

CONSTRUCTION DE LOIS D'EXPERIENCE EN PRESENCE D'EVENEMENTS CONCURRENTS

APPLICATION A L'ESTIMATION DES LOIS D'INCIDENCE D'UN CONTRAT DEPENDANCE

Quentin GUIBERT¹

Frédéric PLANCHET²

ISFA - Laboratoire SAF

Université de Lyon - Université Claude Bernard Lyon 1

Autorité de Contrôle Prudentiel et de Résolution³

Prim'Act⁴

Résumé:

Cet article propose d'illustrer la mise en œuvre de méthodes d'estimation non paramétriques, introduites dans le cadre de modèles multi-états markoviens avec censure, pour construire des lois d'expérience applicables en présence de plusieurs événements concurrents. Cette situation se présente en pratique en assurance dépendance lorsqu'il est nécessaire de distinguer les lois d'incidence par pathologie. Aussi, plutôt que d'appliquer des techniques utilisées usuellement par les praticiens et consistant à observer marginalement chaque cause d'entrée en dépendance, l'approche décrite permet d'estimer globalement l'ensemble des lois d'entrée par cause et de correctement appréhender l'interdépendance entre chacune d'elle. Ce travail fournit une comparaison des résultats obtenus par ces deux approches au niveau de la provision à constituer afin de pouvoir justifier d'une approche marginale, plus simple à mettre en œuvre en pratique.

Mots-clés : modèle multi-états, risques concurrents, durée de maintien marginale, processus markovien, estimation non paramétrique, assurance dépendance, taux d'incidence.

¹ Quentin Guibert est doctorant à l'ISFA et économiste chercheur au sein de l'ACPR. Contact : quentin.guibert@acpr.banque-france.fr.

² Frédéric Planchet est professeur à l'ISFA et actuair associé chez Prim'Act. Contact : frederic.planchet@univ-lyon1.fr.

³ ACPR-CC 66 2771, 61 Rue Taitbout, 75436 Paris Cedex 09, France

⁴ Prim'Act – 42 avenue de la Grande Armée - 75017 Paris - France

Abstract:

This article presents a non-parametric markov multi-state approach adapted to censored data to construct inception rates with competing risks. This situation arises for LTC insurance when multiple pathologies are likely to cause disability. Instead of using classical techniques applied by practitioners based on latent failure times, our approach allows to estimate globally all inception rates and taking account adequately of dependence between the latent times. This article compares the results obtain on reserves with the both approaches in order to justify if the latent failure times approach is relevant.

Keywords: multi state model, competing risks, latent failure time model, Markov process, non-parametric estimation, Long Term Care insurance, inception rates.

Les points de vue exprimés dans cet article n’engagent que leurs auteurs et n’expriment pas nécessairement la position de l’Autorité de Contrôle Prudenciel et de Résolution, ni de la Banque de France.

1. INTRODUCTION

L’assurance dépendance permet d’offrir une couverture financière en cas de perte d’autonomie ou de mobilité entraînant des difficultés pour accomplir des actes de la vie quotidienne. En France, elle s’adresse aux personnes âgées définies comme étant partiellement ou totalement dépendantes et vient compléter, majoritairement sous forme de rente forfaitaire, les aides publiques. Le lecteur intéressé pourra consulter les travaux de Courbage et Roudaut [2011] pour davantage de précision concernant le dispositif français ou la thèse de Plisson [2009] concernant le développement de ce marché dans l’hexagone.

La description des contrats d’assurance dépendance et l’examen des problématiques de tarification et de provisionnement associées ont fait l’objet de nombreux développements dans la littérature actuarielle (cf. le chapitre 6 de l’ouvrage d’Haberman et Pitacco [1998], la revue de la littérature associée ou encore Deléglise *et al.* [2009]). Ces développements s’appuient généralement sur l’utilisation de modèle multi-états markoviens (cf. le chapitre 5 de Denuit et Robert [2007] ainsi que la revue de la littérature associée), à l’instar des modèles utilisés en assurance incapacité-invalidité. Plus récemment, l’utilisation de modèles semi-markoviens dans ce champ de l’assurance de personnes a été envisagée (cf. le chapitre 6 de Denuit et Robert [2007] ou Christiansen [2012]) et se justifie théoriquement par le fait que la mortalité des assurés dépendants est à la fois fonction de l’âge et de la durée de maintien.

La construction de bases d'expérience adaptées à ces modèles se révèle en assurance dépendance être un exercice particulièrement délicat, du fait d'une part des faibles volumes observés jusqu'à aujourd'hui par les assureurs et d'autre part du caractère évolutif de ce phénomène (cf. par exemple sur ce point Ferri et Olivieri [2000], Olivieri et Pitacco [2001] et Rickayzen et Walsh [2002]). Une difficulté supplémentaire consiste avec l'arrivée des dispositifs Solvabilité 2 et IFRS à concevoir ces bases d'expériences avec un certain degré de réalisme, d'homogénéité et selon les risques spécifiques de l'assureur.

Ces difficultés amènent les assureurs à réfléchir sur les techniques de construction de tables permettant d'y répondre tant en amont qu'en aval de l'entrée en dépendance (cf. Tomas et Planchet [2013] pour la construction de loi de survie en dépendance). Aussi, cet article s'intéresse de manière plus spécifique à la construction de lois d'expérience mesurant la survenance de la dépendance lourde. Ce phénomène résulte de pathologies liées à l'âge (maladies neuro-dégénératives, affections cardio-vasculaires, cancers, etc.) constituant autant de causes concurrentes qu'il peut être nécessaire de distinguer puisque certains contrats appliquent des délais de carence par type de pathologie pour limiter les risques d'anti-sélection. Dans cette mesure, il est légitime d'introduire un modèle à risques concurrents pour évaluer les taux d'incidence vers ces différentes causes de sorties.

À l'instar des travaux de Czado et Rudolph [2002] et de Helms *et al.* [2004], nous choisissons d'aborder cette problématique d'inférence à partir d'un modèle multi-états (cf. Andersen *et al.* [2002]), en proposant des estimateurs tenant compte du phénomène de censure à droite observé sur les données. Toutefois et bien qu'il soit théoriquement plus pertinent, ce cadre est potentiellement plus lourd à mettre en œuvre et plus délicat à appréhender par les praticiens. Aussi, nous examinons dans quelle mesure une approche basée sur l'estimation des durées de maintien marginales, plus couramment utilisées, peut se justifier.

Cet article s'organise de la manière suivante. La section 2 précise les notations du modèle et met en exergue les quantités d'intérêts pour le calcul des provisions techniques. Dans la section 3, les méthodologies d'estimation non paramétrique sont décrites. La première utilise le cadre classique des modèles multi-états tandis que la seconde s'appuie sur les durées de maintien marginales. Une application numérique basée sur un portefeuille de contrats dépendance est ensuite présentée dans la section 4. Elle s'attache en particulier à évaluer le risque d'estimation associé à la mise en œuvre de l'approche multi-états pour construire des lois d'expérience. Finalement, la section 5 conclut l'étude.

2. MODELISATION

Cette section présente le cadre de base permettant de modéliser l'incidence d'un contrat dépendance en présence de risques concurrents.

2.1 Notation

Sur un espace probabilisé (Ω, \mathcal{A}, P) , l'état occupé par l'assuré est représenté à partir d'un processus markovien non-homogène $(X_t)_{t \geq 0}$, muni de sa filtration engendrée $(\mathcal{F}_t)_{t \geq 0}$, et est supposé continu à droite avec limite à gauche. Ce processus prend ses valeurs dans un espace d'états $\mathcal{S} = \{e_0, e_1, \dots, e_n\}$ fini, où e_0 correspond à l'état de « cotisant » ou de « valide » et e_j correspond à la « $j^{\text{ème}}$ cause de sortie » pour $j = 1, \dots, n$. Les transitions possibles entre ces états sont représentées par le schéma de la Figure 1.

