1. Nicholas Metropolis et Stanislaw Ulam, « The Monte Carlo Method », Journal of the American Statistical Association, vol. 44, no 247, septembre 1949, p. 335-341

Un intervalle de confiance à 95% peut donc être obtenu pour l'estimation :

$$[\frac{1}{N}\sum_{i=1}^{N} X_i - 1.96\frac{\sigma}{\sqrt{N}}, \frac{1}{N}\sum_{i=1}^{N} X_i + 1.96\frac{\sigma}{\sqrt{N}}]$$

où la variance empirique de X peut être obtenue par :

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i^2 - (\frac{1}{N}\sum_{i=1}^{N} X_i)^2)$$

7.2 Application : simulation du compte de résultat

7.2.1 Intérêt de la simulation de Monte Carlo

La simulation de Monte Carlo appliquée à ce problème présente des avantages en comparaison d'une méthode déterministe.

Elle permet de mesurer l'impact de certains paramètres sur la sinistralité du portefeuille ainsi que le calcul d'intervalle de confiance pour les grandeurs estimées par l'obtention de la distribution des éléments simulés.

De plus, cette méthode prend en compte le caractère évolutif de l'état de dépendance des assurés (passage de dépendance partielle vers dépendance totale).

Nous pouvons ainsi obtenir, certes en fonction d'hypothèses, l'ensemble des réalisations possibles, mais aussi la probabilité de réalisation de chacun de ces évènements. Nous pouvons ainsi calculer des valeurs comme la Value at Risk, nous indiquant le montant étant dépassé qu'avec une faible probabilité.

7.2.2 Réalisation

La simulation a été réalisée grâce au langage Visual Basic for Applications. Un générateur aléatoire décrit dans la suite a été utilisé pour simuler une variable aléatoire entre 0 et 1.

La simulation proposée permet de simuler la vie d'un assuré. Elle est effectuée pour chaque assuré composant le portefeuille. Une simulation du portefeuille consiste en la simulation de chaque assuré composant le portefeuille, autonome comme dépendant. Cette opération est répétée autant de fois que nécessaire pour que la loi des grands nombres puissent être appliquée.

Nous distinguons la simulation d'un assuré dépendant et d'un assuré autonome.

— Simulation 1 : En premier lieu, la simulation d'un assuré autonome

1. La simulation est initialisée par la vérification de la survie de l'assuré autonome.
   (a) Si l'assuré survit, alors nous vérifions s'il entre en dépendance.
      i. S'il tombe en dépendance, nous passons à la simulation 2
      ii. Si l'assuré ne tombe pas en dépendance, alors la simulation boucle, nous retournons au point 1. de la Simulation 1, en incrémentant l'âge de l'assuré.
   (b) Si l'assuré décède, la simulation est terminée.

— Simulation 2 : Simulation d'un assuré dépendant (simulation de la cause de dépendance)
   1. Nous vérifions d'abord pour quel motif l'assuré tombe en dépendance.
      (a) Si l'assuré tombe en dépendance pour cause d'accident, peu importe la durée de souscription de la police, nous l'affectons en dépendance et nous passons à la simulation 3
      (b) Si l'assuré tombe pour cause de maladie
         i. Si l'assuré est présent dans le portefeuille depuis plus de 1 an, alors nous passons à la simulation 3
         ii. S'il est présent depuis moins de 1 an, alors le délai de carence pour ce type de dépendance n'est pas encore terminé, l'assuré n'est pas couvert, la simulation s'arrête.
      (c) Si l'assuré tombe pour cause de démence
         i. Si l'assuré est présent dans le portefeuille depuis plus de 3 ans, alors nous passons au point 2.
         ii. S'il est présent depuis moins de 3 ans, alors le délai de carence pour ce type de dépendance n'est pas encore terminé, l'assuré n'est pas couvert, la simulation s'arrête.

— Simulation 3 : Simulation d'un assuré dépendant (simulation du type de dépendance)
   1. La simulation est alors initialisée par l'affectation de l'assuré en dépendance partielle ou en dépendance totale selon la proportion observée de dépendance partielle dans le portefeuille.
      (a) Si l'assuré entre en dépendance partielle, nous passons à la simulation 4.2.
      (b) Sinon il est dépendant total, nous passons aussi à la simulation 4.1.

— Simulation 4 : Simulation d'un assuré dépendant (simulation de la vie de l'assuré dépendant)
   1. En cas de dépendance totale, nous évaluons sa probabilité de décéder :

(a) S'il décède, la simulation s'arrête.

(b) Sinon la simulation boucle, nous retournons au point 1 de cette simulation

2. En cas de dépendance partielle, nous évaluons dans un premier temps si l'assuré décède en dépendance partielle ou pas (selon le même principe que la dépendance totale, mais avec la mortalité des dépendants partiels).

   (a) S'il décède, la simulation s'arrête.

   (b) Sinon la simulation continue et nous regardons si l'assuré passe en dépendance totale ou non.

      i. Si l'assuré n'entre pas en dépendance totale, alors la simulation boucle, l'âge est incrémenté et nous retournons au point 2 de cette simulation

      ii. Sinon il reste en dépendance partielle, la simulation boucle aussi, l'âge est incrémenté et nous retournons au point 1 de cette simulation

En voici un schéma explicatif :

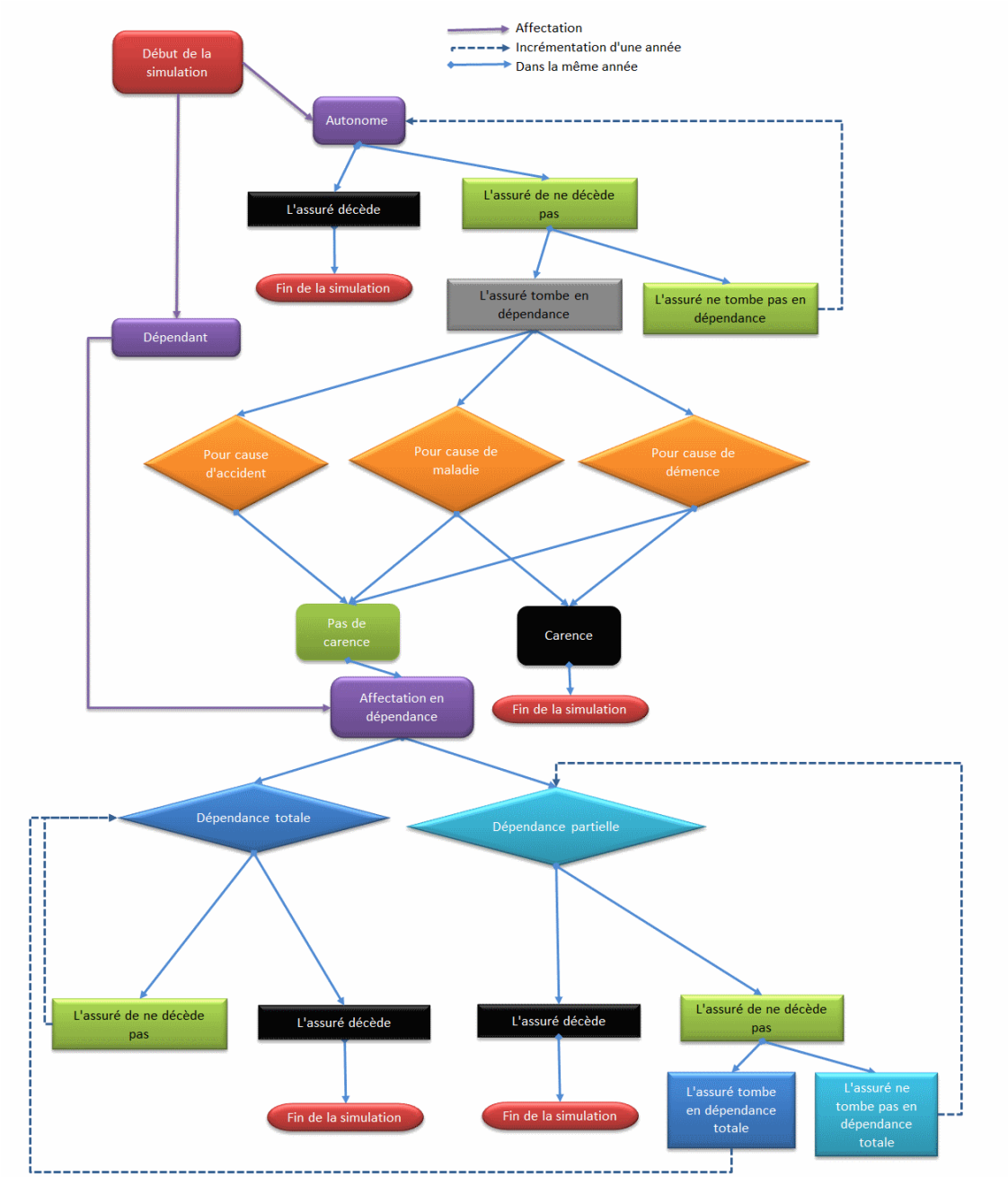

Illustration 7.1 : *Schema de la simulation de Monte Carlo*

A la fin de la simulation, si l'assuré est resté autonome, alors nous considérons qu'il a payé sa prime. De même, nous calculons la provision pour risque croissant afférant à sa police. Dans un but de prudence, nous considérons que l'assuré ne paye sa prime qu'à la fin de l'échéance quand bien même le paiement de la prime est à terme à échoir. En effet une grande majorité des assurés paye de manière mensuelle ou trimestrielle. Ainsi, ne pouvant simuler de manière précise où intervient le décès ou l'entrée en dépendance de l'individu dans l'année, sauf hypothèse de répartition intra annuelle sur les probabilités, nous considérons que l'assureur ne percevra la prime que si l'assuré n'est pas décédé ou entré en dépendance. Dans le cas contraire,

soit le contrat se termine, soit l'assureur devra verser une rente. Nous faisons ainsi l'hypothèse que le décès ou la dépendance survient en début de période, afin de ne pas sur évaluer le montant des cotisations perçues.

En revanche, si l'assuré est entré ou reste en dépendance, alors la prestation due lui est versée, et nous calculons la provision mathématique pour la rente qui lui sera versée dans le futur.

La simulation est réitérée pour chaque assuré du portefeuille, autonome comme dépendant, jusqu'au maximum entre un horizon de temps défini et la mort de l'individu. Une simulation du portefeuille consiste en la simulation de chaque assuré du portefeuille. La simulation du portefeuille est répétée un certain nombre de fois afin d'assurer la convergence, c'est à dire que l'ajout d'une simulation ne modifie pas significativement les résultats. Nous pouvons nous baser sur le critère de la variance : si la variance de l'estimation ne diminue plus avec l'ajout d'une simulation du portefeuille supplémentaire, nous considérons que la simulation a convergé.

