

Mesurer la prudence d'une table d'expérience

Formation SEPIA du
23/11/2010



Aymric KAMEGA Actuaire
akamega@winter-associes.fr

Le secteur de l'assurance est aujourd'hui confronté à une triple mutation :

- prudentielle avec l'avènement du futur cadre prudentiel européen qui résultera du projet Solvabilité 2 ;
- comptable avec les normes IFRS en général, et avec la préparation de la seconde phase de la norme internationale IFRS consacrée aux contrats d'assurance en particulier ;
- du reporting financier avec le recours de plus en plus massif aux méthodes de Market Consistent Embedded Value de valorisation de compagnie d'assurance.

Ces trois référentiels s'inscrivent dans une même logique d'uniformisation internationale (à tout le moins européenne) et de transparence. Pour cela, la référence au marché est omniprésente (ce principe n'est pas sans poser de problèmes conceptuels et opérationnels).



INTRODUCTION GENERALE



WINTER
& ASSOCIÉS

En particulier, concernant le passif :

- les provisions techniques doivent être évaluées à partir de la meilleure estimation des flux de trésorerie futurs, ce qui nécessite de prendre en compte l'expérience du portefeuille assuré ;
- le recours à un quantile d'ordre élevé conduit à porter une attention nouvelle aux queues de distribution des provisions.

De manière générale, ce support ne doit pas être isolé des précisions orales données lors de la présentation.

SOMMAIRE

SEPIA du 23/11/2010 :
Mesurer la prudence
d'une table
d'expérience

1. Présentation des différents risques

1.1 Risques mutualisables (risque d'échantillonnage)

1.2 Risques systématiques (risque d'estimation, risque de modèle)

2. Mesure du risque d'estimation

2.1 Intervalles de confiance

2.2 Mesure sur les taux ajustés

2.3 Mesure sur les espérances de vie

2.4 Mesure sur les provisions

3. Réduction du risque d'estimation en cas d'hétérogénéité

3.1 Impact d'une réduction d'échantillon

3.2 Choix du modèle d'hétérogénéité

3.3 Évolution des risques d'estimation selon les modèles



1. PRESENTATION DES DIFFERENTS RISQUES

1.0. Introduction

Au regard de la différence de modélisation entre un aléa frappant différemment les diverses têtes du portefeuille et un aléa frappant de manière systématique l'ensemble du portefeuille, il est indispensable de décrire les risques en fonction de leur caractère mutualisable ou systématique.

Il convient également de noter qu'un risque mutualisable et un risque systématique présentent un impact différent sur la distribution de l'engagement :

- la prise en compte des risques mutualisables ne modifie pas la valeur du *best estimate*, mais introduit de la volatilité dans la distribution de l'engagement ;
- la prise en compte des risques systématiques modifie la valeur du *best estimate* et introduit de la volatilité dans la distribution de l'engagement.



1. PRESENTATION DES DIFFERENTS RISQUES

1.1. Risques mutualisables

Les risques mutualisables associés à la probabilité de payer le flux correspondent aux fluctuations d'échantillonnage, conséquences de la taille finie du groupe entraînant une fluctuation de l'espérance empirique (valeur observée de la somme des flux futurs actualisés) autour de l'espérance mathématique (valeur théorique de la somme des flux futurs actualisés si le modèle est correctement spécifié et ajusté). On considère ici que la charge de sinistres est distribuée normalement autour de son espérance.

L'impact des risques mutualisables sur le coefficient de variation de l'engagement dépend du type de contrat observé et dépend de la taille du portefeuille.

En pratique, l'impact sera, par définition, d'autant plus faible que la taille du portefeuille est grande.



1. PRESENTATION DES DIFFERENTS RISQUES

1.2. Risques systématiques

Les risques systématiques associés à la probabilité de payer le flux correspondent aux erreurs sur les lois de survenance et de maintien et peuvent provenir de deux sources :

- Une incertitude sur l'estimation de la loi d'expérience (l'estimation des lois d'expérience n'étant qu'une approximation de la « vraie loi » sous-jacente).
- Le modèle retenu est faux, suite à un mauvais choix de spécification, ou suite à une déformation de la loi sous-jacente au fil du temps.

Contrairement aux risques mutualisables, les risques systématiques (risque d'estimation et risque de modèle) présentent un impact identique sur le coefficient de variation de l'engagement, et ce quelle que soit la taille du portefeuille.



1. PRESENTATION DES DIFFERENTS RISQUES

1.2. Risques systématiques (focus sur le risque d'estimation)

Les risques d'estimation sont d'autant plus importants que la taille de l'échantillon retenu pour construire la loi d'expérience est faible. Lorsque les lois d'expérience sont construites à partir d'une base de données d'un volume significatif, le risque d'estimation devient secondaire au regard du risque de modèle.

Lors de la mise en place de lois *best estimate*, le risque d'estimation doit en particulier être considéré dans les trois cas suivants :

- mortalité prospective : les portefeuilles de rentes sont de petites tailles ;
- invalidité : il s'agit ici de construire des lois avec de petits effectifs ;
- autres risques : dans le cadre d'une segmentation destinée à rendre compte de l'hétérogénéité du portefeuille.

2. MESURE DU RISQUE D'ESTIMATION

2.1. Intervalles de confiance

Les taux de décès bruts présentent des irrégularités directement liées aux fluctuations d'échantillonnage.

Les intervalles de confiance permettent d'encadrer les taux de décès théoriques réels au titre de ces fluctuations d'échantillonnage.

On cherche ici un intervalle de confiance pour q_x (taux de décès réel et théorique à l'âge x dans la population) à partir de q_x (taux de décès observé à l'âge x dans un échantillon exposé au risque de taille R_x).

En retenant l'estimateur des taux bruts de Hoem, qui considère entre autres que les décès suivent une loi binomiale, et en considérant que R_x est grand, on a :

$$Q_x \sim N \left(q_x ; \sigma_x = \sqrt{\frac{q_x(1-q_x)}{R_x}} \right).$$

2. MESURE DU RISQUE D'ESTIMATION

2.1. Intervalles de confiance

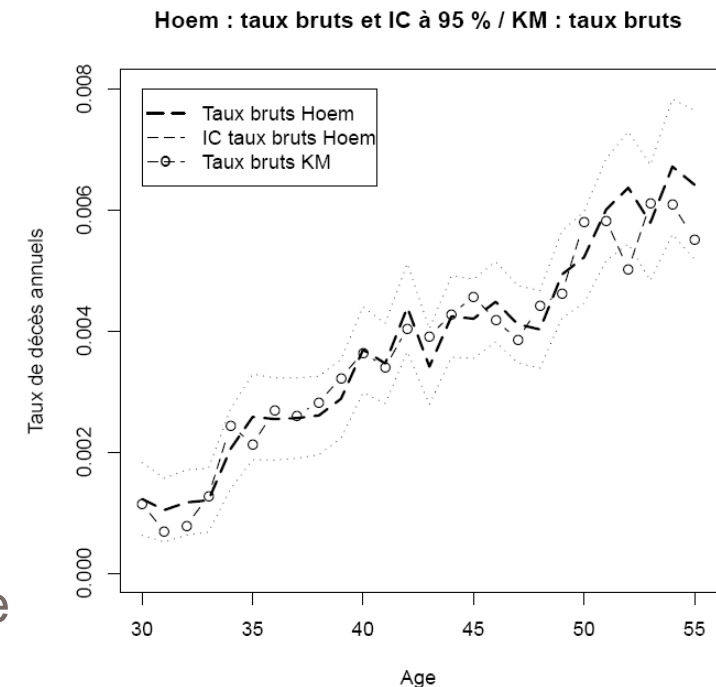
On en déduit que l'expression approchée des bornes de l'intervalle de confiance de q_x est :

$$q_x \pm u_{\alpha/2} \sqrt{\frac{q_x(1-q_x)}{R_x}}$$

où $u_{\alpha/2}$ désigne le quantile d'ordre $\alpha/2$ de la loi normale centrée réduite et $1-\alpha$ désigne le niveau de confiance.