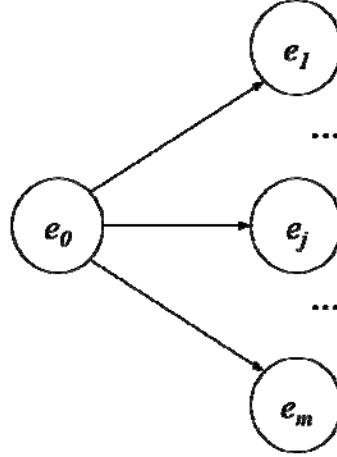


Figure 1 : Modèle multi-états associé à l'incidence en dépendance

Comme il est classique de le faire dans la littérature relative aux modèles multi-états (cf. par exemple le chapitre 2 de Andersen *et al.* [1993]), nous caractérisons le processus $(X_t)_{t \geq 0}$ à partir de ces probabilités de transition

$$p_{0j}(s, t) = P(X_t = j | X_s = e_0) \text{ pour } 0 \leq s \leq t \text{ et } j \in \mathcal{S} ; \quad (1)$$

et de ces intensités de transition telles que pour tout $t \geq 0$

$$\begin{aligned} \mu_{0j}(t) &= \lim_{\Delta t \rightarrow 0} \frac{p_{0j}(t, t + \Delta t) - p_{0j}(t, t)}{\Delta t} \text{ pour } j \neq e_0 ; \\ \mu_{00}(t) &= - \sum_{j \neq e_0} \mu_{0j}(t). \end{aligned}$$

Les fonctions d'intensité cumulée de transition correspondantes sont alors obtenues

pour tout $t \geq 0$ et $j \in \mathcal{S}$ via :

$$\begin{aligned} M_{0j}(t) &= \int_0^t \mu_{0j}(\tau) d\tau \text{ pour } j \neq e_0 ; \\ M_{00}(t) &= -\sum_{j \neq e_0} M_{0j}(t). \end{aligned} \quad (2)$$

D'autre part, en notant T la durée de maintien dans l'état « cotisant » et $V = X_T$ l'état du modèle à cet instant, nous introduisons les fonctions de répartition marginales définies pour tout $t \geq 0$ par

$$F_j(t) = P(T \leq t, V = j), \text{ pour } j \neq e_0 ; \quad (3)$$

auxquelles on associe les durées de maintien marginales $T_{0j} = \inf_{t \geq 0} (X(t) = j)$ pour $j \neq e_0$

qui vérifient

$$T = \min_{j \neq e_0} (T_{0j}).$$

En particulier, on note que les fonctions de répartition marginales s'écrivent pour tout $t \geq 0$

$$F_j(t) = p_{0j}(0, t) = \int_0^t p_{00}(0, \tau) \cdot \mu_{0j}(\tau) d\tau, \text{ pour } j \neq e_0.$$

2.2 Provisions techniques

De manière simplifiée, l'expression des flux futurs actualisés en date $t \geq t_0$ associés à un assuré cotisant à l'âge x en t_0 s'écrit (cf. par exemple le chapitre 1 d'Haberman et Pitacco [1998] ou Christiansen [2012])

$$\begin{aligned} \Lambda_x(t) &= \sum_{j \neq e_0} \int_{t-t_0}^{+\infty} \frac{\delta(t_0 + \tau)}{\delta(t)} \cdot 1_{\{X_{(x+\tau)} = 0, X_{x+\tau} = j\}} \cdot c_j(x + \tau) d\tau \\ &\quad - \int_{t-t_0}^{+\infty} \frac{\delta(t_0 + \tau)}{\delta(t)} \cdot 1_{\{X_{x+\tau} = 0\}} \cdot b(x + \tau) d\tau \end{aligned} ; \quad (4)$$

où $c_j(x)$ représente la prestations versées à l'âge x en cas de survénance de la j^{me} cause de sortie et $b(x)$ correspond à la prime versée tant que l'assuré demeure dans l'état « cotisant ». Le coefficient d'actualisation et le prix d'un zéro-coupon s'écrivent de manière classique à partir du taux sans risque $(r_t)_{t \geq 0}$ tel que pour tout $0 \leq t \leq T$

$$\delta(t) = \exp\left(-\int_0^t r_\tau d\tau\right) \text{ et } B(t, T) = \mathbb{E}_t \left[\frac{\delta(T)}{\delta(t)} \right].$$

Dans la suite, nous nous intéresserons à la provision *best estimate* associée à un contrat s'écrivant en considérant l'espérance de (4), en supposant que des flux de primes et de prestations sont indépendants du risque de taux

$$\begin{aligned} \mathbb{E}[\Lambda_x(t) | X_x = e_0] &= \sum_{j \neq e_0} \int_{t-t_0}^{+\infty} B(t, t_0 + \tau) \cdot p_{00}(x, x + \tau) \cdot \mu(x + \tau) \cdot c_j(x + \tau) d\tau \\ &\quad - \int_{t-t_0}^{+\infty} B(t, t_0 + \tau) \cdot p_{00}(x, x + \tau) \cdot b(x + \tau) d\tau \end{aligned} \quad (5)$$

L'équation (5) peut s'écrire en temps discret sous la forme suivante pour $t \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}[\Lambda_x(t) | X_x = e_0] &\approx \sum_{j \neq e_0} \sum_{k=t-t_0}^{+\infty} B(t, t_0 + k + 1) \cdot p_{00}(x, x + k) \cdot q_j(x + k) \cdot c_j(x + k) \\ &\quad - \sum_{k=t-t_0}^{+\infty} B(t, t_0 + k) \cdot p_{00}(x, x + k) \cdot b(x + k) \end{aligned} \quad (6)$$

où $q_j(x) = p_{0j}(x, x + 1) = P(T \leq x + 1, V = j | T > x)$.

Aussi, l'évaluation de la provision *best estimate* requiert de réaliser l'estimation des lois de survie $p_{00}(x, x + 1)$ et d'incidence $q_j(x)$ pour tout $j \in \mathcal{S}$.

3. ESTIMATION DES TAUX D'INCIDENCE BRUTS

Dans l'optique de construire des lois de transition réalistes, le choix des techniques d'inférence du modèle est une question primordiale. En présence de risques concurrents, une littérature abondante traite des problématiques d'inférence, parmi laquelle se distingue notamment des approches basées sur les durées de maintien marginales (cf. par exemple Prentice *et al.* [1978]), sur un modèle à hasard proportionnel comme le modèle classique de Fine et Gray [1999] ou encore sur des approches multi-états (cf. Andersen *et al.* [2002]).

En assurance de personnes, il est relativement courant que les praticiens utilisent des techniques basées sur les durées marginales, obtenues à partir de l'estimateur de Kaplan et Meier [1958] ou par positionnement par rapport à une référence externe, et une hypothèse d'indépendance entre celles-ci pour estimer les taux d'incidence bruts. Or ces choix peuvent apparaître critiquables puisqu'ils ne permettent pas de correctement appréhender la dépendance entre les durées de maintien marginales (cf. Tsiatis [1975], Prentice *et al.* [1978] ou encore la section 3.3 de Beyersmann *et al.* [2011] pour un développement sur ce point). Une réflexion similaire sur la pertinence des estimateurs utilisés par les praticiens est développée par Gooley *et al.* [1999] dans le domaine des sciences médicales.