7.3 Validation des résultats

Afin de valider les résultats obtenus par la simulation, nous procédons à différents contrôles de cohérence.

Le calcul du nombre d'assurés autonomes, dépendants, décédant sans être dépendants, entrant en dépendance, ou décédant en dépendance est effectué pour chaque année simulée. De même, nous calculons l'âge moyen de ces individus. Ces deux éléments nous permettent de contrôler la cohérence du nombre de décès (autonome ou dépendant) et d'entrée en dépendance. En effet, en notant :

- $x^{j,c}$ l'âge moyen du portefeuille simulé l'année j, pour la catégorie c (avec c = autonome ou dépendant)
- $N_{j,c}$ l'effectif sous risque l'année j, pour la catégorie c
- $D_{j,c}$ le nombre de sortie l'année j pour la catégorie c
- q_x^a la probabilité de décès entre l'âge x et x+1 d'un autonome
- q_x^i la probabilité de décès entre l'âge x et x+1 d'un dépendant
- i_x la probabilité d'entrée en dépendance entre l'âge x et x+1 d'un autonome

Les grandeurs suivantes devraient vérifier :

$$q_x^a \cdot N_{j,autonome} \simeq D_{j,autonome}$$

,

$$q_x^i \cdot N_{j,dépendant} \simeq D_{j,dépendant}$$

et

$$i_x \cdot (1 - q_x^a) \cdot N_{j,autonome} \simeq N_{j,dépendant} - (N_{j-1,dépendant} - D_{j-1,dépendant})$$

7.4 Limites du simulateur intégré dans VBA

Les limites de cette modélisation repose sur l'utilisation du générateur aléatoire d'Excel. Ce dernier est connu pour son manque "d'aléatoire" et largement décrié par la communauté scientifique [2]. Nous préférons alors utiliser le générateur aléatoire de Mersenne et Twister [3], considéré comme un des générateurs aléatoire les plus puissants. Nous utilisons le code écrit par Makoto Matsumoto et Takuji Nishimura, développeurs de cet algorithme, et traduit en VBA par Pablo Mariano Ronchi.

Dans la partie suivante, nous présenterons les simulations obtenues par ce procédé grâce aux estimations des différentes probabilités obtenues précédemment.

2. McCullough, B. D., and David A. Heiser. "On the accuracy of statistical procedures in Microsoft Excel 2007." Computational Statistics Data Analysis 52.10 (2008) : 4570-4578.

3. Matsumoto Makoto, and Takuji Nishimura. "Mersenne twister : a 623-dimensionally equidistributed uniform pseudo-random number generator." ACM Transactions on Modeling and Computer Simulation (TOMACS) 8.1 (1998) : 3-30.

Partie 8

Résultats des simulations

8.0.1 Simulation des différents éléments

Dans la suite nous appelons :

Pour les simulations concernant la mortalité des autonomes :

- l'estimation centrale, obtenue par régression logistique des taux bruts de mortalité des observations sur les taux de décès de la table TH-TF,
- l'estimation inférieure, où cette fois-ci l'intervalle de confiance inférieur à 95% a été régressé sur les mêmes taux que ceux utilisés pour l'estimation centrale,
- l'estimation supérieure, où cette fois-ci l'intervalle de confiance supérieur à 95% a été régressé sur les mêmes taux que pour l'estimation centrale,
- l'estimation Coale, consistant en la partie ajustée par régression logistique pour la partie avant la clôture, et un ajustement par la méthode de Coale et Kisker pour les âges après la clôture.

Pour les simulations concernant l'incidence :

- l'estimation centrale, obtenue par régression logistique des taux bruts d'entrée en incidence des observations sur les taux de décès de la table TH-TF,
- l'estimation inférieure, où cette fois ci l'intervalle de confiance inférieur à 95% a été régressé sur les mêmes taux que pour l'estimation centrale,
- l'estimation supérieure, où cette fois ci l'intervalle de confiance supérieur à 95% a été régressé sur les mêmes taux que pour l'estimation centrale.

Pour les simulations concernant la mortalité des dépendants :

- l'estimation centrale, constituée par le lissage de Whittaker Henderson des taux bruts estimés pour la partie connue de la table, et la table TH-TF pour la partie inconnue,
- l'estimation inférieure, constituée par l'intervalle de confiance inférieur à 95% des taux bruts estimés par la méthode de Kaplan Meier et se poursuivant par la table TH-TF après la clôture,
- l'estimation supérieure, constituée par l'intervalle de confiance supérieur à 95% des taux bruts estimés par la méthode de Kaplan Meier et se poursuivant par la table TH-TF après la clôture,

- l'estimation Bis, constituée de la même façon que l'estimation centrale, sauf que la TH-TF est majorée par un coefficient SMR.

Modèles retenus Les modèles présentés sont les estimations effectuées lors de la partie sur l'estimation des lois.

Mortalité des autonomes Pour la mortalité des autonomes, nous avons retenu 2 modèles, à savoir l'ajustement de la table mixte TH-TF par le modèle relationnel avec extrapolation des taux de mortalité aux âges élevés par prolongation du modèle et le même ajustement mais avec une extrapolation des taux de mortalité aux âges élevés grâce au modèle de Coale et Kisker.

Nous n'avons pas retenu l'abattement de la table par le coefficient SMR, moins précis que le modèle relationnel, puisque ne dépendant pas de l'âge. Nous n'avons pas retenu le lissage des taux bruts par la méthode de Whittaker Henderson, celle ci n'offrant pas une extrapolation directe des taux, et surtout nécessitant un lissage afin d'être utilisable.

Les estimations obtenues par les deux modèles cités ci-dessus seront comparées, puisque nous avions remarqué que l'extrapolation par le modèle de Coale et Kisker entraînait des taux de mortalité significativement plus élevés que ceux obtenus par prolongation du modèle relationnel.

Entrée en dépendance Pour la loi d'entrée en dépendance, nous avons retenu le modèle relationnel afin d'abattre la table mixte TH-TF, en remarquant qu'un facteur commun caractérisait la mort et l'entrée en dépendance. Cette méthode, intuitive, permet l'obtention de taux d'incidence lisses et une extrapolation, en comparaison du lissage des taux bruts.

Mortalité des dépendants Concernant la mortalité des dépendants, nous avons retenu deux modèles. Les deux se caractérisent par une base commune. En effet, ceux ci sont composés d'abord d'une estimation des taux bruts, obtenue via la méthode de Kaplan et Meier. Puis, pour estimer la mortalité des dépendants pour lesquelles nous n'avions pas d'observations, nous avons utilisé la mortalité de la table mixte TH-TF, pour le premier modèle la table TH-TF telle quelle, pour le deuxième une majoration de cette table par un coefficient de type SMR. Pour rappel, l'étude de la mortalité des dépendants se fait par âge à l'entrée en dépendance.

N'ayant aucune connaissance de la mortalité des dépendants pour des durées longues, ces deux modèles ont été retenus afin de déterminer si l'hypothèse prise quant à la mortalité des dépendants après une certaine période a une incidence ou si c'est principalement l'estimation des taux de mortalité pour les premières années qui est essentiel.

Nous comparerons les résultats obtenus par ces deux modèles dans la suite.

Intervalle de confiance inférieur et supérieur à 95% Pour chacun de ces modèles, et afin de profiter pleinement de la puissance des simulations, nous avions déterminé un intervalle de confiance supérieur et inférieur à 95%. Cet intervalle nous servira à déterminer un scénario optimiste et un scénario pessimiste. L'estimation centrale se veut comme la meilleure estimation possible de l'élément simulé.

Horizon de simulation Nous procédons à une simulation sur un horizon de 25 années. Ainsi, nous allons tenter de projeter la mortalité des autonomes du portefeuille, l'entrée en dépendance de ces mêmes individus et la mortalité des dépendants jusqu'en 2037.

Simulation du stock de contrat Nous simulons le stock de contrats, c'est à dire que nous considérons que plus aucune adhésion n'est effectuée après le date de début de la simulation. En effet, il n'est pas nécessaire de simuler de nouvelles adhésions puisque nous cherchons à visualiser les possibles pertes liées au stock de contrats. Si nous nous apercevons d'une sous tarification de ce stock, il faudra chercher des solutions pour redresser le résultat de celui-ci, et modifier la tarification des futurs contrats afin que ceux ci ne présentent pas un résultat futur déficitaire.

Présentation préliminaire de 1000 simulations Nous présentons dans la suite les résultats pour 1000 simulations du portefeuille par année, ce qui est équivalent à 1000 simulations par assuré, le portefeuille comportant environ 46 000 contrats actifs, cela représente ainsi près de 46 000 000 de simulations par années.

Nous présentons le résultat des éléments simulés pour le scénario central. Les éléments simulés sont les cotisations, les prestations, les PRC et les PMR. Il s'agit d'une vision réassureur, ainsi ces résultats ne représentent pas la réalité de la cédante, mais uniquement celle du réassureur (transformés pour des raisons de confidentialité). Toutefois, le sort de l'assureur est identique au sort du réassureur puisqu'il s'agit d'un traité en Quote Part. Chacune des trajectoires que nous présentons correspond à la réalisation pour l'ensemble du portefeuille d'un algorithme tel que présenté dans la partie précédente pour chacun des assurés durant 25 ans.