Le graphique ci-contre présente les taux de décès bruts estimés selon la méthode de Hoem, les intervalles de confiance à 95 % associés, ainsi que les taux bruts déduits de l'estimation de la fonction de survie selon Kaplan Meier (KM).

On note que les taux bruts estimés selon l'approche KM sont compris dans les bornes de l'intervalle de confiance des taux bruts estimés par Hoem (hormis à 52 ans).



2. MESURE DU RISQUE D'ESTIMATION

2.2. Mesure sur les taux de décès ajustés

2.2.1. Méthode 1 : simulation directe des taux bruts

On s'appuie sur un modèle de Brass, c'est-à-dire un modèle à référence externe tel que $y_x = a \times z_x + b + \varepsilon_x$, avec $y_x = \ln\left(q_x / (1 - q_x)\right)$ et $z_x = \ln\left(q_x^{ref} / (1 - q_x^{ref})\right)$.

On cherche à présenter ici le risque d'estimation au titre des fluctuations d'échantillonnage, c'est-à-dire au titre des données uniquement. Dans ce contexte, on suppose qu'il n'y a pas de risque de modèle. Deux approches peuvent être envisagées.

Dans une **première méthode** on considère des taux bruts estimés selon l'approche de Hoem.

$$\text{On a alors : } Q_x \sim N\left(q_x; \sqrt{\frac{q_x(1-q_x)}{R_x}}\right).$$

2. MESURE DU RISQUE D'ESTIMATION

2.2. Mesure sur les taux de décès ajustés

2.2.1. Méthode 1 : simulation directe des taux bruts

On génère alors k simulations des taux bruts de décès selon cette loi, et pour chaque simulation k , on détermine une estimation des paramètres $\theta^k = (a^k, b^k)$.

On en déduit ensuite k réalisations des taux ajustés (on parle alors de taux simulés) :

$$q_x(\hat{\theta}^k) = \frac{\exp(a^k z_x + \hat{b}^k)}{1 + \exp(a^k z_x + \hat{b}^k)} .$$

Le risque d'estimation peut alors être mesuré par le coefficient $c(\Psi_x) = \frac{\Psi_x}{q_x(\hat{\theta})}$, où

$$q_x(\hat{\theta}) = \frac{\exp(a \times z_x + \hat{b})}{1 + \exp(a \times z_x + \hat{b})} \text{ et } \Psi_x = \sqrt{E \left[\left(q_x(\hat{\theta}^k) - q_x(\hat{\theta}) \right)^2 \right]} .$$

2. MESURE DU RISQUE D'ESTIMATION

2.2. Mesure sur les taux de décès ajustés

2.2.2. Méthode 2 : simulation des résidus

La **seconde approche** présentée pour mesurer le risque d'estimation s'appuie sur la simulation des résidus (de laquelle on déduit des réalisations de taux bruts, puis des réalisations de taux ajustés au titre des fluctuations d'échantillonnage – on parle alors de taux simulés –).

Après avoir tester l'adéquation de la distribution des « vrais » résidus observés du modèle à la loi normale, on calcule empiriquement la moyenne et l'écart-type de la distribution des résidus puis on génère sur ces bases des erreurs aléatoires selon la loi normale.

On a ainsi pour différents scénarios k , $y_x^k = a \times z_x + \hat{b} + e_x^k$. Les « taux de décès bruts »

associés aux k scénarios sont alors ici $q_x^k = \frac{\exp(a \times z_x + \hat{b} + e_x^k)}{1 + \exp(a \times z_x + \hat{b} + e_x^k)}$.

2. MESURE DU RISQUE D'ESTIMATION

2.2. Mesure sur les taux de décès ajustés

2.2.2. Méthode 2 : simulation des résidus

Pour chaque scénario k , on calcule à partir de ces « taux bruts » une estimation des paramètres $\theta^k = (a^k, b^k)$.

On en déduit ensuite k réalisations des taux ajustés (on parle alors de taux simulés) :

$$q_x^k(\hat{\theta}^k) = \frac{\exp\left(a^k z_x + \hat{b}^k\right)}{1 + \exp\left(a^k z_x + \hat{b}^k\right)} .$$

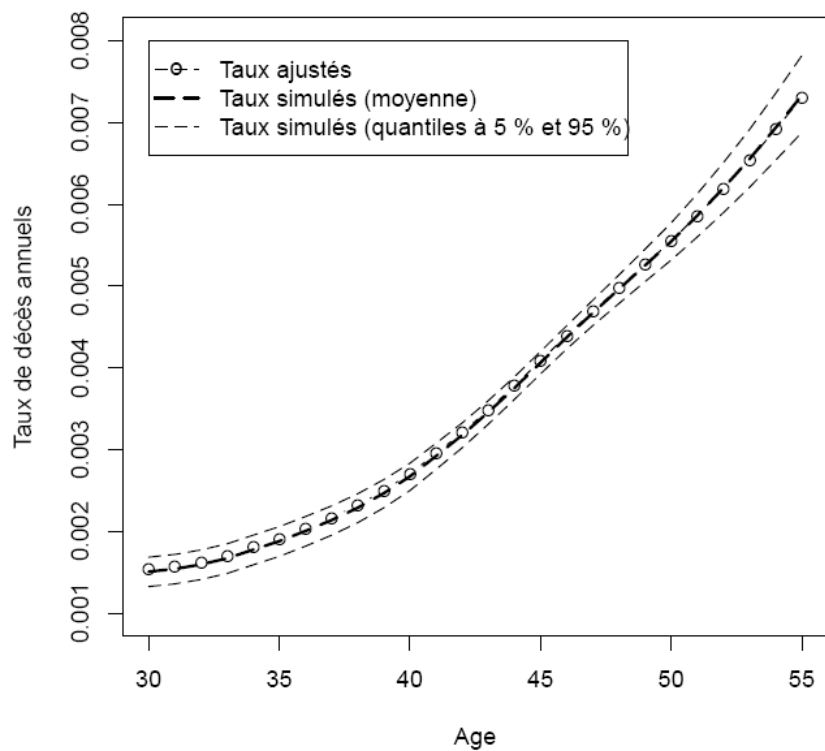
Le risque d'estimation peut alors être mesuré par le coefficient $c(\Psi_x) = \frac{\Psi_x}{q_x(\hat{\theta})}$.

2. MESURE DU RISQUE D'ESTIMATION

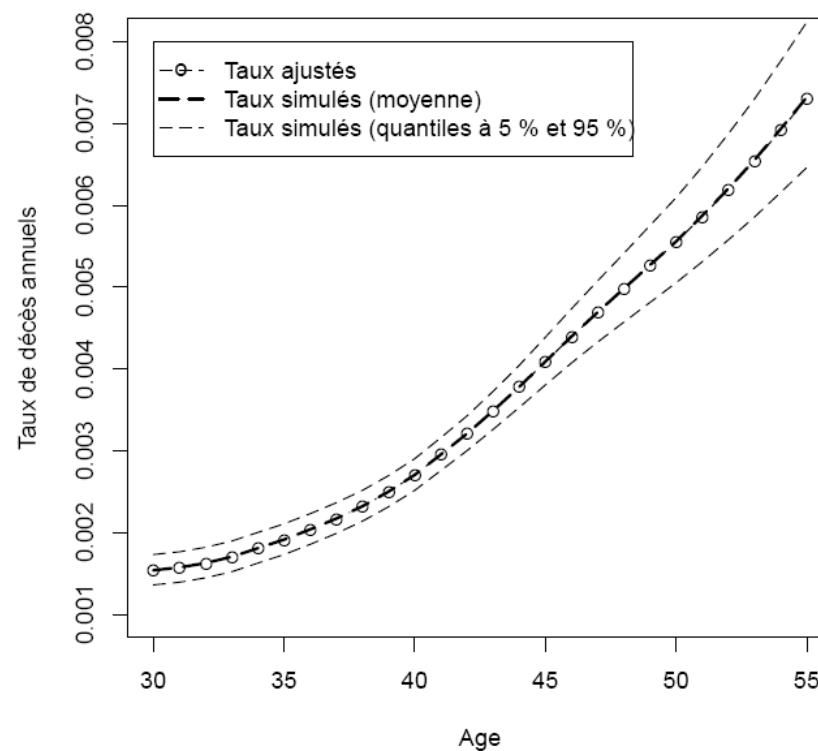
2.2. Mesure sur les taux de décès ajustés

2.2.3. Illustration des taux de décès ajustés et simulés

Simulation directe des taux bruts



Simulation des résidus



2. MESURE DU RISQUE D'ESTIMATION

2.2. Mesure sur les taux de décès ajustés

2.2.3. Illustration des mesures de risques d'estimation (moyenne)

Méthode de simulation des taux bruts (risque d'estimation)	Mesure du risque d'estimation (en % des taux ajustés)
Méthode 1 : simulation « directe » des taux bruts	4,07%
Méthode 2 : simulation des résidus	5,59%

La mesure du risque d'estimation issue de la simulation des résidus majore en moyenne de plus de 35 % celle issue de la simulation directe des taux bruts (les résultats se basent sur 15 000 simulations).