Cette section se concentre sur une méthodologie d'estimation basée sur l'utilisation de modèles multi-états et reprend les éléments essentiels de la méthodologie basée sur les durées de maintien marginales. Les deux approches ainsi proposées tiennent compte du fait que les observations de contrats dépendance font l'objet d'une censure à droite et d'une troncature à gauche, à l'inverse du classique modèle de régression poissonien très utilisé

des actuaires (cf. CMIR12 [1991] pour un cas d'utilisation classique de ce modèle ou la section 2.6 de Denuit et Robert [2007] pour une application à un modèle à risques concurrents).

Dans la suite, nous considérons un échantillon $(X_t^m)_{t \geq 0, m=1, \dots, M}$ de M individus indépendants. On considère également $(C^m)_{m=1, \dots, M}$ les dates de censure, supposées indépendantes de l'état du modèle et non informatives, associées à chaque individu ainsi que les couples $(\bar{T}^m, V^m)_{m=1, \dots, M}$ tels que $\bar{T}^m = T^m \wedge C^m$ et $V^m = X_{\bar{T}^m}^m$.

3.1 Estimation multi-états

Dans le cadre markovien, la théorie des processus de comptage et celle des martingales permettent d'introduire l'estimateur de Nelson-Aalen des intensités de transition cumulées (2) et l'estimateur d'Aalen-Johansen des probabilités de transition (1). Pour une présentation théorique de ces estimateurs et de leurs principales propriétés, le lecteur intéressé pourra se reporter aux sections IV 1 et IV 4 d'Andersen *et al.* [1993] ou à Andersen et Keiding [2002].

Ainsi, en notant $\bar{T}^{(m)}$ la statistique d'ordre m de $(\bar{T}^m)_{m=1, \dots, M}$, l'estimateur des intensités de transition s'écrit

$$\hat{M}_{0j}(t) = \sum_{\{m, \bar{T}^{(m)} \leq t\}} \frac{I_{\{V^{(m)}=j\}}}{Y_0(\bar{T}^{(m)})} \text{ pour tout } j \neq e_0 \text{ et pour } t \geq 0 ; \quad (7)$$

où $Y_j(t)$ correspond au nombre d'individus non censurés présents dans l'état $j \neq e_0$ en date $t \geq 0$.

De cette manière, l'estimateur des taux d'incidence bruts s'écrit à partir de l'équation (7). En adaptant l'expression de l'estimateur d'Aalen-Johansen (cf. section IV.4.1.7 d'Andersen *et al.* [1993]) au cas particulier d'un modèle à risque concurrents, il est possible d'estimer les taux d'incidence par

$$\hat{q}_j(t) = \sum_{\{m, t < \bar{T}^{(m)} \leq t+1\}} \hat{S}(\bar{T}^{(m)} -) \cdot \frac{I_{\{V^{(m)}=j\}}}{Y_0(\bar{T}^{(m)})} \text{ pour tout } j \in \mathcal{S} \text{ et pour } t \geq 0 ; \quad (8)$$

où $\hat{S}(t)$ correspond à l'estimateur de Kaplan-Meier de la fonction de survie de T . À ce titre, il convient de noter que les taux de maintien $p_{00}(x, x+1)$ dans l'état sont naturellement estimés à partir de l'estimateur de Kaplan-Meier.

Les propriétés asymptotiques de l'estimateur (8) permettent de construire des

intervalles de confiance asymptotiques. De plus, il présente l'avantage de permettre une prise en compte de covariables - le sexe serait par exemple un critère de segmentation pertinent selon Gauzère *et al.* [1999] - en faisant l'hypothèse d'un modèle à hasard proportionnel (cf. section VII.2 d'Andersen *et al.* [1993] pour davantage de précisions).

3.2 Estimation par une méthode marginale

La principale problématique relative à l'analyse des durées de maintien marginales provient du fait que la structure de dépendance reliant celles-ci n'est pas identifiable et ne peut être estimée. En effet, les durées marginales sont artificielles et ne peuvent par conséquent être directement observées.

Ainsi, les praticiens font usuellement l'hypothèse que les durées de maintien marginales sont indépendantes, ce qui permet d'écrire tout $t \geq 0$ et $j \neq e_0$

$$P(T_{0_j} > t) = \exp\left(-\int_0^t \mu_{0_j}(\tau) d\tau\right). \quad (9)$$

L'expression (9) permet de vérifier que $P(T > t) = \prod_{j \neq e_0} P(T_{0_j} > t)$ ainsi que

l'équation (3). Il convient par ailleurs de préciser que l'hypothèse d'indépendance ne peut être testée. Au surplus, le lecteur intéressé par l'analyse plus précise de la dépendance entre les durées de maintien marginale pourra se reporter à Carrière [1994] et Carrière [1995].

En segmentant l'échantillon observé par cause de sortie, la quantité (9) est alors estimée à partir de l'estimateur de Kaplan-Meier soit

$$\hat{q}_j^*(t) = \hat{P}(T_{0_j} \leq t+1 | T_{0_j} > t). \quad (10)$$

Toutefois, lorsque les volumes de données à dispositif ne permettent d'offrir une robustesse suffisante à cette estimation, les assureurs peuvent recourir à une approche paramétrique ou à un positionnement par rapport à une référence externe (cf. Deléglise *et al.* [2009]).

L'hypothèse d'indépendance n'étant pas valable, les taux d'incidence $q_j(t)$ sont estimés pour tout $t \geq 0$ et $j \neq e_0$ par le praticien en introduisant des priorités entre les différents estimateurs de $q_j^*(t)$ ce qui se traduit, pour un ordre de priorité fixé $(j_1, \dots, j_n) \in \mathbb{N}^n$, par la formulation

$$\left\{ \begin{array}{l} \tilde{q}_{j_1}(t) = \hat{q}_{j_1}^*(t) \\ \tilde{q}_{j_2}(t) = \hat{q}_{j_2}^*(t) \cdot (1 - \hat{q}_{j_1}^*(t)) \\ \dots \\ \tilde{q}_{j_n}(t) = \hat{q}_{j_n}^*(t) \cdot \prod_{k=1}^{n-1} (1 - \hat{q}_{j_k}^*(t)) \end{array} \right. \quad (11)$$

L'expression (11) réintroduit de manière arbitraire un lien de dépendance entre les lois de durée de maintien marginales et permet de vérifier pour tout $t \geq 0$ la relation

$$1 - \sum_{k=1}^n \tilde{q}_{j_k}(t) = \prod_{k=1}^n (1 - \hat{q}_{j_k}^*(t)) = \hat{p}_{00}(t, t+1). \quad (12)$$

Ainsi, la loi de maintien dans l'état de « cotisant » est indifférente au choix du vecteur de priorité. De surcroît, l'estimateur des taux d'incidence \tilde{q}_j pour $j \neq e_0$ est compris, quel que soit l'ordre de priorité retenu, entre les deux bornes b_j^- et b_j^+ définies telles que pour tout $t \geq 0$

$$b_j^-(t) = \hat{q}_j^*(t) \cdot \prod_{\substack{k \neq j \\ k \neq e_0}} (1 - \hat{q}_k^*(t)) \leq \tilde{q}_j(t) \leq \hat{q}_j^*(t) = b_j^+(t) \quad (13)$$

En outre, la borne b_j^+ est parfois appelée probabilité de transition partielle (cf. exemple IV.4.1 d'Andersen *et al.* [1993]) et correspondrait au taux d'incidence estimé si une seule cause de sortie était considérée.