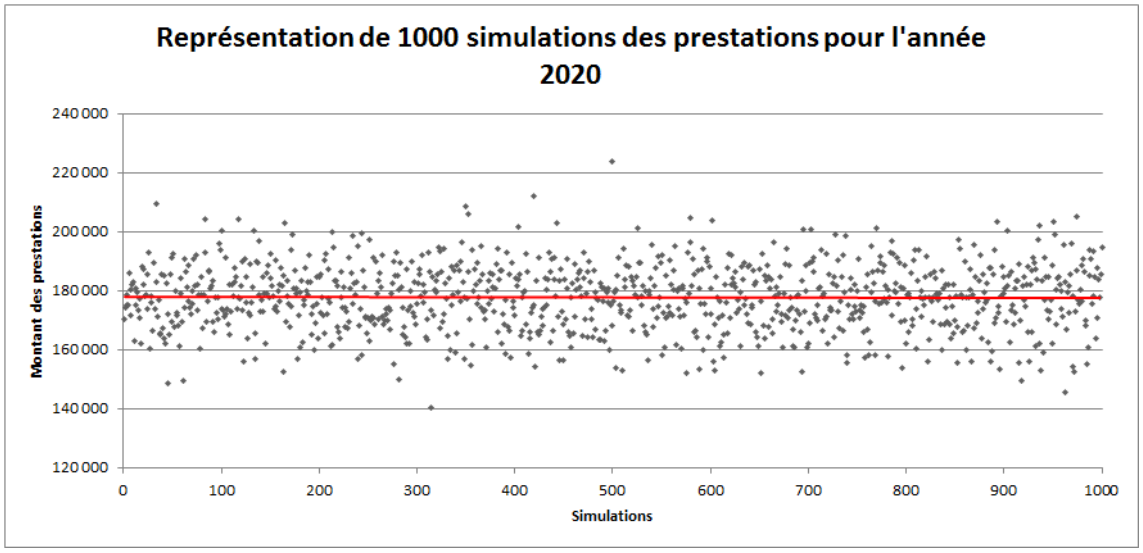

Illustration 8.1 : *Exemple de 1000 simulations des prestations pour l'année 2020*

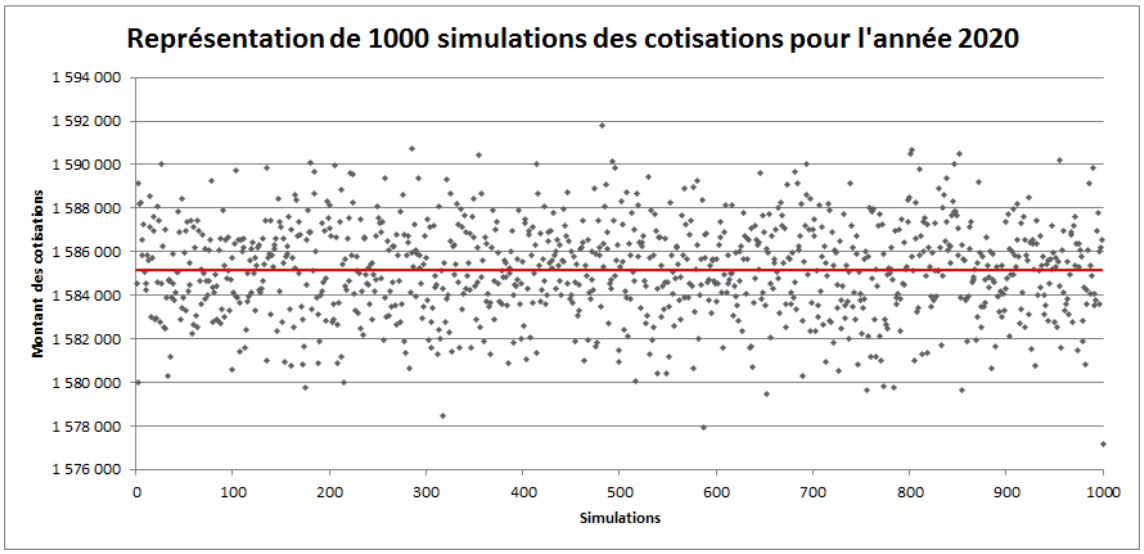

Illustration 8.2 : *Exemple de 1000 simulations des cotisations pour l'année 2020*

Nous pouvons ainsi observer l'intérêt des simulations : chacune des simulations représente une des réalisations possible pour les éléments présentés. De même les points isolés représentent tout autant de valeurs extrêmes pouvant être prises par les prestations ou les cotisations.

8.0.2 Illustration de la convergence

Afin de choisir le nombre de simulations nécessaire à la convergence des estimations sur un horizon de 25 ans, nous utilisons comme mesure synthétique la moyenne sur 25 ans de la variance des éléments simulés en fonction du nombre de simulation. Plus précisément, par exemple pour les cotisations C_a^s de l'année a et étant le résultat de la $s^{ème}$ simulation, nous avons la mesure de la variance V :

$$V = \frac{1}{H}\sum_a [\frac{1}{N-1}\sum_s (C_a^s - \bar{C}_a)^2]$$

avec : $s \in [1, N]$ et $a \in [2014, 2014 + H]$ et $\bar{C}_a = \frac{1}{N} \sum_s C_a^s$
où N est le nombre de simulations et H l'horizon de simulation.

Lorsque la variance ne diminue plus significativement, nous pouvons considérer qu'il y a convergence, c'est à dire que l'ajout d'une simulation n'améliorerait pas la précision des résultats.

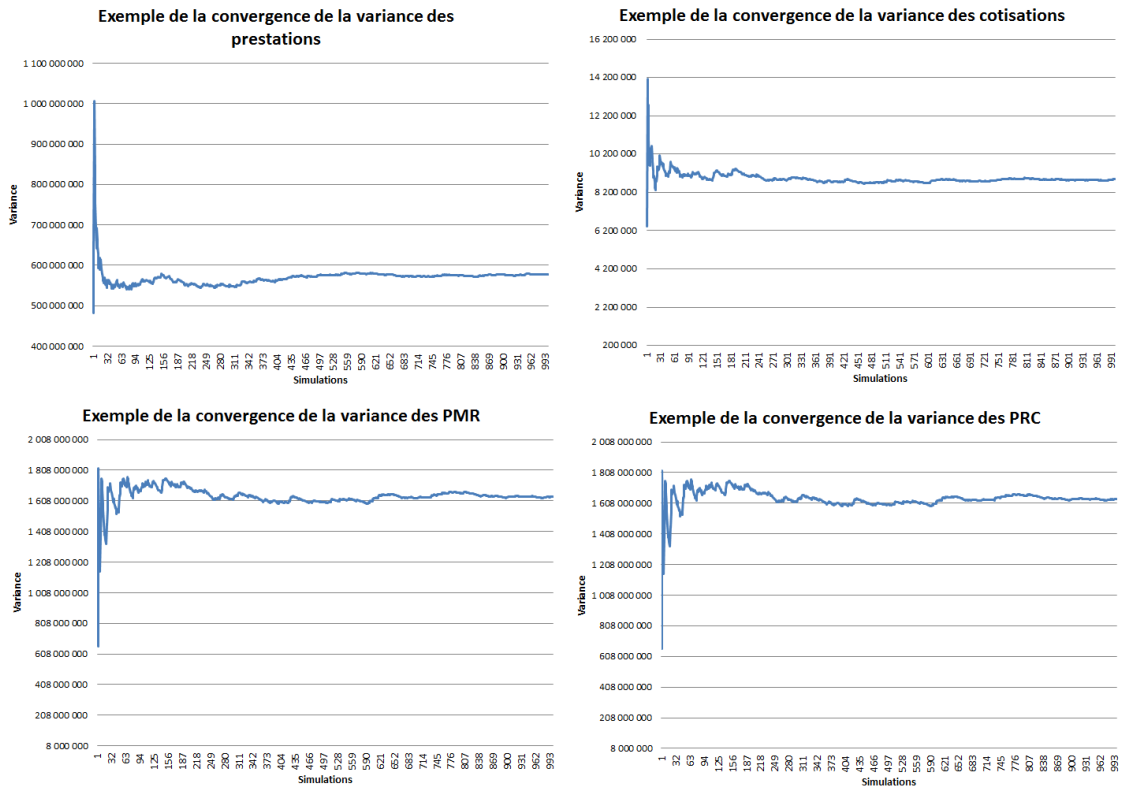

Illustration 8.3 : *Convergence de la variance*

Nous observons ainsi que la convergence est assez rapide (cf. Illustration 8.3). Excepté pour les PRC et les PMR qui nécessitent plus de simulations que les cotisations et les prestations pour que leurs variances convergent, nous constatons qu'après 300 simulations, la variance a convergé de manière stable. Ainsi, dans la suite de cette étude, nous nous servirons de 300 simulations. Notons aussi que les résultats présentés ci-dessus ne concernent que les résultats bruts, sans application de la Quote Part sur les différents éléments. Le but ici est simplement de déterminer le nombre de simulations nécessaires afin d'obtenir dans la suite des résultats cohérents.

Situation initiale des comptes Dans la suite, le montant sur les graphiques renseignés pour l'année 2013 ne correspond pas à un montant simulé, il s'agit des éléments du compte de résultat du réassureur au 31/12/2013 pour l'année passée (transformés).

Différence entre l'estimation centrale et bis des dépendants Nous avions émis un doute quant à la modélisation de la mortalité des dépendants. En effet, nous avions effectué deux estimations : l'estimation centrale et l'estimation bis, l'une sous-estimant, l'autre surestimant la mortalité des dépendants après la clôture de la table de mortalité des dépendants. Nous avions émis l'hypothèse que la différence entre le résultat prospectif simulé par ces deux méthodes, toute chose égale par ailleurs, n'entraînerait pas une différence trop importante dans les résultats, dans la mesure où l'espérance de survie des dépendants est estimée largement inférieure à la durée de vie au moment de la clôture de la table. Nous effectuons ainsi des simulations pour comparer les résultats obtenus par ces deux approches de modélisation.

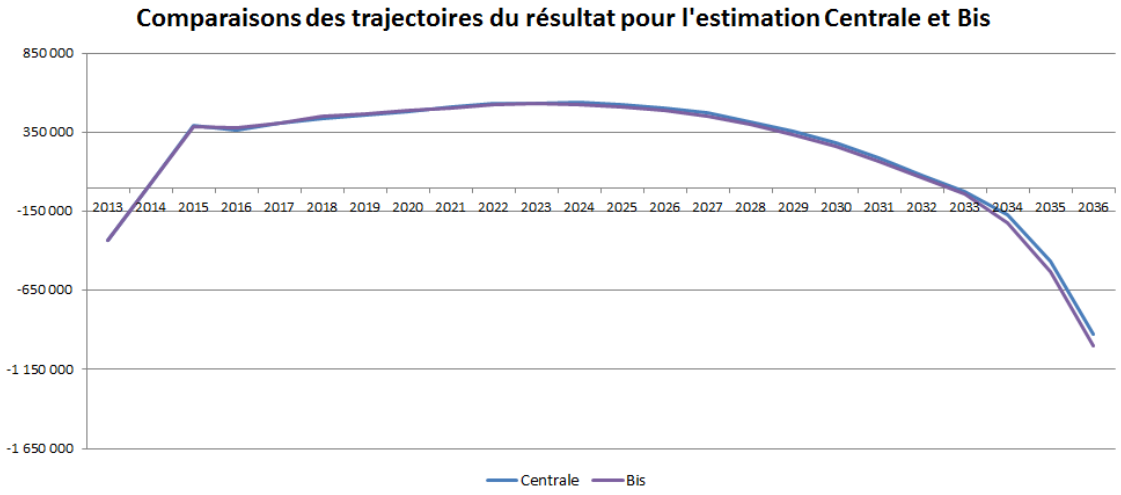

Illustration 8.4 : *Comparaison des simulations du résultat pour l'estimation centrale et l'estimation Bis de la mortalité des dépendants*

Nous nous apercevons que la différence entre les deux estimations n'est pas significative, voir négligeable pour la première dizaine d'années de simulation (cf. Illustration 8.4). L'estimation Centrale présente un résultat légèrement supérieur à l'estimation Bis. Or l'estimation Centrale était censée sur-estimer la longévité des dépendants. Ce résultat contre intuitif provient du lissage effectué et confirme toutefois notre hypothèse que l'estimation des premières années passées en dépendance est cruciale. Nous effectuerons dans la suite les simulations avec l'estimation Centrale, puisque les résultats obtenus avec celle-ci sont peu différents de ceux obtenus avec la simulation utilisant l'estimation Bis.