Ces écarts s'expliquent par les hypothèses retenues : la méthode de simulation des résidus repose largement sur l'hypothèse d'absence d'erreur de spécification du modèle, ce qui ne reflète pas la réalité (en pratique, il existe une erreur de spécification du modèle). Ainsi, pour cette méthode, les mesures de risque d'estimation présentées incluent également un risque de modèle.



2. MESURE DU RISQUE D'ESTIMATION

2.3. Mesure sur les espérances de vie partielles

Outre la présentation et la comparaison des taux de décès, il est nécessaire de comparer les tables de mortalité associées, et à cet effet il convient d'utiliser une fonctionnelle propre à chaque table qui lui associe un nombre positif.

L'espérance de vie est de ce point de vue une fonctionnelle naturelle.

Le graphique suivant présente, pour chaque méthode de simulation, les estimations des fonctions de densité (par l'estimateur à noyau) des espérances de vie résiduelles (entre 30 et 55 ans) établies à partir des taux de décès simulés.

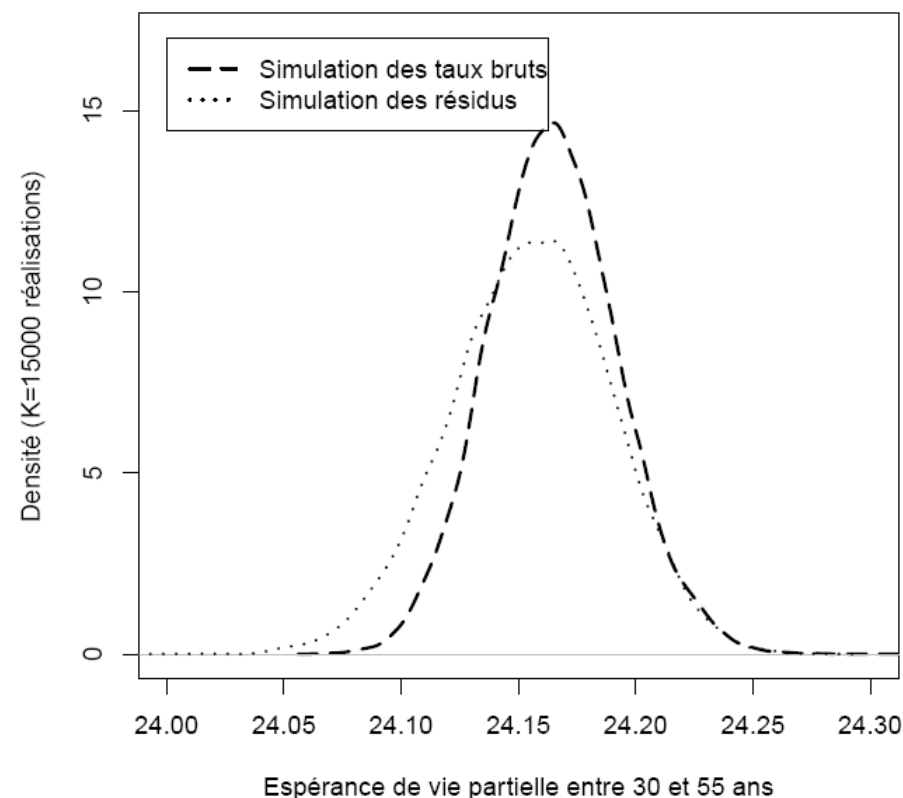
2. MESURE DU RISQUE D'ESTIMATION

2.3. Mesure sur les espérances de vie partielles

Il apparaît sur le graphique ci-contre que la distribution de l'espérance de vie partielle issue de la simulation directe des taux bruts présente une queue moins épaisse que celle issue de la simulation des résidus.

En d'autres termes, les tables de mortalité issues des taux simulés à partir de la simulation des taux bruts présentent une volatilité plus faible (au titre du risque d'estimation) que les tables issues des taux simulés à partir de la simulation des résidus.

Distribution des espérances de vie partielles à 30 ans





2. MESURE DU RISQUE D'ESTIMATION



2.4. Mesure sur les provisions

2.4.0. Introduction

On se place désormais dans un contexte d'assurance et on choisit le montant des provisions comme variable d'intérêt.

On considère plus précisément des provisions relatives à des engagements temporaires au décès (la durée de l'engagement est de d années) et évaluées à partir des taux de décès simulés (l'aléa sur les taux de décès utilisé ici est établi à partir de la simulation directe des taux de décès bruts, car elle permet une meilleure mesure du risque lié aux fluctuations d'échantillonnage).

2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.1. Provisions déterministes

En faisant l'hypothèse de décès en milieu d'année, on en déduit que le montant de la provision pour la réalisation k est : $L_0^k = \sum_{t=0}^{d-1} F_x^k(t) \times (1+r_{t+1})^{-t-1/2}$.

où r_t représente le taux d'actualisation des flux de l'échéance t et avec

$$F_x^k(t) = \begin{cases} C \times q_x(\hat{\theta}^k), & t=0 \\ C \times q_{x+t}(\hat{\theta}^k) \times \prod_{j=0}^{t-1} (1 - q_{x+j}(\hat{\theta}^k)), & 1 \leq t < d \end{cases}$$

L'impact du risque d'estimation sur les provisions peut alors être mesuré par le coefficient $c(\gamma) = \frac{\gamma}{L_0}$, où L_0 est la provision calculée à partir des taux de décès ajustés $q_x(\hat{\theta})$ et en notant $\gamma = \sqrt{E[(L_0^k - L_0)^2]}$.

On note enfin $\bar{L}_0 = \frac{1}{K} \sum_{j=1}^K L_0^j$. Ici, on considère 15 000 scénarios de taux de décès.

2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.2. Provisions stochastiques

Les provisions stochastiques visent à introduire un aléa sur la date de survenance du sinistre (risque mutualisable), au-delà de l'aléa sur le taux de décès (risque systématique). En pratique, cette approche nécessite de simuler des durées de vie restantes, et donc le moment du sinistre, selon la loi des taux de décès.

Pour simuler des durées de vie restantes, on génère pour différents scénarios une variable aléatoire V de loi uniforme sur $[0,1]$ et on ne considère que la partie entière de la durée de vie restante à l'âge x , notée T_x .

T_x est ainsi un entier qui satisfait $T_x p_x \geq V > T_{x+1} p_x$, soit $\prod_{j=0}^{T_x-1} p_{x+j} \geq V > \prod_{j=0}^{T_x} p_{x+j}$ (c'est-à-dire que la variable T_x est la plus grande durée entière précédant le moment du décès).

2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.2. Provisions stochastiques (avec taux de décès ajustés)

Les probabilités de survie utilisées pour simuler les durées de vie restantes sont donc $p_x(\hat{\theta}) = 1 - q_x(\hat{\theta})$.

On considère désormais que le flux aléatoire de prestation à payer en t est

$$F_x(t) = \begin{cases} C, & T_x = t < d \\ 0, & \text{sinon} \end{cases}$$

Le montant aléatoire de l'engagement en 0 s'écrit alors $\Lambda = F_x(t) \times \left(1 + r_{T_x+1}\right)^{-T_x - 1/2}$.

