

Mesure du risque d'estimation associé à une table d'expérience

Version 1.0 du 24/11/2010

Aymric Kamega ^β

Frédéric Planchet ^{*}

Université de Lyon - Université Claude Bernard Lyon 1
ISFA ^γ
WINTER & Associés ^λ

Résumé

Cette étude présente des outils de mesures de risque d'estimation à partir d'un modèle de Brass (modèle à référence externe). Ces outils sont dans un premier temps appliqués à la mesure du risque associé à la construction d'une loi d'expérience à partir d'une population globale (i. e. sans tenir compte de l'hétérogénéité). Dans un second temps, ces outils sont appliqués à la mesure du risque associé à la construction d'une loi d'expérience à partir d'une sous-population de la population globale, segmentée au titre de la prise en compte de l'hétérogénéité. Dans ce second scénario il apparaît que le risque d'estimation augmente significativement : la modélisation de l'hétérogénéité à partir d'un modèle indépendant pour chaque sous-population n'est donc pas adaptée car elle majore significativement le risque d'estimation.

MOTS-CLEFS : Hoem, Kaplan-Meier, ajustement, modèle à référence externe (Brass), simulation, risque d'estimation, provision déterministe, provision stochastique, hétérogénéité.

^β Aymric Kamega est doctorant au sein du laboratoire SAF à l'ISFA (EA n°2429), actuariaire chez WINTER & Associés et membre du Collège de Direction de l'EURIA (EURO-Institut d'Actuariat, UBO). Contact : akamega@winter-associes.fr

^{*} Frédéric Planchet est docteur en sciences de gestion, enseignant-chercheur du laboratoire SAF à l'ISFA (EA n°2429) et actuariaire associé chez WINTER & Associés. Contact : fplanchet@winter-associes.fr

^γ Institut de Science Financière et d'Assurances (ISFA) - 50 avenue Tony Garnier - 69366 Lyon Cedex 07 - France.

^λ WINTER & Associés - 55 avenue René Cassin - 69009 Lyon - France.

Sommaire

1. Introduction.....	2
2. Intervalles de confiance	4
2.1. Intervalles de confiance sur les taux bruts (Hoem).....	4
2.2. Intervalles de confiance sur la fonction de survie (Kaplan Meier) ...	5
2.3. Intervalles de confiance : illustrations	6
3. Mesure du risque d'estimation sur les taux ajustés	8
3.1. Méthode 1 : simulation directe des taux bruts	9
3.2. Méthode 2 : simulation des résidus.....	10
3.3. Mesure sur les taux ajustés : illustrations	11
4. Mesure du risque d'estimation sur les provisions	16
4.1. Provisions déterministes	16
4.2. Provisions stochastiques	17
4.3. Mesure sur les provisions : illustrations.....	19
5. Impact de la réduction de l'échantillon sur le risque d'estimation.....	22
5.1. Rappel du contexte (hétérogénéité)	22
5.2. Évolution du risque d'estimation sur les taux de décès ajustés	23
5.3. Évolution du risque d'estimation sur les provisions	26
6. Synthèse et conclusion.....	27
Notations des taux de décès	29
Bibliographie	30

1. Introduction

Comme indiqué dans Planchet [2009] en assurance de personnes, l'évaluation des provisions techniques à partir de la meilleure estimation (*best estimate*) des flux de trésorerie nécessite de prendre en compte l'expérience du portefeuille assuré. À cet effet, la détermination d'une loi d'expérience est nécessaire, que celle-ci soit déterminée complètement à l'aide des données disponibles ou calibrée par un positionnement par rapport à une référence externe (*cf.* Planchet et Leroy [2009]). Il s'avère alors indispensable de prendre en compte l'hétérogénéité de la population observée afin de limiter le biais associé (*cf.* Dreesbeke et al. [1989]) et de conférer ainsi à la loi d'expérience une plus grande robustesse face aux déformations futures de la structure de la population.

En pratique, quelle que soit la sophistication du modèle proposé, la prise en compte de l'hétérogénéité conduit de fait à segmenter la population d'expérience en sous-populations. Plus la segmentation est fine et plus les risques d'estimation et de modèle sont importants (principalement en raison de la baisse du volume d'information et de la multiplication des modèles). En d'autres termes, plus on segmente et plus la volatilité autour des tables ajustées augmente, et il arrive un moment où la segmentation devient inefficace. Par exemple, il apparaît dans Alho [2007] que l'introduction du contexte et des causes de mortalité dans les modèles de mortalité prospective n'améliore pas nécessairement la modélisation sur le plan statistique.

Au surplus, en Europe, le recours à un quantile d'ordre élevé pour la mesure du capital de solvabilité dans le dispositif Solvabilité 2 - et également dans le calcul du coût des risques non couvrables (CRNHR) des MCEV ou dans le calcul de l'ajustement pour risque dans la norme IFRS sur les contrats d'assurance - conduit à porter une attention nouvelle aux risques non mutualisables portés par l'organisme assureur du fait de leur poids potentiel dans la queue de la distribution.