Explication de la hausse rapide du résultat Comme signalé précédemment, le résultat apparaissant en 2013 n'est pas simulé, il s'agit de la réalité des comptes. Nous nous apercevons que celui-ci est négatif, et ce, nous le précisons, depuis un certain nombre d'années. Le produit a d'une part été sous tarifié, mais a aussi souffert d'une baisse constante des taux d'intérêt depuis les années 2000.

Ainsi la forte hausse anticipée du résultat pour les années à venir résulte d'une

hausse de tarif survenue les années précédentes. En effet, ayant pris conscience d'un problème de sous tarification, l'assureur et le réassureur ont décider d'augmenter les cotisations afin d'obtenir à nouveau un résultat positif.

Cependant, la question qui se pose est de savoir si cette hausse des cotisations suffira à pallier la sous tarification. Quant à la baisse du taux technique, actuellement déjà très bas, son potentiel de baisse n'est plus aussi important que par le passé. C'est pourquoi dans cette étude nous nous concentrons particulièrement sur les effets de la mortalité et de l'incidence, même si le taux technique a lui aussi un rôle majeur dans la rentabilité d'une couverte à long terme du type dépendance.

8.0.3 Présentation des différents éléments simulés avec l'estimation centrale

Par la suite, nous ne représentons que 100 trajectoires, afin de pouvoir mieux observer les graphiques.

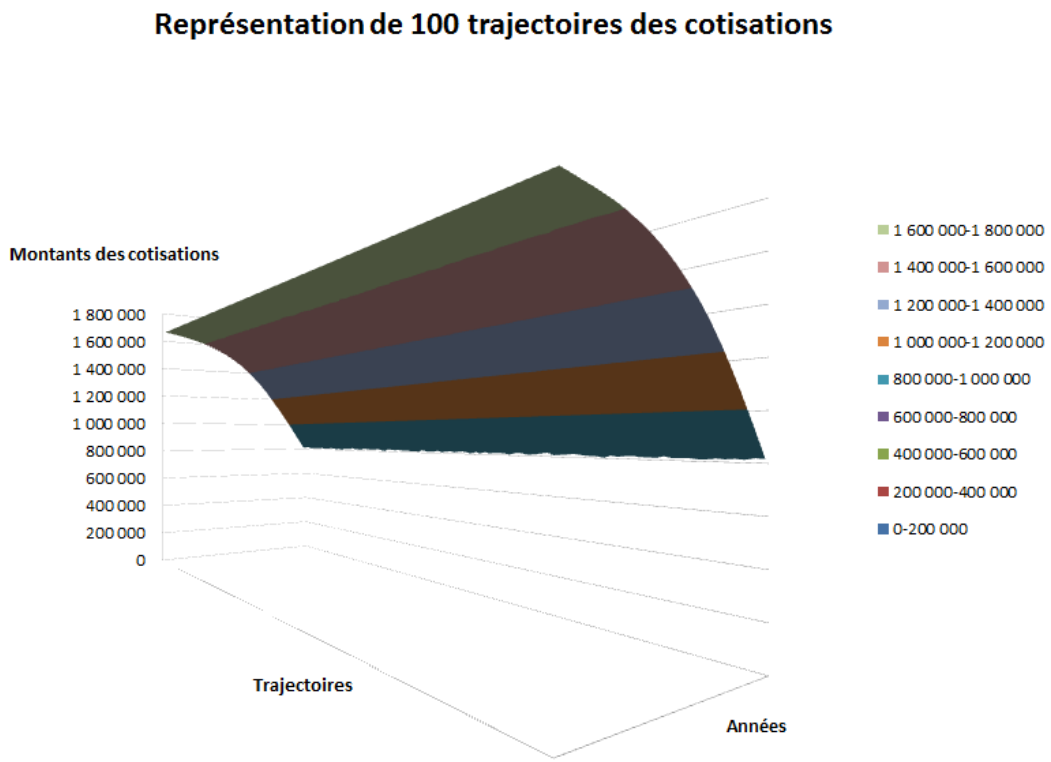

Illustration 8.5 : *Exemple de 100 trajectoires des cotisations sur 25 ans*

Comme nous pouvons le constater, les cotisations ne font que diminuer au fil du temps (cf. Illustration 8.5). Cela s'explique par l'absence de nouvelles adhésions. Au fur et à mesure, les autonomes meurent ou entrent en dépendance, entraînant une baisse du montant des cotisations perçues.

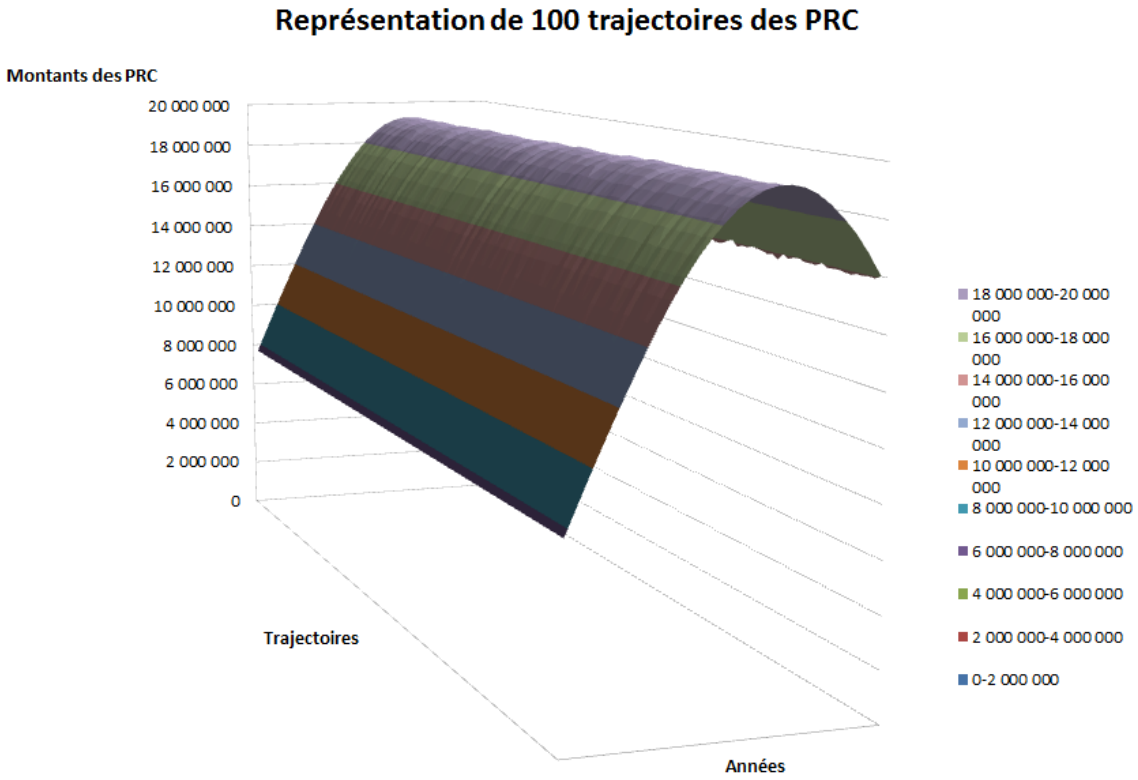

Illustration 8.6 : *Exemple de 100 trajectoires des PRC sur 25 ans*

Les Provisions pour Risques Croissants, ayant pour but d'anticiper la survenue de la dépendance des individus autonomes, évoluent en fonction du nombre d'individus non dépendants et de l'évolution du risque de dépendance. Elles présentent ainsi d'abord une phase croissante, où l'assureur continue de provisionner le risque de dépendance des individus. Puis elles sont ensuite décroissantes : d'une part elles sont libérées pour les individus entrant en dépendance ou décédés et d'autre part, la balance engagement de l'assureur contre engagement de l'assuré redevient en faveur de l'assureur en théorie (cf. Illustration 8.6). Néanmoins il se peut que les prestations réellement versées deviennent supérieures à celles estimées dans les PRC.

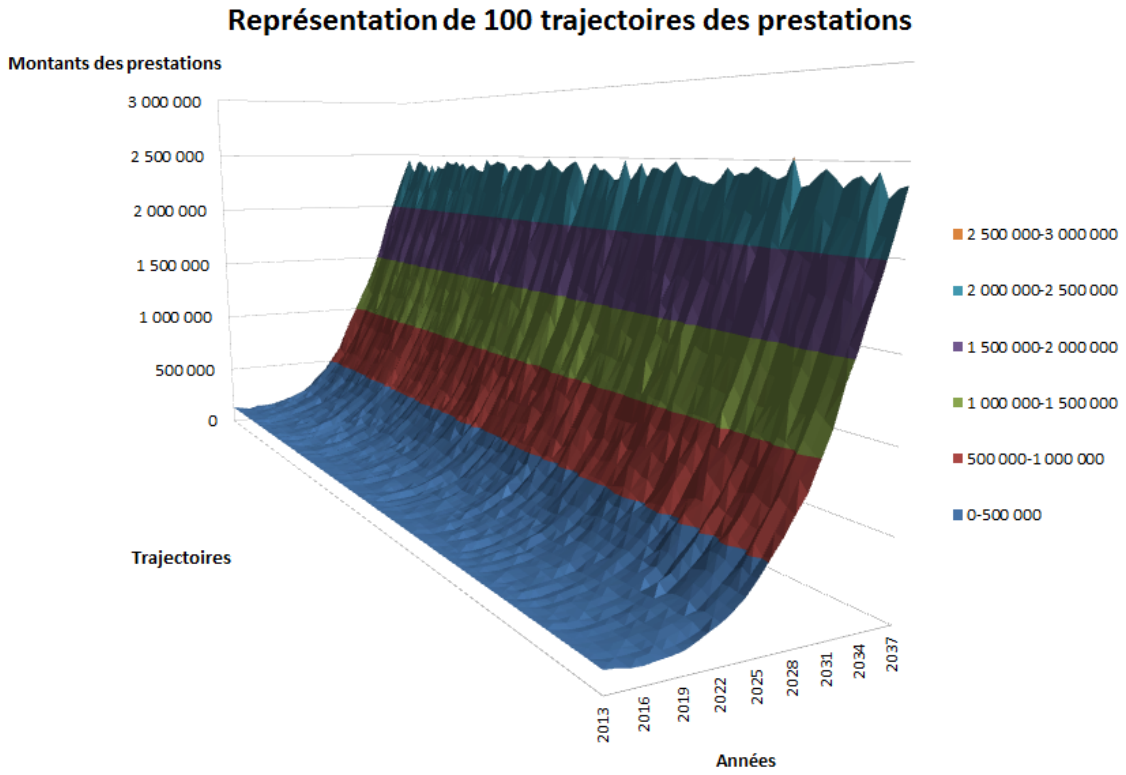

Illustration 8.7 : *Exemple de 100 trajectoires des prestations sur 25 ans*

Quant aux prestations, elles ne font qu'augmenter (cf. Illustration 8.7). En effet, le portefeuille vieillissant par absence de renouvellement de la population, de plus en plus d'assurés sont victimes de dépendance, puisque la probabilité d'entrée en dépendance augmente avec l'âge. Ainsi, l'assureur doit faire face au paiement de plus en plus de prestations à destination des dépendants. Ce montant devrait toutefois atteindre une valeur maximale puis baisser par la suite, puisque les dépendants vont mourir peu à peu et ne seront pas remplacés, l'effectif sous risque étant devenu faible.