4. APPLICATION NUMERIQUE : CONSTRUCTION DES LOIS D'INCIDENCE EN DEPENDANCE

Cette application numérique se propose d'illustrer les résultats obtenus en utilisant la méthodologie d'estimation multi-états décrite dans la section 3.1 sur la population d'un contrat d'assurance dépendance. Après avoir lissé ces résultats à partir du modèle de Whittaker-Henderson, nous les comparons à ceux obtenus en appliquant l'approche marginale décrite dans la section 3.2 et examinons les écarts entre les deux méthodes en termes de provisions techniques. L'évaluation réalisée est effectuée à partir du logiciel R (cf. R Development Core Team [2013]) et plus particulièrement du package *mstate* développé par De Wreede *et al.* [2011]. Les scripts utilisés sont disponibles sur demande.

Cette application n'aborde pas davantage la problématique du choix des méthodes de lissage susceptibles d'être retenues. Toutefois, le lecteur intéressé par ce point pourra se reporter à la récente thèse de Tomas [2013] et aux références bibliographiques reprises par l'auteur, à l'ouvrage classique de Loader [1999] ou l'article de Müller et Wang [1994]

concernant le lissage de l'estimateur de Nelson-Aalen par une méthode de noyau.

4.1 Données

Les données utilisées proviennent de l'observation de données individuelles d'un contrat d'assurance dépendance français pendant environ 10 ans sur la plage d'âges 65 – 90 ans. Ces observations comprennent la présence d'un phénomène de censure à droite, indépendant des dates de changement d'états, et sont tronquées à gauche. En identifiant 4 groupes de pathologies au sein de cette population, un modèle à risques concurrents comprenant 6 causes de sortie est considéré. Celles-ci sont reprises dans le tableau 1.

	Causes de sortie
1	Entrée en dépendance pour pathologie neurologique
2	Entrée en dépendance pour pathologie diverse
3	Entrée en dépendance pour cancer en phase terminale
4	Entrée en dépendance pour démence
5	Décès sans dépendance
6	Résiliation

Tableau 1 : Causes de sortie de l'état « cotisant » d'un contrat dépendance

4.2 Estimation et lissage des taux d'incidence par le modèle multi états

La Figure 2 présente les taux d'incidence bruts obtenus en appliquant l'estimateur de Nelson-Aalen (7) puis celui d'Aalen-Johansen (8) aux données. Ces estimations sont encadrées d'un intervalle de confiance ponctuel de niveau $\alpha = 95\%$ calculé en considérant l'approximation suivante (cf. la formule (4.1.22) d'Andersen *et al.* [1993])

$$\hat{q}_j(x) \cdot \exp\left(\pm \Phi^{-1}\left(\frac{\alpha}{2}\right) \cdot \frac{\hat{\sigma}_j(x)}{\hat{q}_j(x)}\right) \text{ pour l'âge } x \in \llbracket 65; 90 \rrbracket \text{ et pour tout } j \neq e_0 ;$$

où $\Phi^{-1}(u)$ correspond au quantile d'ordre u d'une loi normale centrée réduite et $\hat{\sigma}_j^2(x)$ (cf. la formule (4.4.16) d'Andersen *et al.* [1993]) correspond à l'estimateur de la variance de $\hat{q}_j(x)$.

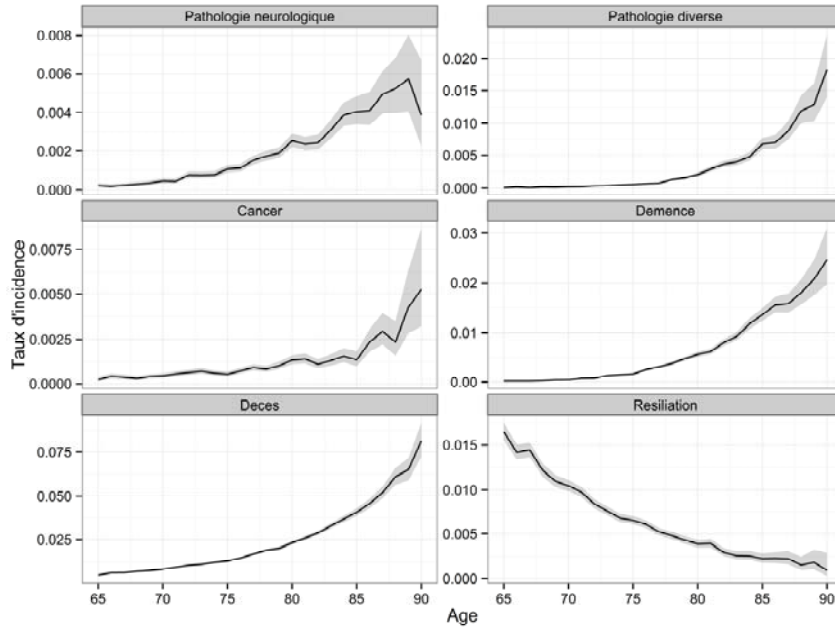


Figure 2 : Estimation des taux d'incidence bruts et intervalle de confiance à 95 %

L'examen des taux bruts ainsi estimés laisse apparaître un niveau d'incertitude plus importante au-delà de 85 ans. Sans surprise, les probabilités d'entrée en dépendance sont croissantes avec l'âge et demeurent relativement faibles en dehors des causes 2 et 4. À l'inverse, les taux de résiliation sont décroissants avec l'âge.

Un lissage de ces taux bruts est appliqué à partir du modèle de Whittaker-Henderson (cf. la section 3.2 du chapitre 5 de Planchet et Thérond [2006]) dont les paramètres (h, z) sont obtenus en minimisant la combinaison linéaire

$$F + h \cdot S ;$$

où F et S correspondent respectivement à des critères de fidélité et de régularité. Ils sont définis tels que

$$F = \sum_{k=1}^l w_k \cdot (\hat{q}_j(x_k) - \tilde{q}_j(x_k))^2 \text{ et } S = \sum_{k=1}^{l-z} (\Delta^z \hat{q}_j(x_k))^2 ;$$

où l correspond à largeur de l'intervalle $[[65;90]]$ et w_k le poids de la $k^{\text{ème}}$ observations.

Les paramètres de lissage optimaux sont obtenus en analysant pour un panel de paramètres donné l'erreur de prédiction relative à chaque jeu de paramètres ainsi qu'en effectuant un diagnostic graphique de la qualité de l'ajustement et des résidus. À ce titre et en fonction des volumes de données sous-jacents, nous utilisons l'information produite en

termes d'erreur de prédiction par les critères suivants pour $x \in \llbracket 65; 90 \rrbracket$ et pour tout $j \neq e_0$:

- le critère de validation croisée, $\frac{1}{l} \cdot \sum_{k=1}^l \left(\frac{\hat{q}_j(x_k) - \tilde{q}_j(x_k)}{1 - \text{infl}_j(x_k)} \right)^2$;
- le critère de validation croisée généralisé, $\frac{l}{(l - p_j)^2} \cdot \sum_{k=1}^l (\hat{q}_j(x_k) - \tilde{q}_j(x_k))^2$;
- l'AIC, $AIC_j = l \cdot \ln \left(\sum_{k=1}^l (\hat{q}_j(x_k) - \tilde{q}_j(x_k))^2 \right) + 2 \cdot p_j$;
- l'AIC corrigé $AIC_j + \frac{2 \cdot p_j \cdot (p_j + 1)}{l - p_j - 1}$;

où p_j correspond au nombre de degrés de liberté du lissage non paramétrique de la $j^{\text{ème}}$ transition, $\text{infl}_j(x)$ correspond à l'influence de l'âge x dans le lissage et $\tilde{q}_j(x)$ correspond au taux d'incidence lissé. Pour d'avantage de précisions concernant les notions de degré de liberté et d'influence dans un cadre de régression non paramétrique, le lecteur intéressé pourra se reporter à Tomas [2011].