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. Elle permet, entre autres, de déterminer un intervalle de confiance sur le calcul d'une provision ou de tout autre indicateur pertinent en fonction du contexte pour une sous-population donnée.

Dans ce contexte, cette étude présente des outils et des illustrations de mesures des risques liés aux données dans le cadre de la construction de tables de mortalité d'expérience à partir d'un modèle à référence externe (modèle de Brass). En particulier, cette étude s'attache à quantifier le risque d'estimation, c'est-à-dire le risque systématique associé à l'utilisation d'une loi de durée dont on sait qu'elle n'est pas exacte du fait des erreurs d'échantillonnage affectant sa construction. L'approche proposée ici ne s'appuie pas sur le cadre général du maximum de vraisemblance, qui propose pourtant des outils naturels de mesure du risque d'estimation, car il n'est pas très bien adapté au contexte des tables d'expérience (on est souvent obligé de recourir à des techniques qui ne sont pas du maximum de vraisemblance).

De manière générale, les outils de mesure présentés ici s'avèrent ainsi particulièrement utiles lorsque les données d'expérience disponibles se révèlent peu volumineuses ou lacunaires.

Plan

L'étude s'attache à mesurer le risque d'estimation (sections 2, 3 et 4), et à quantifier son évolution en cas de baisse de la taille de l'échantillon suite à une segmentation (section 5).

Dans le détail, la deuxième section de l'étude présente les intervalles de confiance associés aux erreurs d'échantillonnage, tant au niveau de l'estimation des taux de décès bruts qu'au niveau de l'estimation de la fonction de survie. Dans la troisième et quatrième section, on s'intéresse ensuite aux mesures du risque d'estimation pour des taux de décès ajustés et dans le cadre d'une problématique de provisionnement, ce qui peut permettre par la suite de définir la fonction de perte. Enfin, dans la cinquième section on segmente la population afin de tenir compte de l'hétérogénéité et on étudie l'évolution du risque d'estimation lors du passage d'une modélisation de la population globale à une modélisation indépendante de chaque segment de la population.

Données

Les illustrations numériques présentées dans cette étude s'appuient sur les données d'expérience utilisées pour la construction des tables de mortalité

réglementaires dans la zone CIMA/FANAF¹ pour les contrats d'assurance en cas de vie et en cas de décès. Les données retenues s'appuient plus précisément sur les données de la population masculine des pays de la sous-région UEMOA, représentés ici par la Côte d'Ivoire, le Mali et le Togo.

Les données utilisées ont été collectées en 2009 et sont présentées en détail dans Planchet et al. [2010]. On retient ici qu'elles sont observées sur les années 2003 à 2006 et comptent des troncatures à gauche (relatives aux entrées après le 01/01/2003) et des censures à droite (relatives aux sorties avant le 31/12/2006 pour un motif autre que le décès).

Dans les sections 2 à 4, toutes les applications numériques sont réalisées à partir de la population d'expérience UEMOA globale. Puis, dans la section 5 on distingue les données par pays afin de prendre en compte l'hétérogénéité à partir d'un modèle indépendant pour chaque pays.

2. Intervalles de confiance

Cette section présente les intervalles de confiance, d'une part pour les taux bruts et d'autre part pour la fonction de survie.

2.1. Intervalles de confiance sur les taux bruts (Hoem)

Les taux de décès bruts, c'est-à-dire les taux de décès observés, présentent des irrégularités directement liées aux fluctuations d'échantillonnage. La description des estimations de taux de décès observés par les seules valeurs estimées est donc insuffisante : il est nécessaire de les décrire par des indicateurs numériques complémentaires. À cet effet, on peut retenir les intervalles de confiance (pour mémoire, dans ce contexte les intervalles de confiance permettent d'encadrer les taux de décès théoriques réels au titre des fluctuations d'échantillonnage).

En d'autres termes, 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). À cet effet, on retient l'estimateur de taux bruts de Hoem, qui considère entre autres que les décès suivent une loi binomiale et que l'on a, pour tout $0 < t < s < 1$, l'approximation suivante : ${}_{s-t}q_{x+t} \approx (s-t)q_x$ (cf. Planchet et Thérond [2006] pour une présentation complète des hypothèses de calculs de cet estimateur). Deux approches sont envisageables pour la détermination des intervalles de confiance dans ce cas.

¹ Zone couvrant les pays membres de la CIMA et représentés par des sociétés d'assurance ou de réassurance de la FANAF, à savoir le Bénin, le Burkina Faso, la Côte d'Ivoire, le Mali, le Niger, le Sénégal et le Togo (c'est-à-dire les pays de l'UEMOA, hors Guinée Bissau) ainsi que le Cameroun, la Centrafrique, le Congo Brazzaville, le Gabon, le Tchad (c'est-à-dire les pays de la CEMAC, hors Guinée Équatoriale).

Si R_x est grand (c'est-à-dire $R_x q_x > 5$ et $R_x(1 - q_x) > 5$, cf. Saporta [2006]), on a $Q_x \sim N\left(q_x; \sigma_x = \sqrt{\frac{q_x(1 - q_x)}{R_x}}\right)$ et on 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.