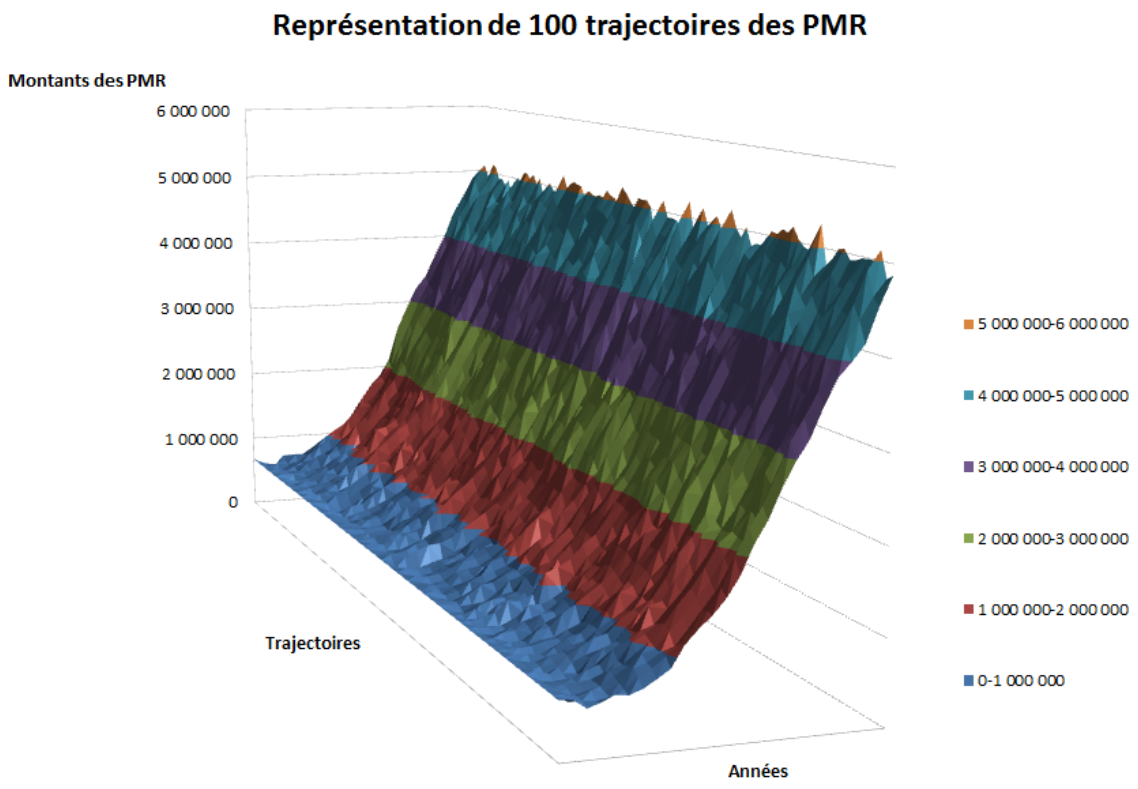

Illustration 8.8 : *Exemple de 100 trajectoires des PMR sur 25 ans*

Les Provisions Mathématiques de Rente, ayant pour but d'anticiper le montant des prestations futures versées à l'individu dépendant, évoluent en réponse du nombre d'individus en dépendance. Puisqu'il y a de plus en plus d'individus en dépendance, il est alors normal que l'assureur ait à constituer des provisions de plus en plus importantes (cf. Illustration 8.8). Elles devraient toutefois diminuer à terme, puisque les dépendants décèdent.

Remarquons aussi la moins grande homogénéité des différents trajectoires concernant les prestations et le PMR en comparaison des cotisations et des PRC. Elle s'explique par un nombre d'assurés en dépendance faible en comparaison du nombre d'individus autonomes. L'effet de mutualisation observé précédemment pour les cotisations n'est plus à l'oeuvre ici.

Pour l'ensemble de ces graphiques, nous constatons une grande homogénéité des simulations. En effet, puisque le portefeuille comporte environ 46 000 individus, une première mutualisation est effectuée intra-simulation. Les "mauvais résultats" d'un assuré compensent les "excellents résultats" d'un autre. Ce que nous entendons par mauvais et excellents résultats correspond aux résultats se trouvant respectivement

en-dessous ou au-dessus de la moyenne prévue. Par exemple si nous avions prévu qu'un assuré autonome avait une espérance de vie résiduelle sans dépendance de 10 ans, mais qu'il en vit 15 sans dépendance, il s'agit d'un bon résultat pour la compagnie. A l'inverse, si cet individu entre en dépendance au bout de 5 ans, il s'agit d'un mauvais résultat.

8.0.4 Simulation des différents hypothèses concernant les autonomes

Nous procédons à la simulation proposée précédemment en comparant les résultats obtenus en fonction des différents scénarios pour la mortalité des autonomes, à savoir le scénario central, le scénario inférieur, le scénario supérieur et le scénario Coale et Kisker. Nous comparons les résultats de ces différents scénarios toutes choses égales par ailleurs, en utilisant le scénario central pour l'incidence et la mortalité des dépendants (totale et partielle).

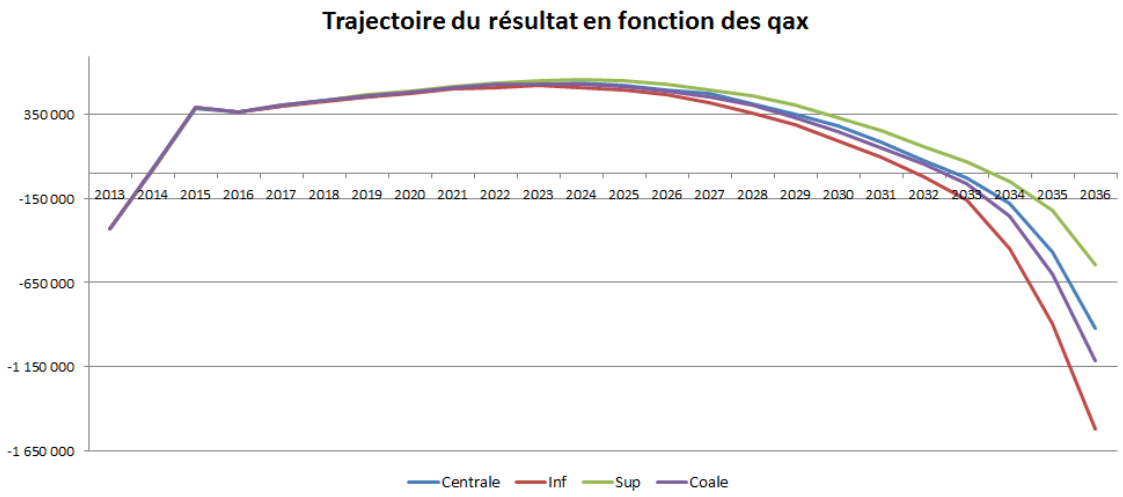

Illustration 8.9 : *Trajectoire du résultat en fonction des qax*

Nous constatons aussi que les différentes hypothèses ont peu d'impact sur le résultat de moyen terme. Néanmoins, à long terme, la faible mortalité anticipée aux grands âges par l'intervalle de confiance inférieure entraîne les pertes les plus importantes. La simulation avec l'intervalle de confiance supérieur, donc estimant une mortalité forte des individus autonomes, est l'hypothèse pour laquelle le résultat à long terme est le plus élevé.

Cette hiérarchisation s'explique par l'estimation d'une exposition au risque plus grande si les assurés autonomes ont une mortalité plus faible.

Puisque la loi d'entrée en dépendance est particulièrement convexe, indiquant une probabilité de plus en plus forte d'entrée en dépendance, plus important sera l'exposition, notamment aux âges élevés, plus important sera le nombre d'individus entrant en dépendance, toutes choses égales par ailleurs.

Nous observons ainsi que la tarification du produit ne semble pas avoir anticipé une durée de vie en autonomie aussi longue que celle utilisée par nos simulations.

8.0.5 Simulation des différents hypothèses concernant l'incidence

Cette fois, nous comparons les différentes hypothèses concernant l'entrée en dépendance des individus. Nous comparons ainsi :

- l'estimation centrale ;
- l'estimation inférieure ;
- l'estimation inférieure.