Les paramètres de lissage retenus ainsi que le nombre de degrés de liberté sont présentés dans le Tableau 2.

Causes de sortie	Degré de liberté	h	z
Pathologie neurologique	6,59	4	2
Pathologie diverse	12,09	7	10
Cancer	8,21	6	4
Démence	9,98	4	6
Décès	9,98	4	6
Résiliation	5,79	7	2

Tableau 2 : Paramètres de lissage par le modèle Whittaker-Henderson

La Figure 3 présente les taux d'incidence lissés par Whittaker-Henderson ainsi que les résidus et les résidus de Student permettant de détecter les taux bruts extrêmes.

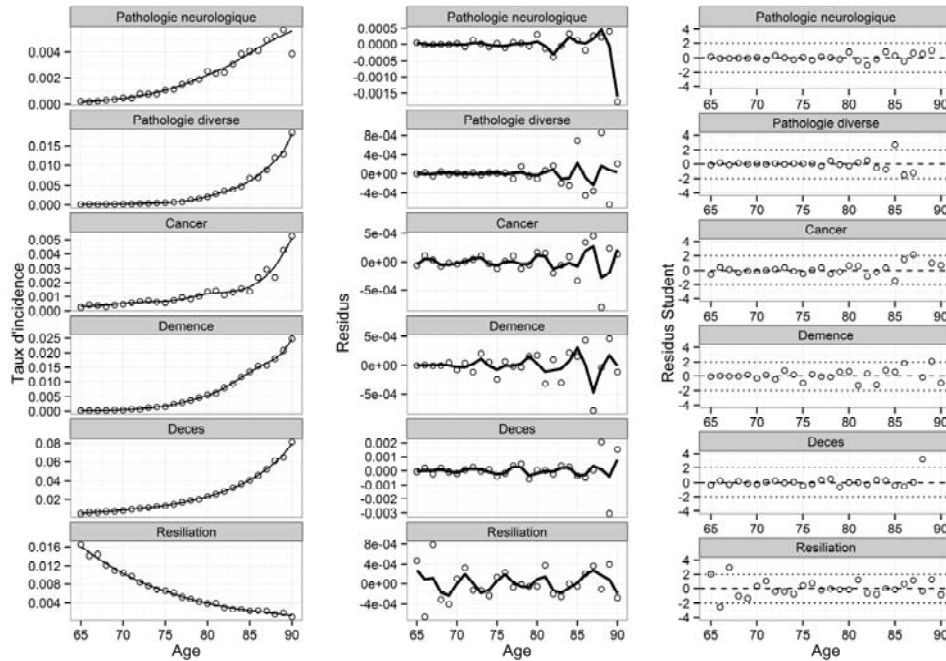


Figure 3 - Taux d'incidence lissés, résidus et résidus de Student du lissage

Les taux d'incidence lissés présentent des tendances d'évolution relativement régulières. En dehors d'un point visiblement aberrant à 90 ans concernant les pathologies neurologiques et dont le poids a été atténué lors du lissage, l'analyse des résidus et des résidus de Student permet de faire apparaître les classes d'âges pour lesquelles une force de lissage plus ou moins grande est appliquée. À l'examen de ces résultats, il apparaît que les taux d'incidence en dépendance sont davantage lissés au-delà de 85 ans, se traduisant par un risque d'estimation plus important pour ces âges. Les taux de décès présentent quant à eux un ajustement de bonne qualité, les âges avec la plus grande incertitude étant ceux au-delà de 87 ans. Par ailleurs concernant les taux de résiliation, une force de lissage plus importante est appliquée avant 70 ans et dans une moindre mesure au-delà de 85 ans par rapport aux autres âges.

Un ensemble des tests classiques est finalement effectué afin de valider la qualité du lissage appliqué. Le Tableau 3 fournit une présentation des résultats de ces tests dont l'examen permet de conclure d'une qualité de lissage satisfaisante.

		Pathologie neurologique	Pathologie diverse	Cancers	Démence	Décès	Résiliation
Résidus de Student	>2	1	4	2	2	3	3
	>3	1	3	1	1	3	0
Kolmogorov-Smirnov	Test	0,08	0,08	0,08	0,08	0,04	0,08
	p-valeur	1	1	1	1	1	1
χ^2	Test	19,41	13,23	25,16	17,21	10,07	19,98
	p-valeur	0,61	0,58	0,91	0,69	0,18	0,59
R^2	Valeur	0,95	1	0,97	1	1	1

Tableau 3 : Évaluation de la qualité du lissage

4.3 Mesure du risque d'estimation

Ce paragraphe décrit les résultats obtenus par la mise en œuvre d'une approche *bootstrap* non paramétrique (cf. Efron [1979]) sur les observations pour mesurer les risques d'estimation. Ces risques apparaissent lors de l'estimation des taux d'incidence bruts et lors du lissage de ces mêmes taux. Pour chaque ré-échantillonnage, l'estimateur d'Aalen-Johansen (8) est réévalué et la procédure de lissage décrite dans le paragraphe 4.2 est réappliquée. Pour cette application, $K = 1\,000$ itérations sont effectuées.

La Figure 4 fournit un positionnement de la moyenne des taux d'incidence simulés par rapport à ceux estimés au paragraphe 4.2. Elle représente également les intervalles de confiance simulés (quantile 5 % et 95 %) afin d'apprécier la robustesse de l'estimation.

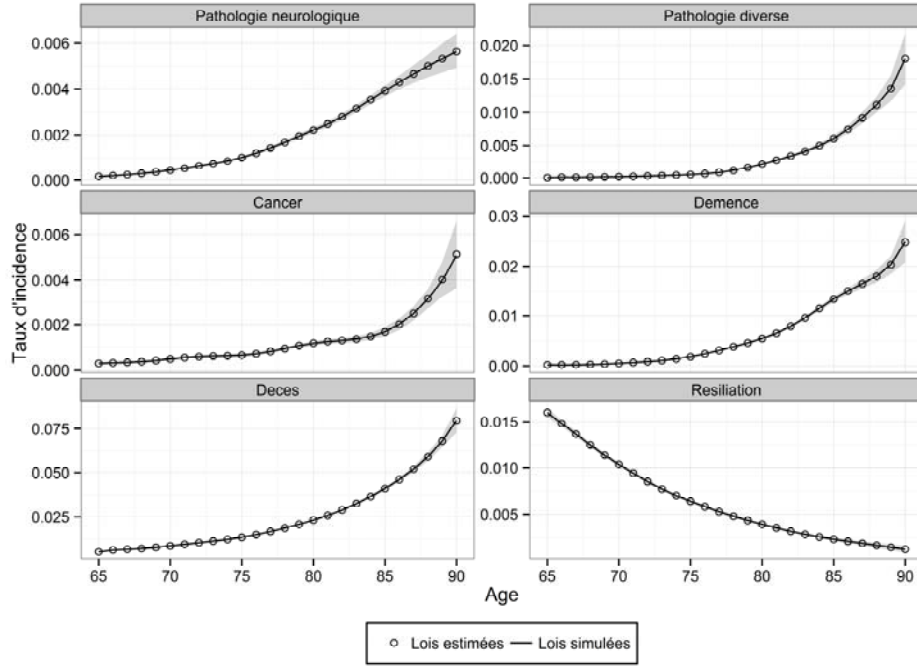


Figure 4 : Taux d'incidence simulés et intervalle de confiance à 95 %

Il ressort de l'analyse que la dispersion des taux d'incidence en dépendance excède assez largement celle des taux de décès ou de résiliation. L'incertitude sur les taux d'incidence en dépendance apparaît significative au-delà de 85 ans.