Si R_x n'est pas assez grand, on peut calculer l'intervalle de confiance à distance finie à partir d'une procédure itérative intégrant la probabilité associée au nombre de décès (cf. Planchet et Théron [2006] pour plus de précisions sur cette alternative). Ce cas nous intéresse peu en pratique.

2.2. Intervalles de confiance sur la fonction de survie (Kaplan Meier)

L'estimateur de la fonction de survie Kaplan Meier présente l'avantage de ne retenir aucune hypothèse sur la répartition des décès sur $[x, x + 1]$. En outre, il permet également d'obtenir des intervalles de confiance et à cet effet, on s'appuie sur l'estimateur de Greenwood de la variance de l'estimateur de Kaplan Meier : $V[S(x)] = S(x)^2 \gamma(x)^2$ (où $S(x)$ représente l'estimateur de

la fonction de survie à l'âge x et $\gamma(x) = \sqrt{\sum_{i=1}^x \frac{d_i}{R_i(R_i - d_i)}}$, avec d_i le nombre de décès observés à l'âge i).

Dans la littérature, on retrouve généralement deux types d'intervalles de confiance pour la fonction de survie (cf. par exemple Klein et Moeschberger [2005]) : les intervalles de confiance ponctuels, c'est-à-dire les intervalles de confiance à un âge x fixé, et les bandes de confiance, c'est-à-dire les intervalles de confiance pour tous les âges compris entre x_m et x_M .

Dans un contexte d'analyse des taux de décès, et en notant que $q_x = 1 - \frac{S(x+1)}{S(x)}$, on comprend que le choix des bandes confiance pour la

fonction de survie est plus approprié. Aussi, on présente ici les bandes de confiance linéaires pour la fonction de survie (ces bandes présentant l'avantage d'être symétriques, contrairement à celles issues de transformations logarithmiques par exemple, cf. Klein et Moeschberger [2005]). Nair définit en 1984 ces intervalles par :

$$S(x) \pm c_\alpha(a_{x_m}, a_{x_M}) S(x) \gamma(x),$$

où $a_X = \frac{p \times \gamma(X)^2}{1 + p \times \gamma(X)^2}$ (p étant la taille de l'échantillon) et $c_\alpha(a_{x_m}, a_{x_M})$

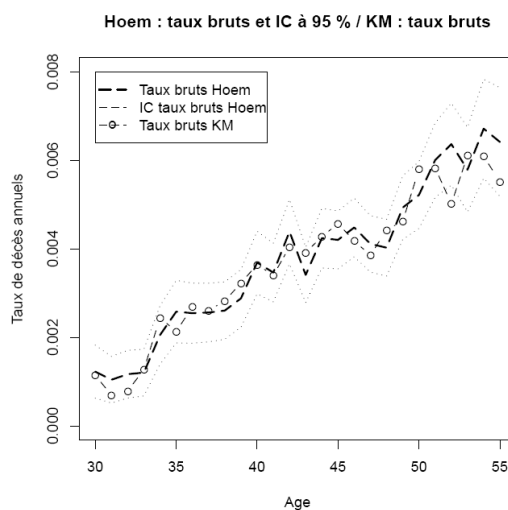
désigne le coefficient de confiance (issu de tables présentées en annexe de Klein et Moeschberger [2005]). On note que les bandes de confiance linéaires ci-dessus présentent l'avantage d'être proportionnelles aux intervalles de confiance ponctuels linéaires (pour mémoire ces intervalles ponctuels s'écrivent $S(x) \pm u_{\alpha/2} S(x) \gamma(x)$, 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, cf. par exemple Planchet et Thérond [2006] ou Martinussen et Scheike [2006]), contrairement aux bandes de confiance linéaires proposées par Hall et Wellner en 1980 (cf. Klein et Moeschberger [2005]).

2.3. Intervalles de confiance : illustrations

Pour les applications numériques, on retient $x_m = 30$ et $x_M = 55$ et on considère à ce stade les données d'expérience globale de l'UEMOA. Par ailleurs, on note qu'avec les données de l'étude et pour un niveau de confiance à $1-\alpha = 95\%$, $c_\alpha(a_{x_m}, a_{x_M})$ est majoré par 2,543 (avec $a_{x_m} = 0,717$ et $a_{x_M} = 0,761$, cf. Klein et Moeschberger [2005]). Aussi, pour l'illustration des bandes de confiance de la fonction de survie présentée en 2.3.2, on suppose que $c_\alpha(a_{x_m}, a_{x_M}) = 2,543$.