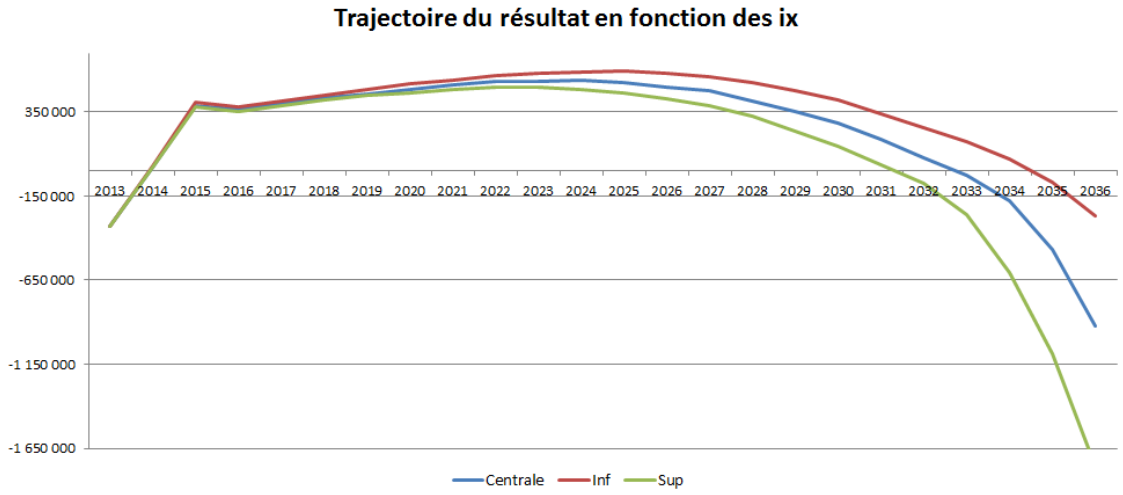

Illustration 8.10 : *Trajectoire du résultat en fonction des ix*

Les conclusions à court et moyen terme restent identiques aux résultats précédents, à savoir une hausse du résultat à court terme, entraînée par une hausse des cotisations, ainsi qu'une faible différenciation des simulations des différentes estimations à moyen terme.

Toutefois, à long terme, la simulation effectuée au moyen de l'estimation supérieure des taux d'incidence, c'est à dire une prévision d'entrée en dépendance plus forte que dans les autres estimations, fait apparaître un risque de perte important.

En revanche, la simulation au moyen de l'intervalle de confiance supérieur montre un résultat quasiment à l'équilibre au terme de la durée considérée. Une nouvelle fois, la tarification du produit ne semble pas avoir anticipé pour les âges élevés un nombre d'entrées en dépendance aussi fort que celui que nous estimons.

8.0.6 Simulation des différents hypothèses concernant les dépendants

Nous simulons à présent le résultat en fonction des différentes lois obtenues pour la mortalité des dépendants. Nous nous servons de l'intervalle de confiance inférieur

et supérieur de l'estimation que nous avions obtenue.

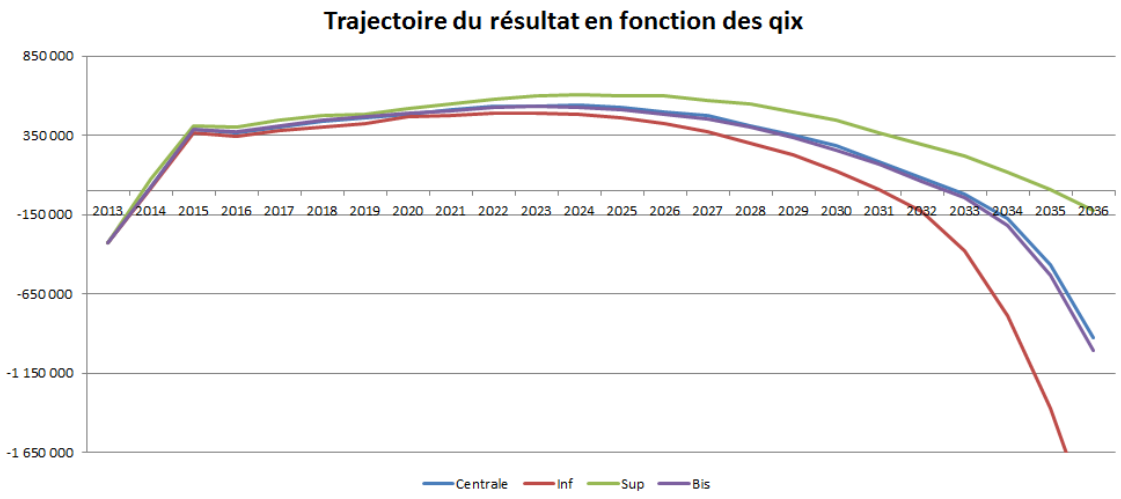

Illustration 8.11 : *Trajectoire du résultat en fonction des qix*

Comme attendu, une mortalité plus faible des dépendants entraîne un résultat plus faible. C'est ce que nous observons avec la simulation obtenue avec l'intervalle de confiance inférieure concernant la mortalité des dépendants, entraînant un résultat à long terme nettement plus bas que pour les simulations utilisant l'estimation centrale et l'estimation par l'intervalle de confiance supérieur.

8.0.7 Simulation de différentes hypothèses de taux de rendement

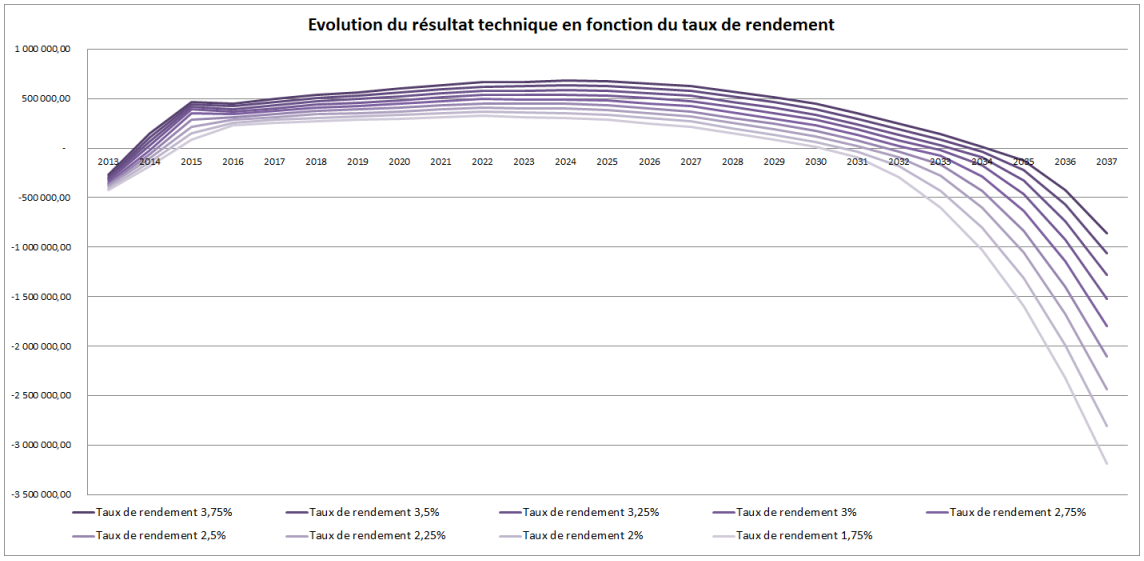

Illustration 8.12 : *Trajectoire du résultat en fonction du taux de rendement des actifs financiers*

L'impact du taux de rendement des actifs financiers est tout autant considérable que l'impact des différentes lois de probabilités. Ce dernier avait été pris égal à

3% dans ce qui précède. Nous constatons alors que ce dernier pourrait permettre de limiter fortement les pertes futures. Au contraire, il les aggraverait s'il venait à diminuer.

Au regard des éléments présentés dans ce qui précède, il apparaît que la tarification de cette garantie ne permet pas de garantir un équilibre : après certain nombre d'années, d'après nos simulations, l'assureur et le réassureur vont se trouver en situation de perte. Cette situation est provoquée par la combinaison de plusieurs éléments.

D'une part, la mortalité des autonomes a été sur-estimée. Ainsi l'exposition au risque de dépendance de l'assureur a été sous-estimé pour les années futures, puisque anticipant un nombre de décès sans dépendance plus important que ce que nous prévoyons.

D'autre part, la sur-estimation de la mortalité des autonomes combine son effet néfaste avec la sous estimation de la loi d'entrée en dépendance. Non seulement le nombre d'individus soumis au risque de dépendance a été sous-estimé, mais également le nombre d'individus entrant en dépendance parmi cet effectif soumis au risque.

De plus, il n'est pas certain que la mortalité des autonomes n'ait pas été sur-estimée. Mais pour confirmer ces hypothèses, une comparaison précise des lois utilisées pour la projection avec les lois utilisées pour la tarification devrait être effectuée.

Quant au taux de rendement des actifs financiers, il pourrait permettre une limitation des pertes.

8.0.8 Solution préventive de la perte future

Constatant le risque de perte future important, mais survenant à une échéance assez longue, nous proposons la mise en place d'un fond d'égalisation. Ce fond d'égalisation, va permettre de mettre en réserve les gains importants qui seront probablement réalisés les années à venir. Sa dotation est constituée de 75% du résultat, si ce dernier est positif.

Nous avons ainsi un nouveau compte de résultat :

Compte annuel de résultat :
+ Cotisation cédées - Indemnités de gestion de la cédante - Frais de gestion du réassureur
+ PRC à l'ouverture - PRC à la clôture
- Prestations réglées cédées + PMR à l'ouverture - PMR à la clôture
+ Produits financiers techniques
- Dotation au fond d'égalisation = X% du résultat +Appurement par le fond d'égalisation
= Résultat de l'exercice

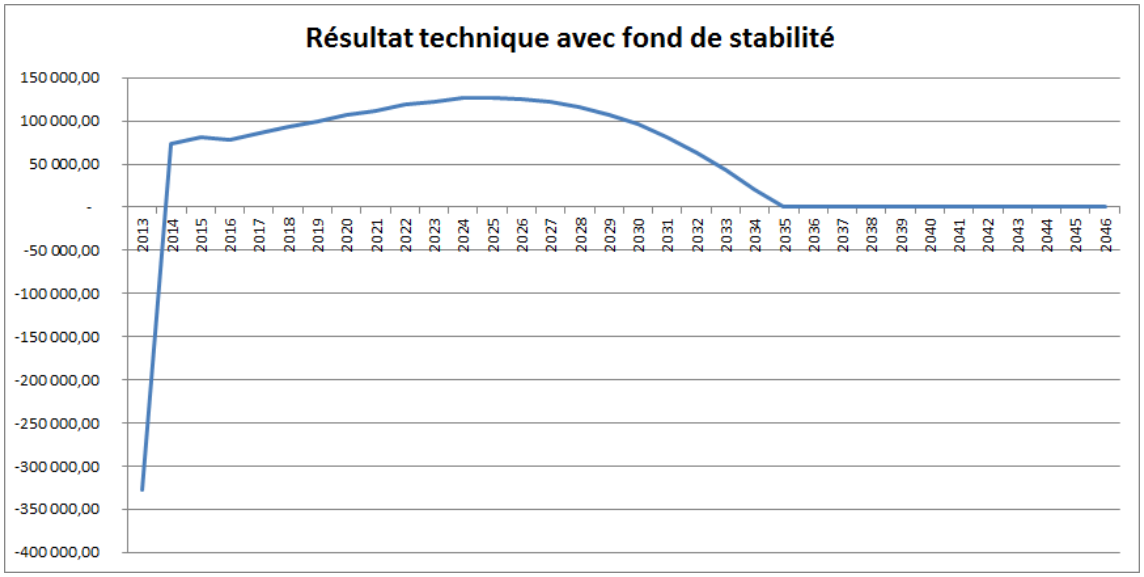

Illustration 8.13 : *Résultat technique simulé avec fond de stabilité*

Ainsi, cette solution offre un répit d'un peu plus de 10 ans au réassureur. Il aurait été intéressant de calibrer la dotation au fond d'égalisation de telle façon à ce qu'il soit entièrement vidé lors de l'extinction de la dernière police. Néanmoins, cette évènement survenant aux alentours de 2070, nous préférons ne pas prêter caution à des simulations si lointaines. La situation sera à surveiller au fil des années.