Une manière de quantifier le risque d'estimation associé à chaque âge consiste à introduire le coefficient de dispersion

$$c(\Psi_{jx}) = \frac{\sqrt{\sum_{k=1}^K (\tilde{q}_j^k(x) - \tilde{q}_j(x))^2}}{\tilde{q}_j(x)} \text{ pour tout } x \in \llbracket 65; 90 \rrbracket \text{ et pour tout } j \neq e_0 ;$$

où $\tilde{q}_j^k(x)$ correspond au taux d'incidence issu du $k^{\text{ème}}$ ré-échantillonnage.

La Figure 5 représente les coefficients de dispersion mesurant le risque d'estimation de chaque loi d'incidence. Elle permet de détecter la présence d'un risque d'estimation significatif entre 65 et 70 ans qui n'était pas visible sur la Figure 4.

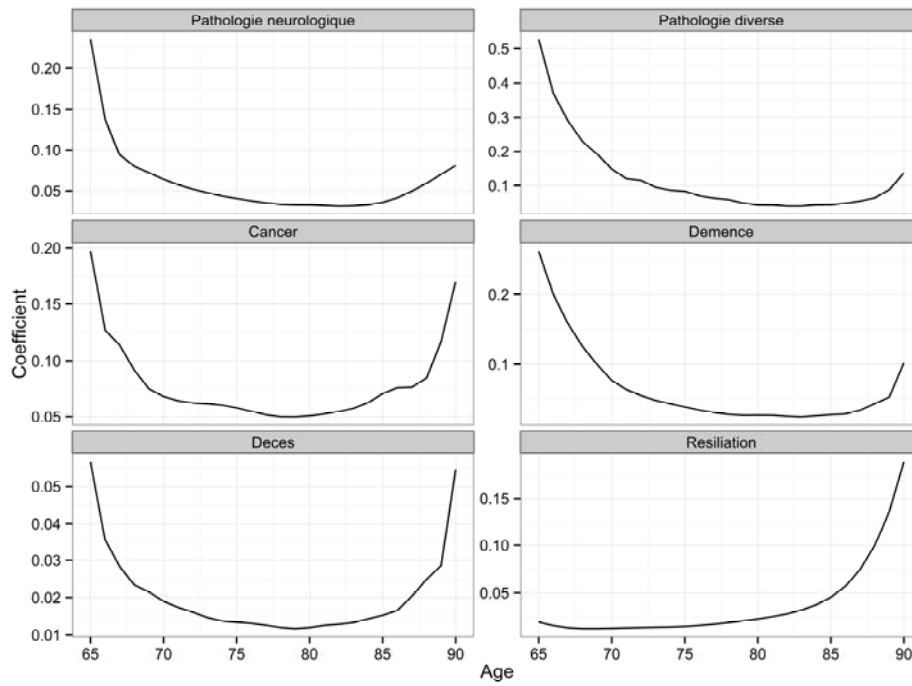


Figure 5 : Risque d'estimation sur les taux d'incidence

La moyenne des coefficients de dispersion calculés est par ailleurs récapitulée dans le Tableau 4. Il en ressort qu'un risque d'estimation non négligeable est constaté sur les taux d'incidence en dépendance, notamment concernant les pathologies diverses.

Causes de sortie	Coefficient de dispersion moyen
Pathologie neurologique	6,01 %
Pathologie diverse	12,12 %
Cancer	7,94 %
Démence	6,53 %
Décès	2,05 %
Résiliation	3,76 %

Tableau 4 : Risque d'estimation moyen sur les taux d'incidence

Enfin, le risque d'estimation est étudié en mesurant son impact sur le montant des provisions techniques. Pour cette évaluation, des hypothèses volontairement simplistes sont

retenues afin de donner une meilleure appréciation des résultats au lecteur. À cet effet, nous retenons un contrat viager réduit prévoyant le versement d'un capital constant d'une unité monétaire en cas d'entrée en dépendance et utilisons un taux d'actualisation constant de 4,00 %. Le montant de provision technique est alors évalué à partir de l'équation (6) et la distribution de $\Lambda_x(t_0)$ est obtenue en considérant le seul risque d'estimation. Incidemment, une approche prenant en compte le risque mutualisable relatif aux fluctuations d'échantillonnage pourrait aisément être intégré à l'approche *bootstrap* mise en œuvre ici en considérant la méthodologie décrite dans Guibert *et al.* [2010] et appliquée par exemple par Planchet et Tomas [2013].

La Figure 6 présente l'estimateur du noyau de la densité $\Lambda_x(t_0)$ pour différents âges. Au regard du faible nombre de tirages effectués, il convient toutefois de préciser que l'analyse des queues de distributions devra être effectuée avec prudence.

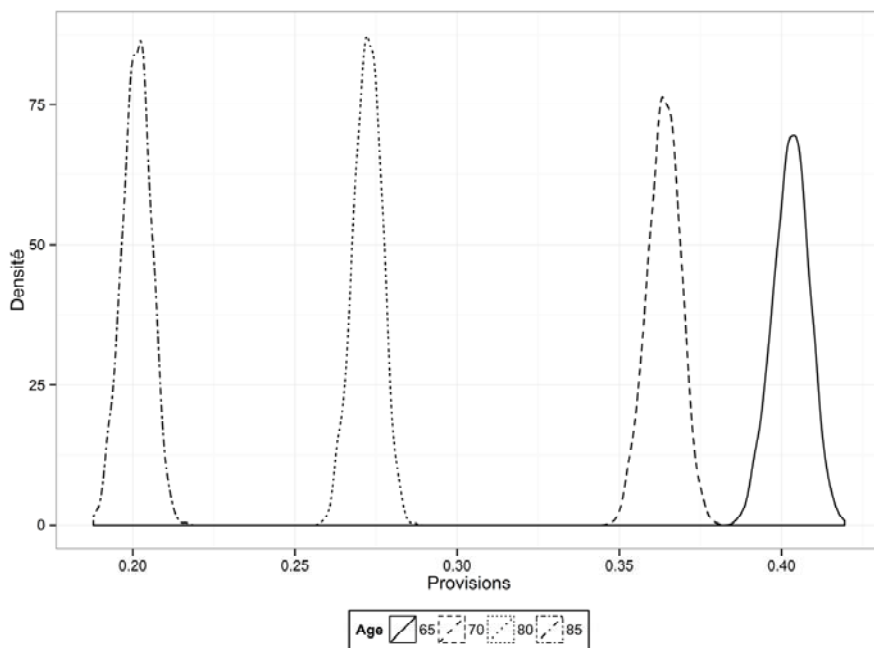


Figure 6 : Distribution des provisions techniques en fonction de l'âge de l'assuré cotisant

À la lumière de cette analyse, le risque d'estimation apparaît relativement important et devra être pris en compte avec soin dans l'évaluation des provisions techniques d'une part et dans le cadre de la gestion des risques d'autre part.

4.4 Comparaison de l'approche multi-états et de la méthode marginale

Finalement, une comparaison entre les deux approches décrites dans la section 3 pour estimer les taux d'incidence bruts est présentée. L'objectif est ici de mesurer le biais introduit par l'utilisation d'une méthode marginale par rapport à l'utilisation d'un modèle multi-états, théoriquement mieux adapté mais pouvant être plus délicat à manipuler. Au préalable, les taux d'incidence (10) sont estimés à partir de l'estimateur de Kaplan-Meier puis lissés sur la base du modèle de Whittaker-Henderson en appliquant une procédure similaire à celle décrite dans le paragraphe 4.2.