2.3.1. Illustrations et comparaisons pour les taux bruts

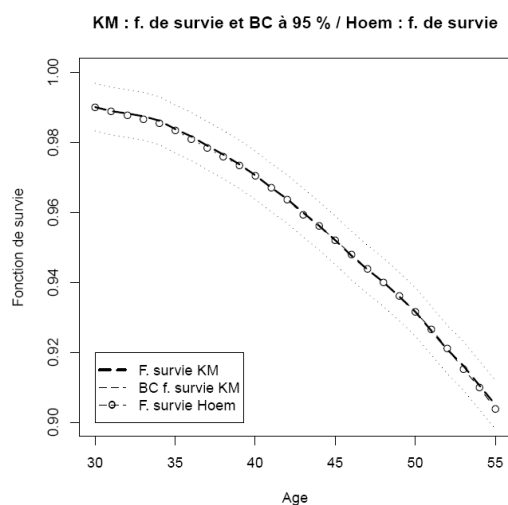
Le graphique ci-dessous présente les taux de décès bruts estimés selon la méthode de Hoem, les intervalles de confiance à 95 % associés, ainsi que les taux bruts déduits de l'estimation de la fonction de survie selon Kaplan Meier.

Figure 1 - Taux de décès bruts et intervalles de confiance (UEMOA)

On note que les deux méthodes d'estimation des taux bruts sont comparables et que les taux bruts estimés selon l'approche Kaplan Meier sont compris dans les bornes de l'intervalle de confiance des taux bruts estimés par Hoem (hormis à 52 ans).

2.3.2. Illustrations et comparaisons pour la fonction de survie

Le graphique ci-dessous présente la fonction de survie estimée selon la méthode de Kaplan Meier, les bandes de confiance à 95 % associées, ainsi que la « fonction de survie » déduite de l'estimation des taux bruts selon Hoem (à noter que dans un souci de comparabilité des deux fonctions de survie, il est supposé ici qu'à l'âge initial $x_m = 30$, la fonction de survie déduite de Hoem est égale à celle issue de Kaplan Meier).

Figure 2 - Fonctions de survie et bandes de confiance (UEMOA)

Il ressort de cette illustration que les deux fonctions de survie sont très proches, ce qui est logique au regard de 2.3.1.

3. Mesure du risque d'estimation sur les taux ajustés

Cette section présente les mesures du risque systématique lié aux fluctuations d'échantillonnage à travers les taux de décès ajustés (il s'agit ainsi du risque d'estimation). À cet effet, 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 $y_x = \ln\left(q_x / (1 - q_x)\right)$, $z_x = \ln\left(q_x^{ref} / (1 - q_x^{ref})\right)$ et $\varepsilon_x \sim N(0; \sigma_\varepsilon)$. On note par ailleurs $\theta = (a, b)$.

Cette étude présente 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, c'est-à-dire que l'on fait l'hypothèse que le modèle est correctement spécifié. Cette hypothèse revient à considérer que les erreurs ε_x présentés *supra* ne correspondent pas à l'erreur de spécification du modèle, mais à l'erreur autour de $y_x = \ln\left(q_x(\hat{\theta}) / (1 - q_x(\hat{\theta}))\right) = a \times z_x + \hat{b}$ au titre des fluctuations d'échantillonnage (la mesure du risque d'estimation présentée dans la sous-section 3.2 est largement basée sur cette interprétation).

On note par ailleurs que la détermination des paramètres $\theta = (a, b)$ est au cœur du dispositif de mesure du risque d'estimation (*cf.* les sous-sections 3.1 et 3.2). Une attention particulière doit donc être accordée à la méthode et à la qualité de l'estimation des paramètres. L'estimation par maximum de vraisemblance peut être biaisée, car la taille de l'échantillon est limitée à $n = x_M - x_m + 1$ observations. En revanche, l'estimation par les moindres carrés n'est pas biaisée, et c'est donc cette approche qui est retenue ici. On rappelle toutefois que l'estimation issue de cette approche n'est pas robuste (étant donnée que l'ajout ou la suppression d'observations dans les données peut modifier sensiblement l'estimation ; *cf.* Meucci [2007] pour une illustration des mesures de robustesse et une démonstration de l'absence de robustesse des estimations par moindres carrés) et qu'il convient ainsi d'analyser les résultats avec prudence.

Pour mémoire les taux bruts présentent des irrégularités directement liées aux erreurs d'échantillonnage (*cf.* section 2). Ainsi, les deux approches présentées ici pour mesurer le risque d'estimation (au titre des fluctuations d'échantillonnage pour mémoire) sur les taux de décès ajustés consistent à générer de manière aléatoire, directement ou indirectement, des taux bruts (suivant la loi adéquate), afin d'en déduire l'impact sur l'estimation des paramètres $\theta = (a, b)$ du modèle de Brass, puis l'impact sur les taux de décès ajustés. La première approche s'appuie sur la simulation directe des taux bruts (*cf.* 3.1) alors que dans la seconde, les « taux de décès bruts simulés » sont issus de la simulation des résidus du modèle (*cf.* 3.2). Dans les deux cas, la technique

de simulation retenue consiste à considérer une méthode de Monte-Carlo pour simuler la distribution d'une loi normale (à partir de la simulation d'une distribution de loi normale centrée et réduite, elle-même déduite d'une simulation de distribution de loi uniforme entre 0 et 1).