Sur la base de g ($g \in [1, G]$) réalisations de T_x , on obtient alors des réalisations $\lambda_1, \lambda_2, \dots, \lambda_G$ de Λ .

On note L_0 la moyenne des provisions stochastiques : $L_0 = \bar{\lambda} = \frac{1}{G} \sum_{i=1}^G \lambda_i$. En pratique, on génère ici $37,5 \cdot 10^6$ de scénarios de durées de vie restantes.

2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.2. Provisions stochastiques (avec taux de décès simulés)

Pour calculer les provisions stochastiques avec les taux simulés, on considère deux sources d'aléas : une première source liée aux taux de décès stochastiques Π (taux de décès $q_x(\hat{\theta}^k)$ pour une réalisation k) et une seconde liée à la loi conditionnelle des engagements stochastiques $\Lambda | \Pi$.

Les provisions stochastiques avec taux simulés sont déterminées à partir des flux suivants : $F_x^k(t) = \begin{cases} C, & T_x^k = t < d \\ 0, & \text{sinon} \end{cases}$. Pour une simulation k donnée des taux de décès ajustés, le montant aléatoire de l'engagement en 0 s'écrit alors : $\Lambda^k = F_x^k(t) \times \left(1 + r_{T_x^k+1}\right)^{-T_x^k - 1/2}$

La provision stochastique moyenne avec taux simulés : $\bar{L}_0 = \frac{1}{K} \sum_{j=1}^K L_0^j = \frac{1}{K} \sum_{j=1}^K \bar{\lambda}^j = \frac{1}{K} \sum_{j=1}^K \left(\frac{1}{G} \sum_{i=1}^G \lambda_i^j \right)$

Les résultats se base ici sur 15 000 simulations de taux de décès, et pour chaque simulation de taux, 2 500 simulations de durées de vie restantes.

2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.3. Illustrations pour une provision de 5 ans

Statistique	Provision déterministe et taux ajustés	Provision stochastique et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)	Provision stochastique et taux simulés (risque d'estimation)
Moyenne	$L_0 = 8,12.10^{-3}$	$L_0 = 8,11.10^{-3}$	$\bar{L}_0 = 7,98.10^{-3}$	$\bar{L}_0 = 7,99.10^{-3}$

Il ressort que la prise en compte du risque systématique impacte légèrement la provision calculée (baisse comprise entre 1,5 % et 1,7 % lors de la prise en compte du risque systématique).

Concernant la dispersion de la distribution de la provision déterministe calculée à partir des taux simulés (risque d'estimation), il apparaît que le coefficient est égal à $c(\Upsilon) = 6,64 \%$.

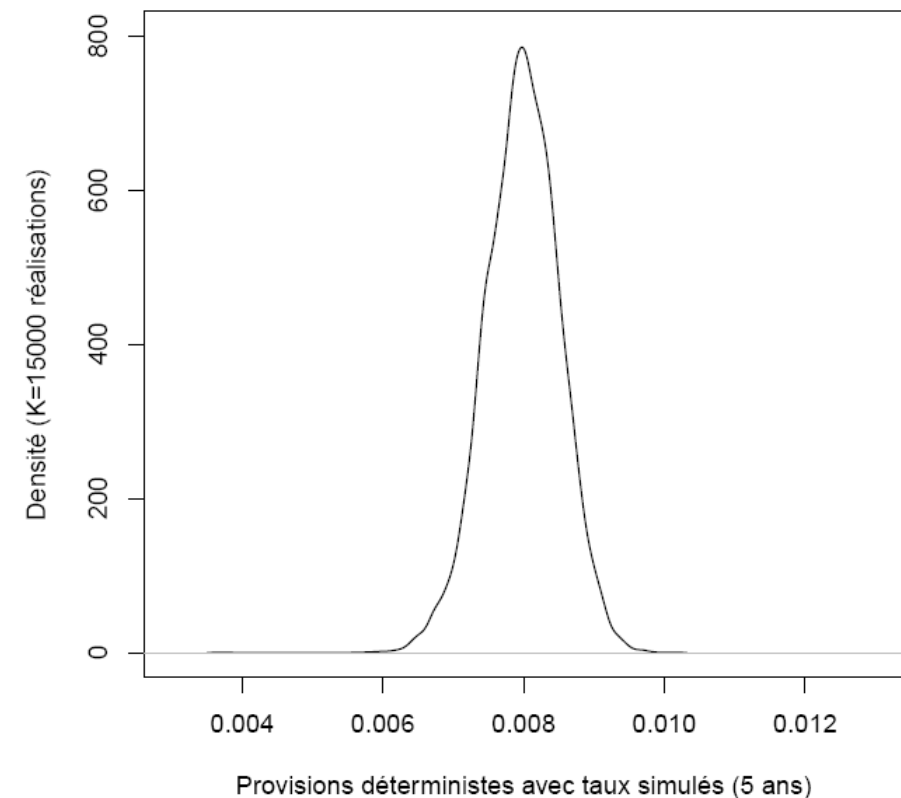
2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.3. Illustrations pour une provision de 5 ans

La dispersion de la distribution peut également être représentée par une estimation de la fonction de densité (par l'estimateur à noyau) de la provision déterministe calculée à partir des taux simulés.

Distribution des provisions (risque systématique)



2. MESURE DU RISQUE D'ESTIMATION

2.4. Mesure sur les provisions

2.4.3. Illustrations pour une provision de 20 ans

Statistique	Provision déterministe et taux ajustés	Provision stochastique et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)	Provision stochastique et taux simulés (risque d'estimation)
Moyenne	$L_0 = 3,83.10^{-2}$	$L_0 = 3,84.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$

Il ressort que la prise en compte du risque systématique impacte de manière limitée la provision calculée (baisse comprise entre 0,8 % et 1 % lors de la prise en compte du risque systématique).

Concernant la dispersion de la distribution de la provision déterministe calculée à partir des taux simulés (risque d'estimation), il apparaît que le coefficient est égal à $c(\Upsilon) = 3,18\%$. La dispersion obtenue ici pour la provision de 20 ans apparaît ainsi près de deux fois inférieure à celle observée pour la provision de 5 ans.

2. MESURE DU RISQUE D'ESTIMATION

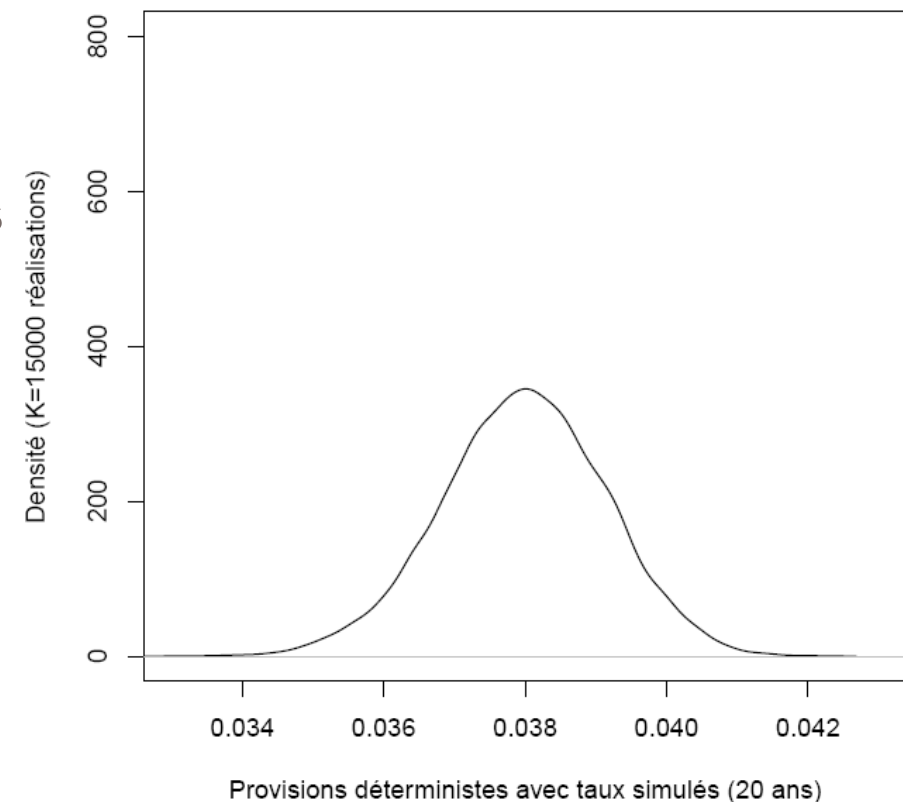
2.4. Mesure sur les provisions

2.4.3. Illustrations pour une provision de 20 ans

La dispersion de la distribution peut également être représentée par une estimation de la fonction de densité (par l'estimateur à noyau) de la provision déterministe calculée à partir des taux simulés.