Afin de constater la variabilité de nos estimations, dans la partie suivante nous calculons la VaR et la TVaR au seuil 95%.

8.0.9 Calcul de la VaR et TVaR

Nous calculons la Value at Risk à 95% des prestations pour chaque année, qui est le montant de prestations qui ne devrait être dépassé qu'avec une probabilité

inférieure à 5%. De même nous déterminons la Tail Value at Risk à 95%, qui quantifie l'espérance de perte sachant que la perte dépasse la Value at Risk à 95%. Ces mesures nous permettent d'obtenir un corridor autour de l'estimation centrale : nous pouvons émettre l'hypothèse que, sous réserve que nos estimations reflètent de la manière la plus juste possible la réalité, le montant des prestations ne dépassera pas un certain seuil, avec une probabilité de se tromper inférieure à 5%. Nous calculons de même la VaR à 5% et la TVaR à 5%.

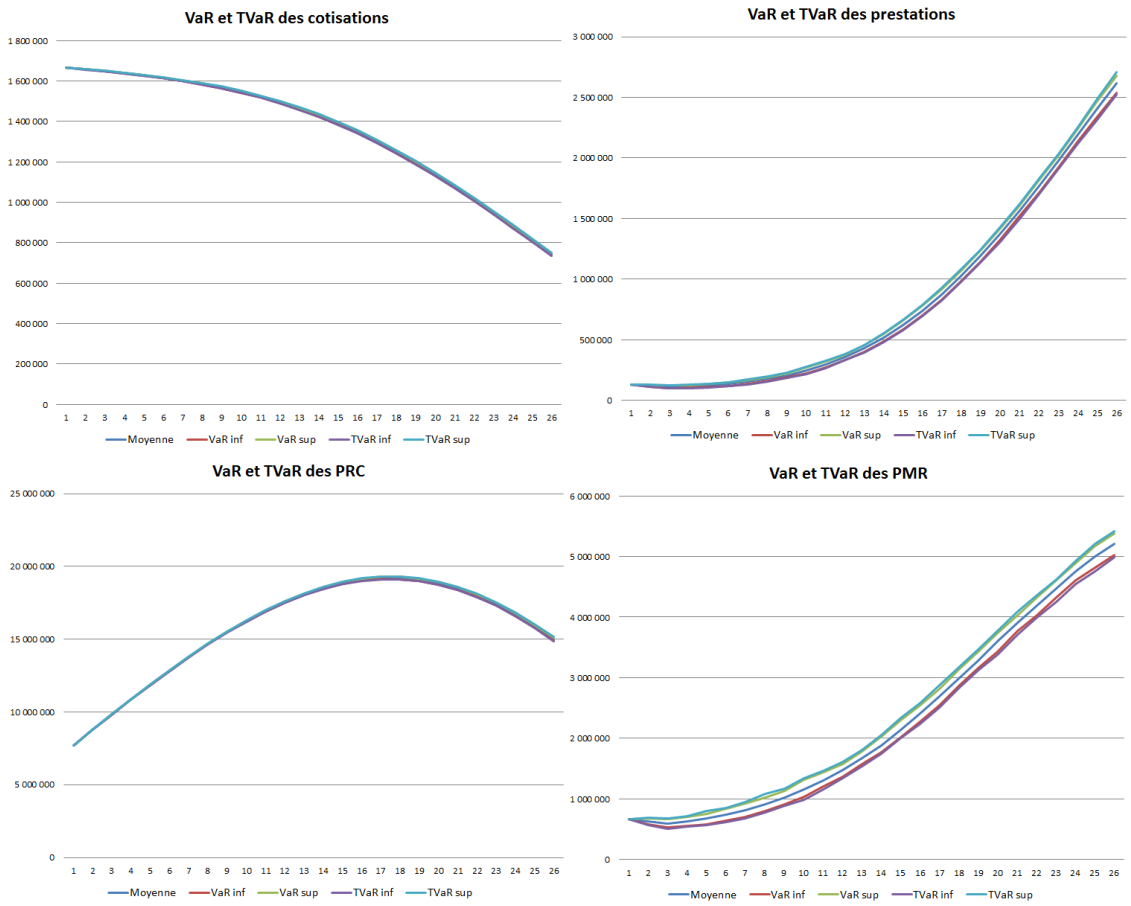

Illustration 8.14 : *VaR, TVaR et moyenne*

Value at Risk, Tail Value at Risk, intervalle de confiance et moyenne sont quasiment confondus pour les simulations des cotisations et des PRC. Les fluctuations éventuelles concernant ces grandeurs ne semblent pas préoccupantes en terme de justesse des simulations.

En revanche, l'estimation concernant les prestations et les PMR présente une plus grande incertitude à en juger par l'écartement de la moyenne avec les valeurs extrêmes. Néanmoins, l'écart est, de manière relative, peu important au regard des montants engagés.

8.0.10 Conclusion des simulations

En conclusion, nous constatons que la récente prise de conscience quant à la sous-estimation des probabilités ayant entraîné une hausse des tarifs devrait porter ses fruits dans les années à venir. En effet, selon nos simulations, le résultat devrait redevenir positif durant les années à venir, et le rester durant de nombreuses années mais redeviendrait négatif à terme, en cause la sous-estimation combinée de la longévité des autonomes et des dépendants ainsi que de l'incidence.

La mise en place d'un fond d'égalisation devrait permettre à l'assureur et au réassureur de protéger leurs résultats. Il s'agit là d'une prévention à long terme.

Partie 9

Conclusion

L'objectif de ce mémoire était de mesurer la rentabilité d'une garantie dépendance commercialisée depuis 14 ans. Nous avons d'abord cherché à estimer les lois propres à ce portefeuille, afin d'être capable d'estimer le nombre de décès des autonomes, le nombre d'entrées en dépendance et la durée de vie en dépendance.

Nous avons alors constaté que la mortalité des individus assurés est bien plus faible que la mortalité des tables TH et TF, que nous avons alors abattues en conséquence afin de projeter au plus juste la mortalité de ces individus, et surtout afin de connaître à chaque instant l'exposition au risque de dépendance.

Par ailleurs, l'estimation de la loi d'entrée en dépendance à montrer une tendance forte : la probabilité d'entrée en dépendance augmente très fortement avec l'âge. Le problème de l'estimation de l'incidence pour les âges non encore observés, principalement après 90 ans, s'est alors posé. Grâce à la modélisation proposée pour estimer l'incidence en fonction de la mortalité d'une table de référence, nous avons pu extrapoler cette incidence aux âges élevés.

En outre, l'estimation des lois de mortalité des dépendants nous a permis de constater la mortalité particulièrement élevée des dépendants lors de la première et deuxième année de dépendance. De plus, il est clairement apparu que la mortalité des individus étant devenus dépendants assez jeunes était plus forte que la mortalité de ceux l'étant devenus plus âgés. Il s'est aussi posé le problème d'estimer la mortalité des individus étant en dépendance depuis une longue période, car n'ayant pas d'observations de ces individus. Nous avons alors utilisé deux méthodes d'alignement sur la mortalité d'une population de référence.

Grâce à ces estimations, nous avons projeté le résultat futur lié à cette garantie, en simulant la vie de chaque assuré du portefeuille, autonome comme dépendant, grâce à des simulations de Monte Carlo.

Ces simulations nous ont permis d'observer une sinistralité future imprévue par les lois utilisées pour tarifier cette garantie à en juger par la dérive négative du résultat.

Cette dérive survient en raison d'une sous-estimation dans la tarification du pro-

duit. Cette sous-estimation concerne le risque de longévité des autonomes, le risque d'entrée en dépendance de ces derniers et le risque de longévité des dépendants. Elle a conduit à une sous estimation de la prime, qui a été compensée par une hausse des cotisations, et à un sous-provisionnement, entraînant des pertes puisque l'assureur n'a pas constitué suffisamment de provisions pour faire face au moment où il y aura un nombre insuffisant d'autonomes en comparaison du nombre de dépendants.

Nous avons alors proposé la mise en place d'un fond de garantie, afin de mettre en réserve le résultat positif attendu durant les années à venir et ainsi combler le manque de provisionnement. Cette solution devrait permettre à long terme d'éviter à l'assureur et au réassureur d'éventuelles pertes.

Ouverture

Dans cette étude, pour l'estimation des différentes lois de probabilités, nous n'avons pas fait de distinction entre homme et femme. Or ce facteur est probablement significatif. Nous n'avons pas effectué ce travail par manque de données et par soucis de robustesse des résultats. Néanmoins il serait intéressant d'effectuer cette segmentation dans les années à venir si les données le permettent.

De même, nous n'avons pas étudié les résiliations de contrat ou réductions. Mais le taux de résiliation très faible présent sur ce type de garantie permet de négliger cet effet.

Par ailleurs, nous n'avons pas pris en compte la variation du taux technique. Celui-ci intervient dans le calcul des Provisions Risques Croissants, les faisant diminuer lorsque le taux augmente, et augmenter lorsque le taux baisse. Il serait alors pertinent d'essayer d'anticiper ce taux, par différents modèles. Il en est de même pour le taux de rendement financier, lequel influe de manière importante sur le résultat. Une modélisation plus fine via des modèles de taux serait intéressante.