3.1. Méthode 1 : simulation directe des taux bruts

On se place dans le cas des taux bruts estimés selon l'approche de Hoem. La première étape consiste alors à générer k simulations ($k \in [1, K]$) des taux de décès bruts, pour tous les âges x ($x \in [x_m, x_M]$), selon

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

L'étape suivante consiste à calculer pour chaque simulation k une estimation des paramètres $\theta^k = (a^k, b^k)$. Pour mémoire, le modèle est de la forme $y_x^k = a^k z_x + b^k + e_x^k$ (où $y_x^k = \ln \left(\frac{q_x^k}{1-q_x^k} \right)$ et q_x^k est une réalisation de Q_x). En retenant un ajustement par la méthode des moindres carrés ordinaires et en notant $n = x_M - x_m + 1$, on a ainsi (cf. par exemple

$$\text{Bourbonnais [2004]} : a^k = \frac{\sum_{x=x_m}^{x_M} y_x^k \times z_x - n \times \bar{y} \times \bar{z}}{\sum_{x=x_m}^{x_M} z_x^2 - n \times \bar{z}^2} \text{ et } \hat{b}^k = \bar{y} - a^k \times \bar{z}.$$

Pour la troisième étape, on note que $\ln \left(\frac{q_x(\hat{\theta}^k)}{1-q_x(\hat{\theta}^k)} \right) = y_x^k = a^k z_x + \hat{b}^k$ et on en déduit k réalisations (au titre des fluctuations d'échantillonnage) 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 $q_x(\hat{\theta}) = \frac{\exp(a \times z_x + \hat{b})}{1 + \exp(a \times z_x + \hat{b})}$ correspond au taux de décès ajusté à partir de l'estimation du taux de décès brut observé à l'âge x (on parle

alors de taux ajusté) et $\psi_x = \sqrt{E \left[\left(q_x(\hat{\theta}^k) - q_x(\hat{\theta}) \right)^2 \right]}$. Ce coefficient

constitue une mesure de dispersion, en pourcentage et au titre du risque d'estimation, des taux de décès simulés autour du taux de décès ajusté. On peut noter au passage qu'on utilise ici comme taux de décès de « base » le taux obtenu avec la valeur *best estimate* du paramètre, qui ne coïncide pas avec sa vision *best estimate* du fait de la non linéarité de la transformation qui lie ces deux grandeurs.

3.2. Méthode 2 : simulation des résidus

Conformément aux indications présentées en introduction de la section 3, les résidus du modèle correspondent dans notre étude à une estimation des erreurs au titre des fluctuations d'échantillonnage (pour mémoire, les résidus du modèle sont $e_x = y_x - a \times z_x - \hat{b}$, où $\hat{\theta} = (a, \hat{b})$ sont les paramètres du modèle estimés par la méthode des moindres carrés ordinaires). Sur ces bases, cette seconde approche 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 –). Par souci de cohérence avec les travaux de la sous-section précédente, les taux de décès ajustés $q_x(\hat{\theta})$, desquels on déduit les « vrais » résidus observés du modèle de Brass, sont déterminés à partir des taux bruts estimés selon l'approche de Hoem (on note toutefois que les taux bruts estimés à partir de la fonction de survie de Kaplan Meier auraient pu être retenus ici).

Pour mesurer le risque d'estimation à partir de la simulation des résidus, la première étape consiste à tester l'adéquation de la distribution des « vrais » résidus observés du modèle à la loi normale. Dans le cas d'échantillons de petites tailles (c'est-à-dire avec $n = x_M - x_m + 1 \leq 50$, cf. Rakotomalala [2008]), ce qui peut être fréquent dans le cas de tables instantanées, le test de Shapiro-Wilk est particulièrement robuste (pour mémoire, ce test est basé sur le rapport entre une combinaison linéaire des étendues successives et l'écart-type). À mesure que le nombre d'observations retenues pour l'ajustement du modèle de référence augmente, on privilégiera le test de d'Agostino, qui devient particulièrement efficace à partir de $n \geq 20$, cf. Rakotomalala [2008] (pour mémoire, le test d'Agostino est basé sur les coefficients d'aplatissement et d'asymétrie).

Dès que la compatibilité à la loi normale est vérifiée, dans une deuxième étape on calcule empiriquement la moyenne et l'écart-type (sans biais) de la distribution des résidus, respectivement notés μ_ε et σ_ε (on a donc

$\mu_\varepsilon = \frac{1}{n} \sum_{i=x_m}^{x_M} e_i$ et $\sigma_\varepsilon = \sqrt{\frac{1}{n-1} \sum_{i=x_m}^{x_M} (e_i - \mu_\varepsilon)^2}$). On utilise alors ces paramètres

pour générer des erreurs aléatoires selon la loi normale, et on a donc pour différents scénarios k ($k \in [1, K]$): $y_x^k = a \times z_x + \hat{b} + e_x^k$, où e_x^k est une réalisation de $\hat{\varepsilon}_x^k$ et $\hat{\varepsilon}_x^k \sim N(\mu_\varepsilon; \sigma_\varepsilon)$. Les « taux de décès bruts » associés aux k scénarios sont alors ici :