L'échelle en ordonnée de ce graphique est identique à celle du graphique précédent, et on note toutefois que le montant moyen de la provision est plus de quatre fois supérieur à celui associé au graphique précédent.

Distribution des provisions (risque systématique)



3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.0. Contexte

Dans le cadre du calcul du *best estimate*, il convient prendre en compte l'hétérogénéité du portefeuille.

En effet, avec une approche trop globale pour l'ensemble du portefeuille, on court le risque de n'être finalement *best estimate* sur aucune des sous-populations significatives composant le portefeuille et de voir la table devenir inadaptée dès que le portefeuille se déforme avec le temps.

Par exemple, en construisant une table de mortalité unisexe, la mortalité des hommes sera sous-estimée et celle des femmes surestimée et toute modification du *sex ratio* conduira à une inadéquation de la table avec le risque.

Quelle que soit la sophistication du modèle, la prise en compte de l'hétérogénéité conduit à segmenter la population d'expérience en sous-populations.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.0. Contexte

A contrario, une segmentation trop fine fait apparaître :

- un risque d'estimation ;
- un risque de modèle.

Le risque d'estimation et le risque de modèle sont des risques systématiques, donc potentiellement dangereux.

Une réflexion s'impose alors sur le choix du niveau de segmentation optimal. La quantification du risque d'estimation associé à la construction de tables d'expérience peut être un critère utile à cet effet.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.1. Impact d'une réduction d'échantillon

3.1.0. Introduction

Les illustrations précédentes ont été réalisées à partir à des données de toute la population UEMOA. Ces données incluent les données de trois pays, présentées dans le tableau ci-dessous (il s'agit de données sur les 30-55 ans).

Homme (pop. ass.)	Exposition	Âge moyen	Taux décès moyen	Taux décès moyen (borne inf. à 95%)	Taux décès moyen (borne sup. à 95%)
CI	549 656	43,9 ans	0,40%	0,38%	0,41%
ML	12 114	42,5 ans	0,22%	0,14%	0,31%
TG	133 779	43,2 ans	0,42%	0,39%	0,46%
UEMOA (CI-ML-TG)	695 549	43,8 ans	0,40%	0,38%	0,41%

L'objectif ici est alors de présenter l'évolution du risque systématique lorsque l'on passe de la population UEMOA à celle d'un pays.

Le pays retenu pour cette illustration est le Togo (TG), pays intermédiaire entre la Côte d'Ivoire (CI) et le Mali (ML) en termes d'exposition au risque.

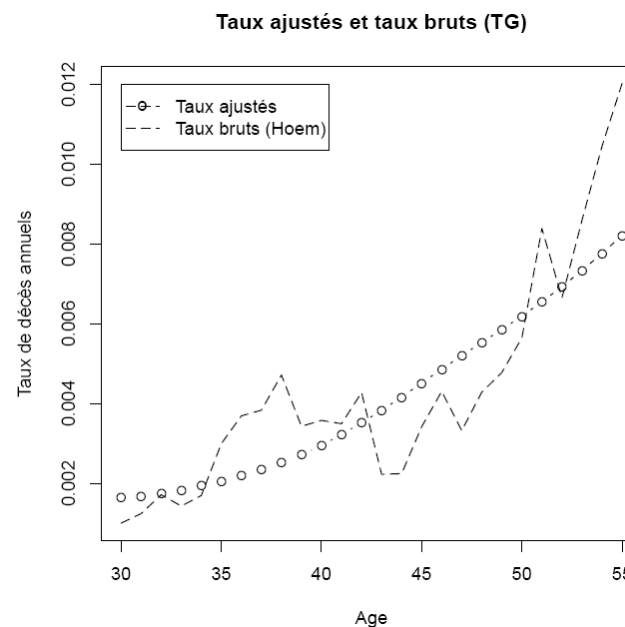
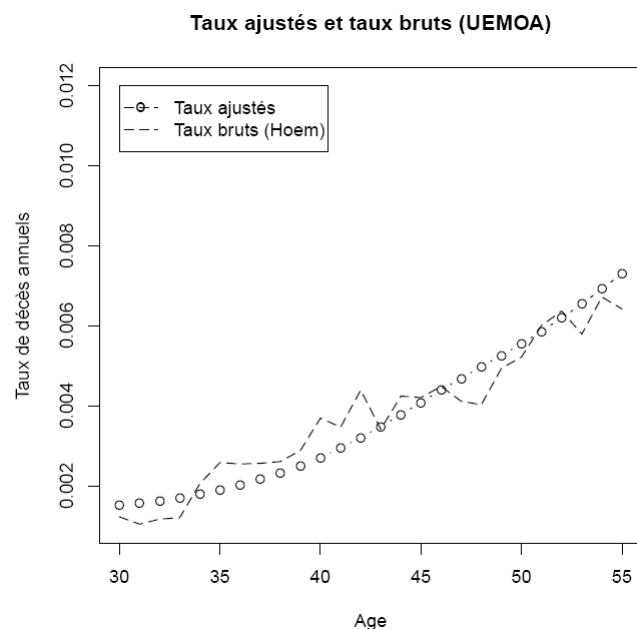
3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.1. Impact d'une réduction d'échantillon

3.1.1. Impact sur les taux de décès

Les graphiques ci-dessous présentent les taux ajustés et les taux bruts pour les données de l'UEMOA et du Togo. Il y apparaît que les fluctuations d'échantillonnage sont logiquement plus importantes avec les données du Togo.



3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

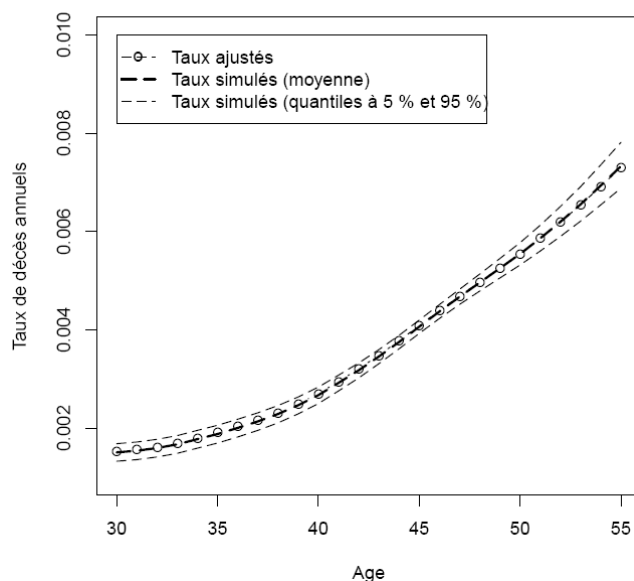


3.1. Impact d'une réduction d'échantillon

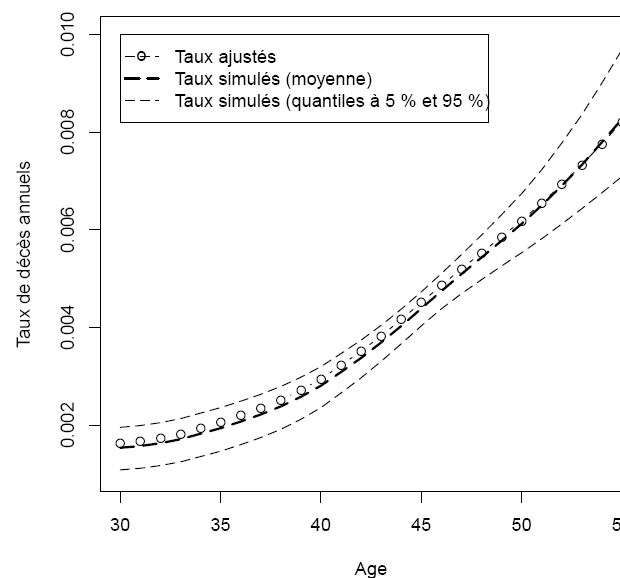
3.1.1. Impact sur les taux de décès

Les graphiques ci-dessous présentent les taux ajustés et les taux simulés pour les données de l'UEMOA et du Togo. Il ressort que la dispersion des taux de décès simulés à partir des données du Togo est plus importante.