Les bornes b_j^- et b_j^+ , définies pour $j \neq e_0$ par l'équation (13) puis lissées, permettant de matérialiser la plage à l'intérieur de laquelle évoluent les taux d'incidence bruts estimés *via* la méthode marginale. La Figure 7 positionne les taux d'incidence bruts et lissés sur la base de l'approche multi-états par rapport à cette plage.

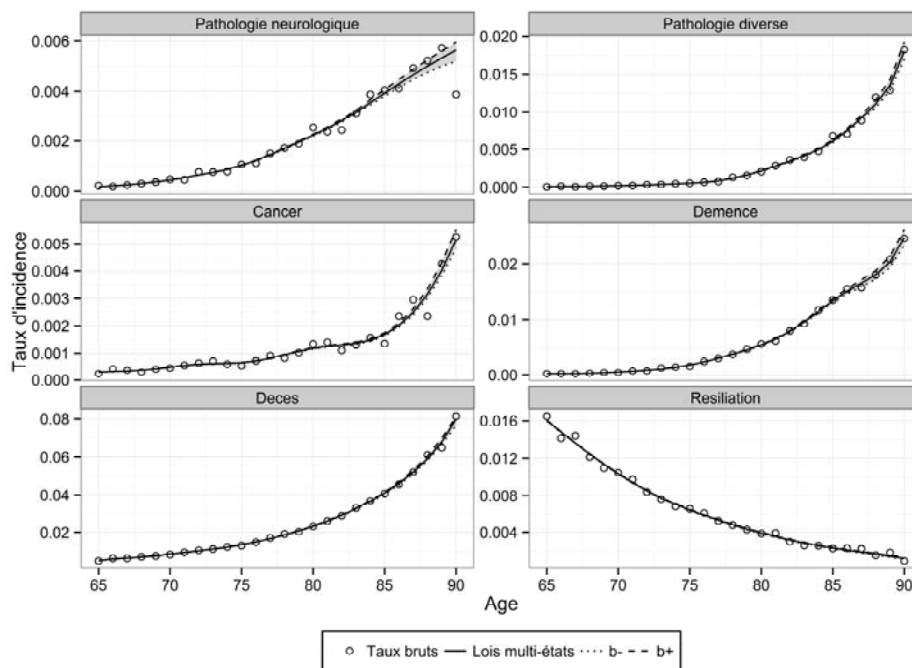


Figure 7 : Positionnement des taux d'incidence multi-états par rapport aux bornes b_j^- et b_j^+

Il apparaît que les taux d'incidence estimés par l'approche multi-états se positionnent approximativement au centre de la plage délimitée par les bornes b_j^- et b_j^+ . Ces bornes sont relativement resserrées en ce qui concerne les taux de décès et de résiliation signifiant que l'approche marginale introduit quelle que soit la convention retenue dans l'équation (11) un écart relativement modeste pour ces transitions. L'écartement de ces bornes est en revanche plus important pour l'incidence en dépendance, notamment en ce qui concerne les pathologies neurologiques.

Afin de mesurer de manière plus précise le biais potentiellement introduit par l'approche marginale, une évaluation des provisions techniques, sur les mêmes hypothèses que celles précédemment introduites, est effectuée pour différents vecteurs de priorité.

Le Tableau 5 fait état de quatre situations testées représentées ensuite sur la Figure 8. Celle-ci permet de positionner le montant de provisions pour ces différentes situations par rapport à celle évaluée à partir de l'approche multi-états.

Situation	Pathologies neurologiques	Pathologies diverses	Cancers	Démence	Décès	Résiliation
1	6	4	5	3	1	2
2	3	2	5	4	1	6
3	1	2	6	5	3	4
4	1	3	2	4	5	6

Tableau 5 : Définition arbitraire de plusieurs ordres de priorité testés

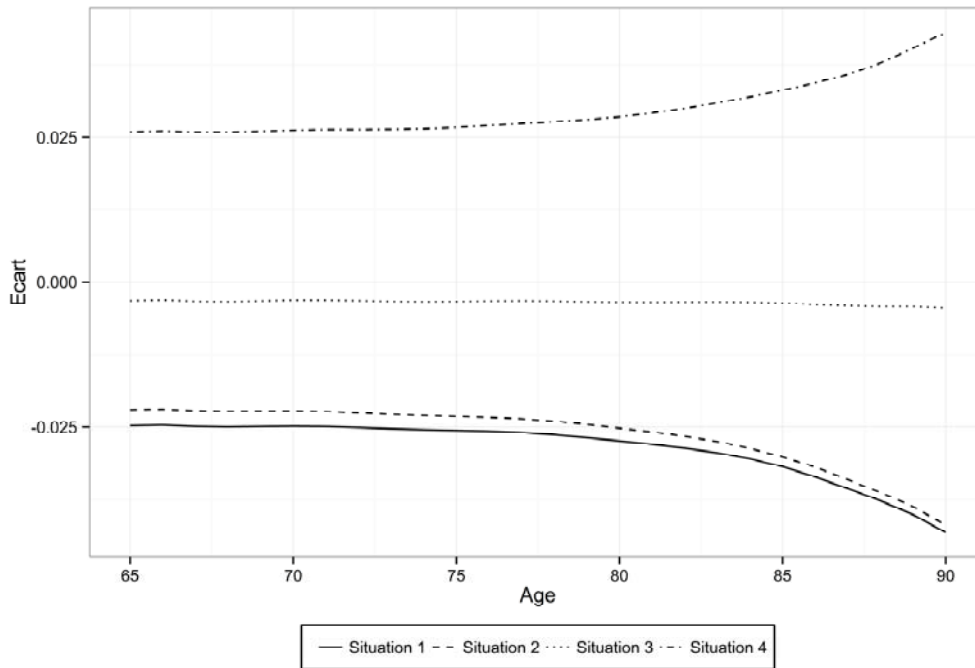


Figure 8 : *Ecart*s relatifs sur les provisions entre les méthodes d'estimation marginale et multi-états

Il ressort que l'ordre des priorités a un impact relativement limité sur le montant des provisions techniques, l'écart observé pouvant être positif ou négatif selon les cas. À 65 ans, il apparaît sur les différentes situations étudiées que l'écart demeure contenu à environ $\pm 2,5\%$, puis à tendance à augmenter avec l'âge.

Par ailleurs, il convient de noter que l'écart semble plus faible lorsque les priorités entre les différentes durées de maintien marginales sont établies de sorte que le décès et la résiliation soient placés « au centre » du vecteur de priorité. Dans la mesure où l'écart reste limité et en répartissant judicieusement l'ordre des priorités entre les différentes durées de maintien, il semble possible pour cette application de retenir une approche marginale pour la construction de lois d'expérience.

5. CONCLUSION

Cette étude présente un cadre pour l'estimation de lois d'incidence en dépendance s'appuyant sur les estimateurs non paramétriques introduits dans le cadre de modèles multi-états avec censure. Cette approche n'étant pas celle habituellement privilégiée par les

actuaire en présence d'évènements concurrents, nous mesurons le risque d'estimation qu'elle engendre et comparons les résultats obtenus à ceux provenant de l'utilisation d'une méthodologie d'estimation simplifiée.

Cette méthodologie, à l'inverse de celle construite à partir de durées de maintien marginales, présente l'avantage d'offrir un cadre plus cohérent en évacuant le problème artificiellement posé par la structure de dépendance reliant chaque loi marginale. Toutefois, si l'utilisation des modèles associés aux causes de sortie de l'état « cotisant » est traitée de manière appropriée en termes d'ordres de priorité, il apparaît en pratique que les lois construites par l'une ou l'autre des deux approches peuvent conduire à observer des résultats relativement comparables en termes de provisionnement sur une garantie dépendance.