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

Pour chaque scénario k , on calcule à partir de ces « taux bruts » une estimation des paramètres $\theta^k = (a^k, b^k)$ en retenant un ajustement par la méthode des moindres carrés ordinaires (à l'image de ce qui est présenté dans la sous-section 3.1). On en déduit alors k réalisations (au titre du risque d'estimation) des taux ajustés (on parle alors de taux simulés) :

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

À l'image de la sous-section 3.1, on calcule le coefficient $c(\psi_x) = \frac{\psi_x}{q_x(\hat{\theta})}$

(avec désormais $\psi_x = \sqrt{E \left[\left(q_x^k(\hat{\theta}^k) - q_x(\hat{\theta}) \right)^2 \right]}$), afin d'avoir une mesure de dispersion, en pourcentage et au titre du risque d'estimation, des taux de décès simulés autour du taux de décès ajusté.

3.3. Mesure sur les taux ajustés : illustrations

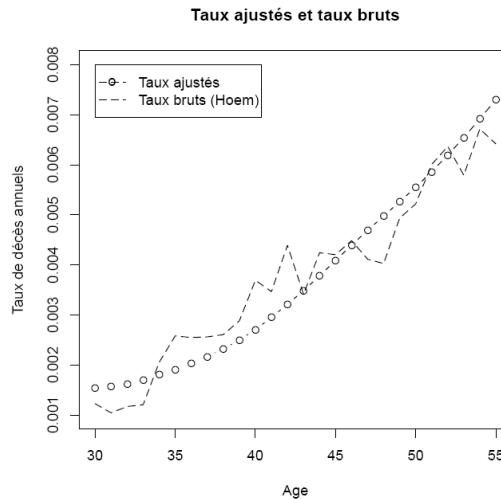
On présente ici une illustration et une comparaison des résultats obtenus selon les méthodes retenues pour la simulation des taux bruts au titre des fluctuations d'échantillonnage (cf. descriptions de ces méthodes en 3.1 et 3.2).

Dans l'ensemble des applications numériques, on s'appuie sur $K = 15\,000$ simulations des taux bruts et on considère que pour chaque simulation k les réalisations de la loi normale centrée réduite $N(0;1)$, desquels on déduit les réalisations appropriées pour chaque méthode, sont exactement les mêmes dans les deux méthodes. On note en outre que dans le cadre de la méthode décrite en 3.1, méthode qui pour mémoire se base sur la simulation directe des taux bruts, les taux bruts sont en pratique générés sous la contrainte $q_x^k > 0$ (lorsque cette condition n'est pas vérifiée pour une simulation k , celle-ci est annulée et reconduite pour les deux méthodes).

Par ailleurs, en pratique l'ajustement du modèle de Brass est réalisé en considérant la table réglementaire française TH00-02 (décès) comme table de référence. Cet ajustement conduit aux estimations suivantes : $a = 0,799$ et $\hat{b} = -1,082$ (avec des probabilités de rejeter à tort l'hypothèse H_0 de nullité du paramètre selon le test de Student respectivement égales à $4,0 \cdot 10^{-12}$ et à 0,7 %). Aussi, il ressort que le R^2 ajusté est égal à 86,5 %. Enfin, on note que la probabilité associée au test Shapiro-Wilk sur les résidus est égale à 55,8 % ; en d'autres termes, on accepte l'hypothèse H_0 de normalité des résidus observés.

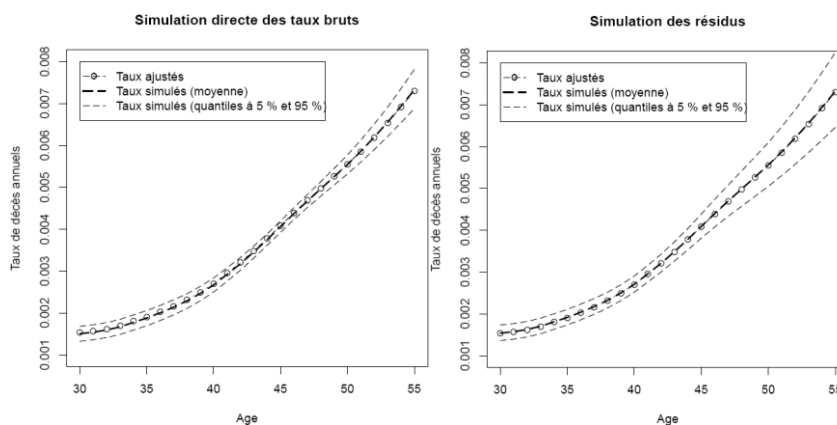
À titre d'information, le graphique suivant présente les taux ajustés et les taux bruts (estimés selon l'approche de Hoem).