Simulation directe des taux bruts (UEMOA)



Simulation directe des taux bruts (TG)



3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.1. Impact d'une réduction d'échantillon

3.1.1. Impact sur les taux de décès

Pour quantifier l'écart de dispersion, le tableau ci-dessous présente une moyenne arithmétique du coefficient $c(\psi_x)$ pour tous les âges $x \in [x_m, x_M]$.

Population	Mesure du risque d'estimation (en % des taux ajustés)
UEMOA (CI-ML-TG)	4,07%
TG	10,03%

La moyenne du coefficient augmente de près de 150 %.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

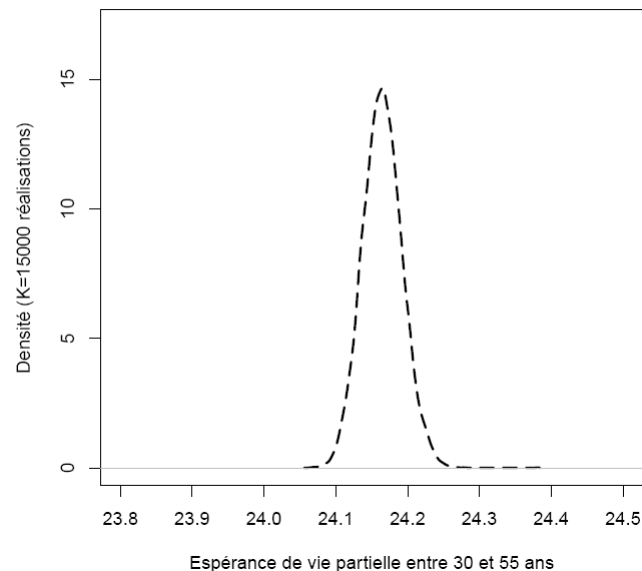


3.1. Impact d'une réduction d'échantillon

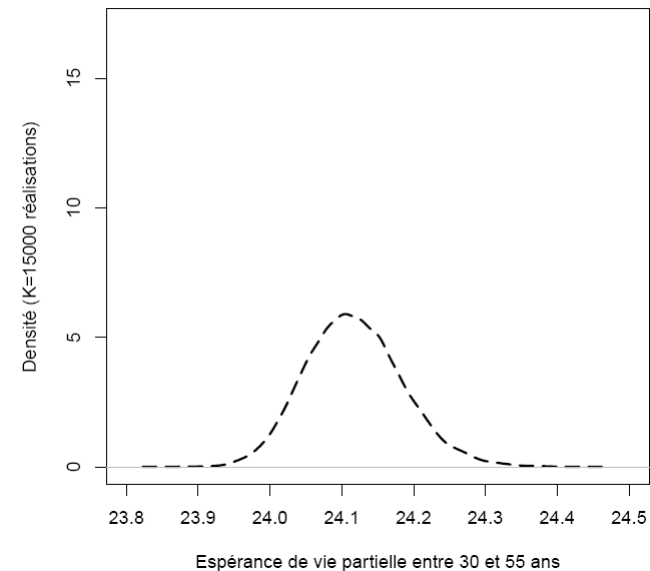
3.1.2. Impact sur les espérances de vie partielles

Dans la continuité des observations ci-dessus, il apparaît sur ces graphiques que la distribution de l'espérance de vie partielle issue de la population UEMOA présente une queue moins épaisse que celle issue de la population du Togo.

Distribution des espérances de vie partielles (UEMOA)



Distribution des espérances de vie partielles (TG)



3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.1. Impact d'une réduction d'échantillon

3.1.3. Impact sur les provisions déterministes

Statistique	UEMOA		TG	
	Provision déterministe et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)	Provision déterministe et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)
Moyenne	$L_0 = 3,83.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$	$L_0 = 4,18.10^{-2}$	$\bar{L}_0 = 4,02.10^{-2}$
Quantile à 0,5 %	Sans objet	$3,48.10^{-2}$	Sans objet	$3,17.10^{-2}$
Quantile à 5 %	Sans objet	$3,60.10^{-2}$	Sans objet	$3,53.10^{-2}$
Quantile à 95 %	Sans objet	$3,99.10^{-2}$	Sans objet	$4,47.10^{-2}$
Quantile à 99,5 %	Sans objet	$4,08.10^{-2}$	Sans objet	$4,70.10^{-2}$
Coefficient	Sans objet	$c(\Upsilon) = 3,18 \%$	Sans objet	$c(\Upsilon) = 7,91 \%$

Avec les données du Togo, la prise en compte du risque systématique impacte significativement la provision calculée (baisse de 3,8 %), alors que l'impact est limité avec les données UEMOA (baisse de 0,8 %).

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.1. Impact d'une réduction d'échantillon

La modélisation de l'hétérogénéité à partir de modélisations indépendantes de sous-populations n'est pas appropriée car elle fait apparaître un risque d'estimation significatif, qui conduit à biaiser le *best estimate* et à épaissir la queue de distribution des engagements (ce qui majore le capital de solvabilité requis dans le projet Solvabilité 2, le CRNHR dans le référentiel MCEV ou l'ajustement pour risque dans le projet norme IFRS sur les contrats d'assurance).

La segmentation du portefeuille peut donc présenter un coût significatif en cas de modélisation indépendante de chaque population.

Pour affiner la segmentation, il pourrait être utile de construire un modèle intégrant l'hétérogénéité à partir de facteurs observables.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.2. Choix du modèle d'hétérogénéité

3.2.0. Introduction

On cite ici trois approches possibles pour modéliser l'hétérogénéité :

- la première démarche consiste à modéliser le comportement de chaque sous-population de manière indépendante ;
- la deuxième démarche consiste à se tourner vers des modèles de durée intégrant des facteurs d'hétérogénéité observables à partir de variables explicatives ;
- la troisième démarche consiste à se tourner vers des modèles de durée intégrant des facteurs d'hétérogénéité inobservables.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.2. Choix du modèle d'hétérogénéité

3.2.1. Modèles à variables explicatives et problèmes de dimension

Lorsqu'un phénomène peut être expliqué par plusieurs variables explicatives, on peut se tourner vers des régressions purement paramétriques. L'avantage est qu'on alors peut facilement trouver des estimateurs convergents, mais l'inconvénient est que ces modèles s'appuient sur de nombreuses hypothèses pour le comportement du phénomène observé et présentent à ce titre un risque important de ne pas être fidèle aux données d'expérience.

Une alternative consiste à se tourner vers des régressions non paramétriques, qui ont l'avantage de s'appuyer sur un nombre limité d'hypothèses pour le comportement du phénomène observé. Ces modèles présentent toutefois un inconvénient connu sous le terme de fléau de la dimension (du mathématicien Richard Bellman), relatif à l'augmentation explosive du volume de données liée à l'ajout de dimensions supplémentaires dans un espace mathématiques.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.2. Choix du modèle d'hétérogénéité

3.2.2. Modèles *single-index* (SIM)

Il est donc nécessaire de réduire la dimension des modèles. La méthode qui est considérée ici est la méthode retenue par Lopez [2007] dans sa thèse sur la réduction de dimension en présence de données censurées : il s'agit de la régression *single index*.