Au final, dans le contexte bien précis présenté ici, on peut donc retenir que l'utilisation d'une approche marginale semble possible pour la construction de table d'expérience, ce qui semble particulièrement utile dans la mesure où une telle approche est naturellement privilégiée par le praticien souhaitant répondre aux difficultés posées par de faibles volumes de données à partir de références externes. Une telle approche peut toutefois conduire à des biais potentiels non négligeables et croissants avec l'âge, selon le traitement retenu pour les différentes causes de sortie.

6. BIBLIOGRAPHIE

ANDERSEN P. K., ABILDSTROM S. Z., ROSTHØJ S. [2002] « Competing risks as a multi-state model », *Statistical Methods in Medical Research*, Vol. 11, No. 2, p. 203-215.

ANDERSEN P. K., BORGAN Ø., GILL R. D., KEIDING N. [1993] *Statistical Models Based on Counting Processes*, Springer-Verlag New York Inc., 767 p., (Springer Series in Statistics), ISBN : 0-378-97872-0.

ANDERSEN P. K., KEIDING N. [2002] « Multi-state models for event history analysis », *Statistical Methods in Medical Research*, Vol. 11, No. 2, p. 91-115.

BEYERSMANN J., SCHUMACHER M., ALLIGNOL A. [2011] *Competing Risks and Multistate Models with R*, Springer, 245 p., (Use R !), ISBN : 9781461420347.

CARRIÈRE J. F. [1994] « Dependent decrement theory », *Transactions of the Society of Actuaries*, Vol. 46, p. 45-74.

CARRIÈRE J. F. [1995] « Removing Cancer when it is Correlated with other Causes of Death », *Biometrical Journal*, Vol. 37, No. 3, p. 339-350.

CHRISTIANSEN M. C. [2012] « Multistate models in health insurance », *AStA Advances in Statistical Analysis*, Vol. 96, No. 2, p. 155-186.

CMIR12. [1991] *The Analysis of Permanent Health Insurance Data*, Continuous Mortality Investigation Bureau, The Institute of Actuaries and the Faculty of Actuaries.

COURBAGE C., ROUDAUT N. [2011] « Long-term care insurance: The French example », *European Geriatric Medicine*, Vol. 2, No. 1, p. 22-25.

CZADO C., RUDOLPH F. [2002] « Application of survival analysis methods to long-term care insurance », *Insurance: Mathematics and Economics*, Vol. 31, No. 3, p. 395-413.

DELEGLISE M. P., HESS C., NOUET S. [2009] « Tarification, provisionnement et pilotage d'un portefeuille Dépendance », *Bulletin Français d'Actuariat*, Vol. 9, No. 17, p. 70-108.

DENUIT M., ROBERT C. [2007] *Actuariat des assurances de Personnes - Modélisation, tarification et provisionnement*, Paris : Economica, 405 p., (Assurance Audit Actuariat), ISBN : 978-2-7178-5329-2.

EFRON B. [1979] « Bootstrap Methods: Another Look at the Jackknife », *The Annals of Statistics*, Vol. 7, No. 1, p. 1-26.

FERRI S., OLIVIERI A. [2000] « Technical bases for LTC covers including mortality and disability projections », *XXXI International ASTIN Colloquium*, Porto Cervo , p. 135-155.

FINE J. P., GRAY R. J. [1999] « A Proportional Hazards Model for the Subdistribution of a Competing Risk », *Journal of the American Statistical Association*, Vol. 94, No. 446, p. 496-509.

GAUZERE F., COMMENGES D., BARBERGER-GATEAU P., LETENNEUR L., DARTIGUES J.-F. [1999] « Maladie et dépendance - Description des évolutions par des modèles multi-états », *Population*, Vol. 54, No. 2, p. 205-222.

GOOLEY T. A., LEISENRING W., CROWLEY J., STORER B. E. [1999] « Estimation of failure probabilities in the presence of competing risks: new representations of old estimators », *Statistics in medicine*, Vol. 18, No. 6, p. 695–706.

GUIBERT Q., PLANCHET F., JUILLARD M. [2010] « Un cadre de référence pour un modèle interne partiel en assurance de personnes », *Bulletin Français d'Actuariat*, Vol. 10, No. 20, p. 5-34.

HABERMAN S., PITACCO E. [1998] *Actuarial Models for Disability Insurance*, Chapman and Hall/CRC, 1^{re} éd., 280 p., ISBN : 0-8493-0389-3.

HELMS F., CZADO C., GSCHLÖBL S. [2004] *Calculation of LTC Premiums based on direct estimates of transition probabilities*, Discussion paper, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

KAPLAN E. L., MEIER P. [1958] « Nonparametric Estimation from Incomplete Observations », *Journal of the American Statistical Association*, Vol. 53, No. 282, p. 457-481.

LOADER C. [1999] *Local Regression and Likelihood*, Springer-Verlag New York Inc., 290 p., (Statistics and Computing Series), ISBN : 0-387-98775-4.

MÜLLER H. G., WANG J. L. [1994] « Hazard Rate Estimation under Random Censoring with Varying Kernels and Bandwidths », *Biometrics*, Vol. 50, No. 1, p. 61-76.

OLIVIERI A., PITACCO E. [2001] « Facing LTC risks », *XXXII International ASTIN Colloquium*, Washington, p. 1-15.

PLANCHET F., THEROND P. E. [2006] *Modèles de durée - Applications actuarielles*, Paris : Economica, 271 p., (Assurance Audit Actuariat), ISBN : 2-7178-5234-4.

PLANCHET F., TOMAS J. [2013] « Uncertainty on Survival Probabilities and Solvency Capital Requirement: Application to LTC Insurance », *Les cahiers de recherche de l'ISFA*, Vol. 2013, No. 4, p. 1-11.

PLISSON M. [2009] *Assurabilité et développement de l'assurance dépendance*, Thèse, Paris : Université Paris Dauphine, 499 p.

PRENTICE R. L. *et autres*. [1978] « The Analysis of Failure Times in the Presence of Competing Risks », *Biometrics*, Vol. 34, No. 4, p. 541-554.

R DEVELOPMENT CORE TEAM. [2013] *R: A Language and Environment for Statistical Computing*, Vienna, Austria, (R Foundation for Statistical Computing), ISBN : 3-900051-07-0.

RICKAYZEN B. D., WALSH D. E. P. [2002] « A Multi-State Model of Disability for the United Kingdom: Implications for Future Need for Long-Term Care for the Elderly », *British Actuarial Journal*, Vol. 8, No. 2, p. 341-393.

TOMAS J., PLANCHET F. [2013] « Multidimensional smoothing by adaptive local kernel-weighted log-likelihood with application to long-term care insurance », *Insurance: Mathematics and Economics*, Vol. 52, No. 3, p. 573-589.

TOMAS J. [2011] « A Local Likelihood Approach to Univariate Graduation of Mortality », *Bulletin Français d'Actuariat*, Vol. 11, No. 22, p. 105-153.

TOMAS J. [2013] *Quantifying biometric life insurance risks with non-parametric smoothing methods*, Dissertation, Amsterdam : Faculty of Economics and Business, 207 p.

TSIATIS A. [1975] « A non identifiability aspect of the problem of competing risks », *Proceedings of the National Academy of Sciences*, Vol. 72, No. 1, p. 20-22.

DE WREEDE L. C., FIOCCO M., PUTTER H. [2011] « mstate: An R Package for the Analysis of Competing Risks and Multi-State Models », *Journal of Statistical Software*, Vol. 38, No. 7.