Figure 3 - Taux de décès ajustés et taux de décès bruts (UEMOA)



3.3.1. Illustrations et comparaisons des taux de décès

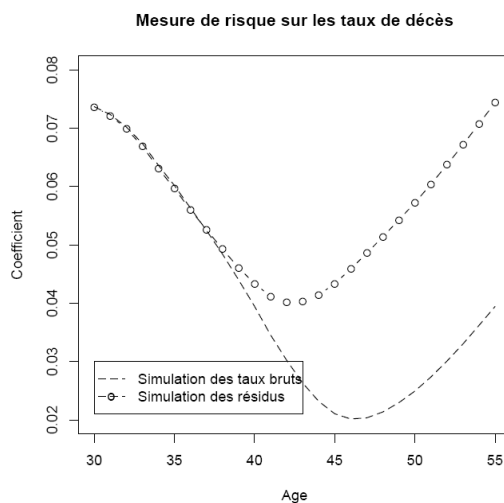
Les graphiques suivants présentent les taux de décès ajustés issus des estimations $\hat{\theta} = (a, \hat{b})$, ainsi que, pour les deux méthodes présentées, les taux de décès simulés au titre des fluctuations d'échantillonnage (dans ce cas, on présente les moyennes et les quantiles à 5 % et à 95 %).

Figure 4 - Taux de décès ajustés et simulés (UEMOA)

Il ressort que la dispersion des taux de décès simulés issus de la simulation des résidus est plus importante que celle des taux simulés issus de la simulation directe des taux bruts (pour préciser ces observations sur les taux de décès, les paragraphes ci-dessous présentent une comparaison des mesures de risques entre les deux méthodes retenues).

3.3.2. Illustrations et comparaisons des mesures de risques sur les taux

Pour quantifier le risque d'estimation selon la méthode de simulation retenue, on présente ci-dessous le coefficient $c(\psi_x)$, introduit dans les sous-sections 3.1 et 3.2 et représentant la dispersion en pourcentage des taux simulés (au titre du risque d'estimation dû aux fluctuations d'échantillonnage) autour des taux ajustés. Dans ce contexte, le graphique suivant présente les valeurs de ce coefficient par méthode et par âge.

Figure 5 - Risque d'estimation sur les taux (par âge) (UEMOA)

Avant de commenter ce graphique, on présente ci-après une moyenne arithmétique du coefficient $c(\psi_x)$ pour tous les âges $x \in [x_m, x_M]$ de l'étude et par méthode de simulation des taux bruts.

Tableau 1 - Risque d'estimation sur les taux (moyenne) (UEMOA)

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 (risque systématique)	$c(\psi) = 4,07 \%$
Méthode 2 : simulation des résidus (risque systématique)	$c(\psi) = 5,59 \%$

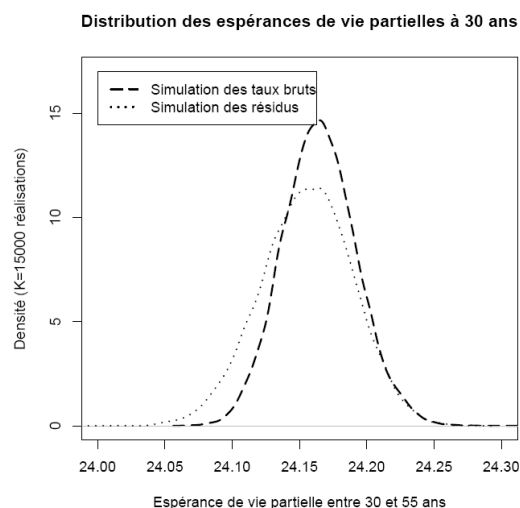
Il ressort de ces résultats que la mesure du risque d'estimation issue de la simulation des résidus majore en moyenne de plus de 35 % celle issue de la simulation directe des taux bruts.

Ces écarts s'expliquent par les hypothèses retenues. En effet, la méthode de simulation des résidus repose largement sur l'hypothèse d'absence d'erreur de spécification du modèle (ce qui revient à considérer que l'erreur du modèle est uniquement due aux fluctuations d'échantillonnage). Bien que cette hypothèse soit justifiée, elle ne reflète pas la réalité et en pratique, il existe une erreur de spécification du modèle. En d'autres termes, pour la méthode de simulation des résidus, les mesures de risque d'estimation présentées dans le graphique et le tableau ci-dessus incluent également une part de risque de modèle (d'où l'écart de mesure de risque observé entre les deux méthodes).

En particulier, et pour revenir à la Figure 5 on note que les valeurs de mesures de risques par âge présentées sont très liées à la structure du portefeuille d'expérience, au modèle et à son calibrage. En effet, d'une part dans le cas de la simulation directe des taux bruts, les mesures de risques sont directement liées à l'exposition au risque et à l'écart entre les taux bruts observés et les taux ajustés. D'autre part, dans le cas de la simulation des résidus, on note que les « taux bruts » simulés ne sont pas directement estimés mais sont issus d'une transformation non linéaire qui est donc de nature à introduire un biais (dans ce cas, les résultats sont donc à apprécier avec prudence).

3.3.3. Illustrations et comparaisons des tables associées aux taux de décès

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 (dans le cadre de l'assurance, le montant de la provision en est une autre, cf. section 4 ci-après). Dans ce contexte, 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.