Les modèles de régression *single index* (SIM) sont définis par :

$$m(z) = E(Y | Z = z) = f(\delta^T z),$$

où Y représente la variable à expliquer de dimension 1, Z représente les variables explicatives de dimension p , m représente une fonction inconnue telle que $m: \mathbb{R}^p \rightarrow \mathbb{R}$, f représente une fonction de lien inconnue telle que $f: \mathbb{R} \rightarrow \mathbb{R}$ et $\delta \in \Theta \subset \mathbb{R}^p$ est un paramètre inconnu de dimension finie.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.2. Choix du modèle d'hétérogénéité

3.2.2. Modèles de Cox et de Lin et Ying

En pratique il apparaît que le modèle multiplicatif de Cox [1972] et le modèle additif de Lin et Ying [1994] sont des cas particuliers du modèle SIM, dans lesquels les hypothèses ne portent pas sur l'espérance conditionnelle mais sur le taux de hasard instantané conditionnel.

En effet, le modèle multiplicatif de Cox peut s'écrire : $\lambda(t|Z=z) = \lambda_0(t)e^{\delta^T z}$, où λ_0 est une fonction supposée inconnue et δ est un paramètre à estimer (on retrouve ainsi facilement l'écriture d'un modèle SIM).

De même, le modèle additif de Lin et Ying peut s'écrire : $\lambda(t|Z=z) = \lambda_0(t) + \gamma^T z$, où λ_0 est une fonction supposée inconnue et γ est un paramètre à estimer.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.2. Choix du modèle d'hétérogénéité

3.2.2. Modèles de Cox et de Lin et Ying

En outre, il apparaît que le modèle multiplicatif de Cox et le modèle additif de Lin et Ying se rejoignent lorsque que λ_0 est constant (donc invariant dans le temps) et que le terme exponentiel dans le modèle multiplicatif (soit $e^{\delta^T z}$) est remplacé par l'expression linéaire $\{1 + \delta^T z\}$; dans ce cas $\gamma = \lambda_0 \delta$.

Les modèles de Cox et de Lin et Ying présentent toutefois des contraintes :

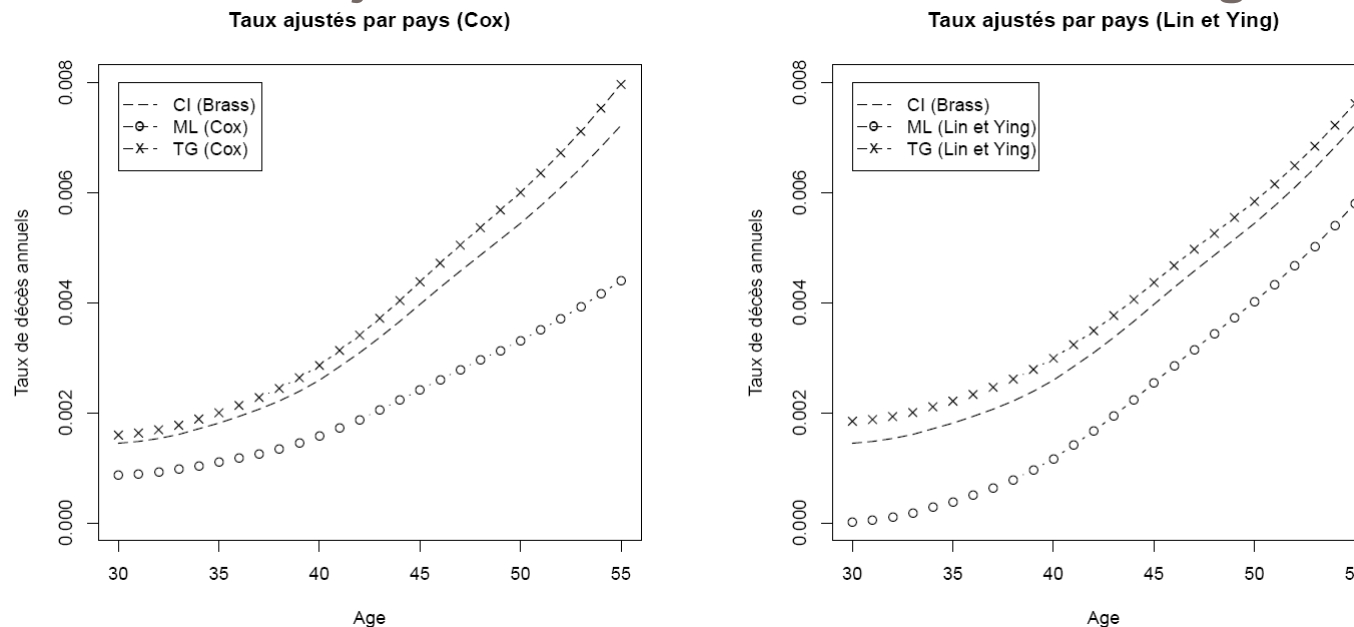
- le modèle de Cox repose sur l'hypothèse que les différences relatives des taux de décès instantanés entre les différents segments sont constantes (hypothèse de proportionnalité) ;
- le modèle de Lin et Ying repose sur l'hypothèse que les différences absolues des taux de décès instantanés entre les différents segments sont constantes.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



3.2. Choix du modèle d'hétérogénéité

3.2.3. Illustration des ajustements de Cox et de Lin et Ying



Dans la continuité des commentaires précédents, il apparaît qu'avec le modèle multiplicatif de Cox, les différences absolues entre les pays augmentent significativement lorsque l'âge augmente, contrairement au modèle de Lin et Ying dans lequel les différences absolues sont constantes pour tous les âges.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.2. Choix du modèle d'hétérogénéité

3.2.3. Comparaison réalisation / prédiction décès

Modèle global UEMOA - H (Brass global)			
Pays	Décès observés	Décès prédits	Différence relative
Côte d'Ivoire	2 188	2 203	0,7%
Mali	27	44	63,8%
Togo	565	511	-9,6%

Modèles intégrant l'hétérogénéité (sans et avec fact. obs.)			
Pays (modèle)	Décès observés	Décès prédits	Différence relative
Côte d'Ivoire (Brass)	2 188	2 144	-2,0%
Mali (Brass*)	27	29	8,4%
Mali (Cox)		26	-3,2%
Mali (Lin et Ying)		26	-4,9%
Togo (Brass)	565	565	-0,1%
Togo (Cox)		548	-3,0%
Togo (Lin et Ying)		550	-2,7%

(*) avec convention pour le traitement des taux de décès bruts nuls.

Le 1^{er} sous-tableau présente les écarts lorsque les taux de décès ajustés sont déterminés sans tenir compte de l'hétérogénéité entre sous-populations. Le 2^{ème} sous-tableau présente les écarts lorsque les taux de décès ajustés sont déterminés avec des modèles intégrant l'hétérogénéité, soit à partir de modèles indépendants pour chaque sous-population (Brass), soit à partir de modèles intégrant l'hétérogénéité à partir de facteurs observables (Cox et Lin et Ying).

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.3. Évolution des risques d'estimation selon les modèles

3.3.1. Impact sur les taux de décès (Togo)

Population	Modèle de Brass (approche 1)	Modèle de Cox (approche 2)	Modèle de Lin et Ying (approche 2)
Togo $c(\psi_{TG})$	9,89 %	6,19 %	6,78 %

L'utilisation du modèle de Brass (modèle indépendant pour chaque sous-population), conduit à un risque d'estimation supérieur à celui obtenu avec le modèle de Cox ou de Lin et Ying (modèles intégrant l'hétérogénéité à partir de facteurs observables). Les résultats se basent sur 1 000 simulations des taux.