Partie 10

Table des abbréviations

AGGIR : Autonomie Gérontologie Groupe Iso-Ressource

AIVQ : Activités Instrumentales de la Vie Quotidienne

APA : Allocation Personnalisée d'Autonomie

AVQ : Activités de la Vie Quotidienne

DP : Dépendance Totale

DREES : Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques

DT : Dépendance Totale

EHPA : Etablissement d'Hébergement pour Personnes Agées

EHPAD : Etablissement d'Hébergement pour Personnes Agées Dépendantes

FFSA : Fédération Française des Sociétés d'Assurance

GEMA : Groupement des Entreprises Mutuelles d'Assurance

GIR : Groupe d'Iso-Ressource

GM : Gompertz-Makeham

HID : Handicaps-Incapacité-Dépendances

IC : Intervalle de Confiance

INED : Institut National d'Etudes Démographiques

INSEE : Institut National de la Statistique et des Etudes Economiques

INSERM : Institut National de la Santé Et de la Recherche Médicale

MAPE : Mean Average Percentage Error

PIB : Produit Intérieur Brut

PMR : Provisions Mathématiques de Rente

PRC : Provisions pour Risques Croissants

PSD : Prestation Spécifique Dépendance

SMR : Standardised Mortality Ratio

TVaR : Tail Value at Risk

VaR : Value at Risk

WH : Whittaker Henderson

Partie 11

Bibliographie

Ouvrages

BLONDEAU J., PARTRAT C. [2003] "La réassurance : approche technique", Economica

DELWARDE A., DENUIT M. [2006] Construction de tables de mortalité périodiques et prospectives, Economica

SAPORTA G. [2006] « Probabilités, analyses des données et statistique », Technip, 2ème édition.

Article

AALEN O. [1978] «Non-parametric inference for a family of counting processes». Ann. Stat. 6, 710-726.

Böhmer, P. Theorie der unabhängigen Wahrscheinlichkeiten Rapports. Mémoires et procès verbaux du septième congrès international d'actuaires. Amsterdam, 1912 (2) : p. 327-43

BRASS W. [1971] « On the Scale of Mortality », Biological aspects of demography, W. Brass (ed.), Londres, Taylor and Francis, 69-110.

COALE A., KISKER E.E. [1990]. Modèles d'extrapolation de la mortalité aux grands âges. Institut des Sciences Actuarielles et Institut de Statistique Université Catholique de Louvain, Louvain-la-Neuve, Belgique.

FLEMING T.R., HARRINGTON D.P. [1991] Counting processes and survival analysis, Wiley Series in Probability and Mathematical Statistics. New-York : Wiley.

GEHAN E.A. [1965] « A generalized Wilcoxon test for comparing arbitrarily singly-censored samples ». Biommetrika, 41, 361-372.

KAMEGA A, P PLANCHET F., Mesure du risque d'estimation associé à une table d'expérience, 2010

KAPLAN, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481

LANGMEIER R. [2000], « Etude de différentes méthodes d'ajustement de tables de mortalité : application aux données d'une compagnie d'assurance », HEC Lausanne

MAKEHAM W. [1860] « On the law of mortality. » Journal of the Institute of Actuaries, 13, 325-358.

MATSUMOTO M., and TAKUJI N., "Mersenne twister : a 623-dimensionally equidistributed uniform pseudo-random number generator." ACM Transactions on Modeling and Computer Simulation (TOMACS) 8.1 (1998) : 3-30.

McCULLOUGH, B. D., and DAVID A. HEISER. "On the accuracy of statistical procedures in Microsoft Excel 2007." Computational Statistics Data Analysis 52.10 (2008) : 4570-4578.

METROPOLIS N, ULAM S., « The Monte Carlo Method », Journal of the American Statistical Association, vol. 44, no 247, septembre 1949, p. 335-341

PETAUTON P. [1996]. Théorie et pratique de l'assurance vie. DUNOD.

PETO R., PETO J. [1972] « Asymptotically efficient rank invariant test procedures (with discussion) ». J. R. Stat. Soc. A, 135, 185-207.

TOMAS, J., PLANCHET, F. (2013). Constructing entity specific projected mortality table : adjustment to a reference. Cahiers de recherche de l'ISFA, 1-32.

WEIBULL W. [1951] « A Statistical Distribution Function of Wide applicability ». Journal of Applied Mechanic, 18, p. 292-297.

Cours

DILLINGER Angélique, Cours sur la dépendance, DUAS 2

SALLAH Kankoé, Méthode actuarielle d'estimation des courbes de survie : principe, différences avec la méthode de Kaplan-Meier

EISELE Karl Theodor, Cours d'analyse de survie, DUAS 2

Partie 12

Annexes

Modifié par Ordonnance n2010-462 du 6 mai 2010 - art. 1

Le contrôle de l'Etat s'exerce dans l'intérêt des assurés, souscripteurs et bénéficiaires de contrats d'assurance et de capitalisation. Sont soumises à ce contrôle :

1 les entreprises qui sous forme d'assurance directe contractent des engagements dont l'exécution dépend de la durée de la vie humaine, s'engagent à verser un capital en cas de mariage ou de naissance d'enfants, ou font appel à l'épargne en vue de la capitalisation et contractent à cet effet des engagements déterminés ;

2 les entreprises qui sous forme d'assurance directe couvrent les risques de dommages corporels liés aux accidents et à la maladie ;

3 les entreprises qui sous forme d'assurance directe couvrent d'autres risques y compris ceux liés à une activité d'assistance.

Les mutuelles régies par le code de la mutualité, les institutions régies par le livre IX du code de la sécurité sociale et à l'article L. 727-2 du code rural et de la pêche maritime ne sont pas soumises aux dispositions du présent code.

Sont également soumises au contrôle de l'Etat les entreprises agréées à la date du 1er janvier 1993 qui font appel à l'épargne en vue de la capitalisation sans souscrire d'engagements déterminés.

Shapiro Wilk Le test de Shapiro Wilk, publié en 1965 par Samuel Shapiro et Martin Wilk, teste l'hypothèse H0 selon laquelle $x_1, ..., x_n$ est issu d'une population normalement distribuée.

La statistique de test est :

$$W = \frac{(\sum_{i=1}^{n} a_i \cdot x_{(i)})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

où $x_{(i)}$ est un échantillon ordonné et a_i une constante générée par la moyenne, les variances et les covariances de la statistique d'ordre de l'échantillon de taille n d'une distribution normale.

R^2 Nous considérons un modèle linéaire gaussien

$$Y = \beta_0 + \beta_1 \cdot X +$$

ϵ

avec $\epsilon \simeq$ Normale(0, σ^2)

Le R^2, défini comme le carré du coefficient de corrélation de X et Y, variable explicative et réponse respectivement, est une mesure de qualité de l'ajustement égale au rapport de la variance effectivement expliquée sur la variance à expliquer :

$$R^2 = \frac{var(\hat{Y})}{var(Y)}$$

où $\hat{Y}$ est le vecteur estimé par la régression.

Test de Student Le principe du test de Student est le suivant : nous voulons déterminer si la valeur d'espérance μ d'une population de distribution normale et d'écart type σ non connu est égale à une valeur déterminée μ_0. Pour ce faire, nous tirons de cette population un échantillon de taille n dont nous calculons la moyenne $\overline{x}$ et l'écart-type empirique s.

Selon l'hypothèse nulle, la distribution d'échantillonnage de cette moyenne se distribue elle aussi normalement avec un écart type $\frac{s}{\sqrt{n}}$.

La variable :

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}},$$

suit alors une loi de Student avec n-1 degrés de liberté où :

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Test de Fisher Nous considérons le même modèle linéaire gaussien que précédement.

Nous cherchons à tester l'hypothèse H0 : $\beta_0 = 0$

Pour tester cette hypothèse, nous utilisons la distribution de Fisher. Nous pouvons montrer que, sous l'hypothèse H0, alors la statistique F_0 égale à :

$$F_0 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-2)}$$

(où est la moyenne de la variable réponse.)

suit une loi de Fisher à 1 degré de liberté au numérateur et n-2 au dénominateur.

H0 est rejetée si :

$$F0 > f_{\alpha,1,n-2}$$

où $f_{\alpha,1,n-2}$ est le quantile d'une loi de Fisher correspondant à la probabilité (1-α) avec α le niveau de significativité.

Value at Risk La Value at Risk au niveau de confiance α est donnée comme le plus petit nombre l tel que la probabilité de perte L excède l est au moins égale à (1-α)

$$VaR_\alpha = inf[l \in \mathbb{R} : \mathbb{P}(L > l) \leq 1 - \alpha]$$

Tail Value at Risk La Tail Value at Risk est, de manière pratique, la perte moyenne sachant qu'une perte survenant qu'avec une probabilité α survient. Elle est égale à :

$$TVaR_\alpha = \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(X)d_\gamma$$

Si la distribution de X est continue alors, alors nous pouvons écrire :

$$TVaR_\alpha = \mathbb{E}[-X|X \leq -VaR_\alpha(X)]$$

Algorithme de la simulation De manière plus formelle la simulation effectuer peut se résumer, en notant :

- x l'âge de l'assuré ;
- t l'ancienneté de sa police ;
- q_x^a la probabilité de décès dans l'année d'un assuré autonome d'âge x ;
- $i_x^{par,tot}$ la probabilité de passage en dépendance partielle ou totale d'un assuré autonome d'âge x dans l'année ;
- $q_x^{i,tot}$ la probabilité de décès dans l'année d'un assuré en dépendance totale ;
- $q_x^{i,par}$ la probabilité de décès dans l'année d'un assuré en dépendance partielle ;
- ratioDT : la probabilité de tomber en dépendance totale sachant que l'individu entre en dépendance ;
- ratioMaladie : la probabilité de tomber en dépendance pour cause de maladie lors de la survenue de la dépendance d(un individu autonome (autre que démence) ;
- ratioDemence : la probabilité d'entrer en dépendance pour cause de démence lors de la dépendance d'un individu autonome ;
- $i_x^{DP->DT}$: la probabilité qu'a un individu en dépendance partielle d'âge x de tomber en dépendance totale dans l'année.

Soit $u_1, ..., u_n$ un vecteur aléatoire suivant une loi uniforme entre 0 et 1, nous avons :

SimulationAutonome(x,t)

- Si $q_x^a < u_1$ alors : fin de la simulation
- Sinon
  - Si $i_x^{par,tot} < u_2$ alors SimulationCarence(t)
  - Sinon : x<-x+1, t<-t+1, SimulationAutonome(x,t)

SimulationCarence(t)

- Si $ratioMaladie < u_3$ et $t < 1$ alors : fin de la simulation
- Si $ratioDemence < u_4$ et $t < 3$ alors : fin de la simulation
- Sinon SimulationTypeDependance(x,t)

SimulationTypeDependance(x,t)

- Si $ratioDT < u_5$ alors SimulationDependanceTotale(x,t)
- Sinon SimulationDependancePartielle(x,t)

SimulationDependancePartielle(x,t)

- Si $q_x^{i,par} < u_6$ alors : fin de la simulation
- Sinon x<-x+1,t<-t+1,
  - Si $i_x^{DP->DT} < u_7$ alors SimulationDependanceTotale(x,t)
  - Sinon SimulationDependancePartielle(x,t)

SimulationDependanceTotale(x,t)

- Si $q_x^{i,tot} < u_8$ alors : fin de la simulation
- Sinon x<-x+1,t<-t+1, SimulationDependanceTotale(x,t)