Par ailleurs, on note que le risque d'estimation est d'autant plus important que la population présente une faible exposition. Ainsi par exemple, le risque d'estimation avec le modèle de Brass pour la Côte d'Ivoire est égal à 4,73 %, et celui du Mali avec le modèle de Cox est égal à 22,27 %.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE



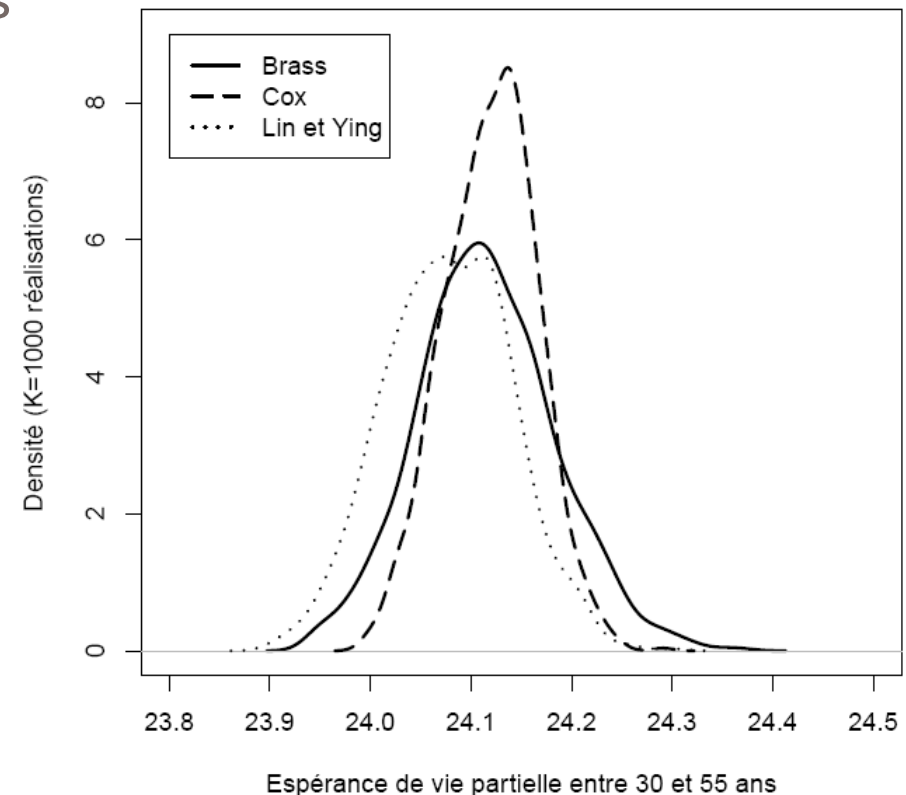
3.3. Évolution des risques d'estimation selon les modèles

3.3.2. Impact sur les espérances de vie partielles (Togo)

Dans la continuité des observations précédentes il apparaît que la distribution de l'espérance de vie partielle issue du modèle de Cox présente une queue moins épaisse que celle issue du modèle de Brass.

En d'autres termes, les tables de mortalité issues des taux simulés à partir du modèle multiplicatif intégrant l'hétérogénéité à l'aide de facteurs observables présentent une volatilité plus faible que celles issues des taux simulés à partir du modèle de Brass utilisé indépendamment pour chaque sous-population.

Dist. des espérances de vie partielles à 30 ans (Togo)



3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.3. Évolution des risques d'estimation selon les modèles

3.3.3. Impact sur les provisions (Togo)

Statistique	Brass		Cox		Lin et Ying	
	Prov. et taux ajustés	Prov. et taux simulés	Prov. et taux ajustés	Prov. et taux simulés	Prov. et taux ajustés	Prov. et taux simulés
Moyenne	$4,18.10^{-2}$	$4,03.10^{-2}$	$4,06.10^{-2}$	$4,01.10^{-2}$	$4,22.10^{-2}$	$4,17.10^{-2}$
Quantile à 0,5 %	Sans objet	$3,18.10^{-2}$	Sans objet	$3,52.10^{-2}$	Sans objet	$3,54.10^{-2}$
Quantile à 5 %	Sans objet	$3,52.10^{-2}$	Sans objet	$3,70.10^{-2}$	Sans objet	$3,74.10^{-2}$
Quantile à 95 %	Sans objet	$4,50.10^{-2}$	Sans objet	$4,34.10^{-2}$	Sans objet	$4,61.10^{-2}$
Quantile à 99,5 %	Sans objet	$4,74.10^{-2}$	Sans objet	$4,50.10^{-2}$	Sans objet	$4,83.10^{-2}$
Coefficient $c(\Upsilon)$	Sans objet	7,91 %	Sans objet	5,15 %	Sans objet	6,41 %

Avec les données du Togo, la prise en compte du risque systématique diminue de 3,6 % la provision calculée lorsque l'on retient le modèle de Brass, alors que la baisse est de 1,2 % pour les modèles de Cox et de Lin et Ying.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.3. Évolution des risques d'estimation selon les modèles

3.3.3. Impact sur les provisions

Les résultats illustrent une diminution de l'impact du risque d'estimation sur les provisions de plus de 50 % avec les modèles de la deuxième approche (Cox et Lin et Ying), par rapport à l'impact observé avec le modèle de la première approche (Brass). En outre, avec les modèles de la seconde approche, la mesure de risque d'estimation diminue de 20 à 35 %.

Néanmoins, dans notre exemple on note en parallèle que le poids du risque d'estimation dans les provisions est comparable au poids du risque de modèle (qui reflète l'évolution de la provision selon le choix de modèle).

Dans la recherche d'un modèle permettant de réduire le risque d'estimation, une attention particulière doit donc également être accordée aux conséquences en termes de risques de modèles.

3. REDUCTION DU RISQUE D'ESTIMATION EN CAS D'HETEROGENEITE

3.3. Évolution des risques d'estimation selon les modèles

Au final, le choix d'un modèle intégrant l'hétérogénéité à partir de facteurs observables présente plusieurs avantages, parmi lesquels on retient :

- la diminution de l'impact du risque d'estimation sur la provision estimée (baisse de plus de 50 % par rapport à l'impact issu du modèle de Brass) ;
- la capacité à modéliser les taux de décès lorsque les données sont sensiblement limitées.

Le choix d'un tel modèle présente toutefois des inconvénients :

- l'hypothèse sur la constance des différences relatives (Cox) ou absolues (Lin et Ying) des taux de décès instantanés ;
- de manière générale, l'impact potentiel en termes de risque de modèle.



BIBLIOGRAPHIE



- Cao H. [2005] *A comparison Between the Additive and Multiplicative Risk Models*, Mémoire d'actuariat, Université Laval.
- Cox D. R. [1972] « Regression Models and Life-Tables », *Journal of the Royal Society. Series B (Methodological)*, Vol. 34, No. 2.
- Hill C., Com-Nougué C., Kramar A., Moreau T., O'Quigley J., Senoussi R., Chastang C. [1990] « Analyse statistique des données de survie », *Inserm, Médecine – Sciences, Flammarion*.
- Klein J. P., Moeschberger M. L. [2005] « Survival Analysis – Techniques for Censored and Truncated Data », *Springer, 2nd edition*.
- Lin D. Y., Ying Z. [1994] « Semiparametric analysis of the additive risk model », *Biometrika*, n. 81.
- Lopez O. [2007] *Réduction de dimension en présence de données censurées*, Thèse de doctorat, Université de Rennes 1.
- Martinussen T., Scheike T. H. [2006] « Dynamic Regression Models for Survival Data », *Springer*.
- Planchet F., Thérond P. [2006] « Modèles de Durée - Applications actuarielles », *Economica*.
- Vaupel J. W., Manton K. G. Stallard E. [1979] « The impact of heterogeneity in individual frailty on the dynamics of mortality », *Demography*, Vol. 16, No. 3.



CONTACTS



Frédéric Planchet

fplanchet@winter-associés.fr

Aymric Kamega

akamega@winter-associés.fr

WINTER & Associés

Bureau de Paris
43-46 avenue de la Grande
Armée
F-75 116 Paris
+33-(0)1-45-72-63-00

Bureau de Lyon
55 avenue René Cassin
F-69 009 Lyon
+33-(0)4-37-37-80-90

<http://www.winter-associés.fr/>
<http://www.ressources-actuarielles.net/>
<http://www.winter-associés.fr/blog> (ou
<http://actudactuaire.typepad.com/laboratoire/>)