Figure 6 - Distribution des espérances de vie partielles (UEMOA)

Dans la continuité des observations et des conclusions des paragraphes 3.3.1 et 3.3.2, il apparaît sur le graphique ci-dessus 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 directe 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. Le tableau ci-dessous confirme cette analyse.

Tableau 2 - Espérances de vie partielles (UEMOA)

Statistique	Avec taux ajustés	Avec taux simulés (méthode 1 : simulation directe des taux bruts)	Avec taux simulés (méthode 2 : simulation des résidus)
Moyenne	24,15668	24,16488	24,15523
Quantile à 0,5 %	Sans objet	24,09700	24,06460
Quantile à 5 %	Sans objet	24,12027	24,09785
Quantile à 95 %	Sans objet	24,21085	24,21022
Quantile à 99,5 %	Sans objet	24,23849	24,23823

Il apparaît effectivement que les distributions de l'espérance de vie partielles sont légèrement plus dispersées lorsque les taux sont simulés à partir de la simulation des résidus.

Par ailleurs, au niveau des moyennes des espérances de vie partielles, il apparaît dans les deux cas l'écart par rapport au calcul à partir des taux ajustés est inférieur à 0,05 %. Dans le cas présent on peut donc considérer que le risque d'estimation n'est pas significatif.

4. Mesure du risque d'estimation sur les provisions

Pour étudier le risque systématique au titre des fluctuations d'échantillonnage (risque d'estimation) sur les tables de mortalité, 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, avec $d \geq 1$) 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, cf. paragraphe 3.3.2). À cet effet, on travaille d'une part avec un montant d'engagement déterministe (dans ce cas, les taux de décès sont les seules sources d'aléas) et d'autre part avec un montant d'engagement aléatoire (ici, les sources d'aléas sont les taux de décès et les durées de vie restantes des assurés, représentant respectivement le risque systématique et le risque mutualisable).

4.1. Provisions déterministes

Lorsque les provisions sont déterministes, l'assureur estime en 0 les flux probables des prestations à payer en t pour un individu d'âge x par (pour la réalisation k des taux de décès simulés) :

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

où C désigne le montant de la prestation.

En faisant l'hypothèse de décès en milieu d'année, on en déduit que le montant de la provision pour la réalisation k est

$$L_0^k = \sum_{t=0}^{d-1} F_x^k(t) \times (1 + r_{t+1})^{-t-1/2}, \text{ où } r_t \text{ représente le taux d'actualisation des}$$

flux de l'échéance t .

L'impact du risque d'estimation sur les provisions peut alors être mesuré par le coefficient $c(\Upsilon) = \frac{\Upsilon}{L_0}$, où L_0 correspond à la provision calculée à partir

des taux de décès ajustés $q_x(\hat{\theta})$ et $\Upsilon = \sqrt{E\left[\left(L_0^k - L_0\right)^2\right]}$. Ce coefficient

permet d'avoir en pourcentage une mesure de dispersion (au titre du risque d'estimation) autour de la provision calculée à partir des taux de décès ajustés.

On note enfin $\bar{L}_0 = \frac{1}{K} \sum_{j=1}^K L_0^j$, la moyenne des provisions déterministes

calculées avec les taux de décès simulés.

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 ajustés (la méthode retenue ici pour quantifier le risque mutualisable est une méthode classique s'appuyant sur des méthodes de simulations ; lorsque de nombreux facteurs de risques sont à prendre en compte, cette méthode peut toutefois devenir délicate à intégrer et on peut dans ce cas retenir par exemple l'approche « semi-analytique » proposée dans Guibert [2010]). Aussi, avant de présenter les outils de mesure du risque d'estimation dans ce contexte (cf. 4.2.3), on présente la méthode de simulation des durées de vie restantes retenue (cf. 4.2.1) et la méthode de calcul des provisions stochastiques dans un cas simplifié (c'est-à-dire dans un cas où l'on considère dans un premier temps les taux de décès ajustés et non simulés, cf. 4.2.2).

4.2.1. Simulation des durées de vie restantes

La méthode retenue pour la simulation des durées de vie restantes dans le cas du modèle de Brass est inspirée de l'approche retenue dans Delwarde et Denuit [2006] dans le cas du modèle d'Heligman-Pollard.

On génère ainsi 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 . La variable T_x est ainsi un entier qui satisfait

$$T_x p_x \geq V > T_{x+1} p_x, \text{ soit } \prod_{j=0}^{T_x-1} p_{x+j} \geq V > \prod_{j=0}^{T_x} p_{x+j} \text{ (c'est-à-dire que } T_x \text{ est la}$$

plus grande durée entière précédant le moment du décès).

4.2.2. Provisions stochastiques et taux de décès ajustés

Dans un premier temps, on évalue les provisions stochastiques en considérant les taux de décès ajustés (il n'y a ainsi aucune illustration du risque d'estimation ici), notés pour mémoire $q_x(\hat{\theta})$ (les probabilités de survie utilisées pour simuler les durées de vie restantes sont donc $p_x(\hat{\theta}) = 1 - q_x(\hat{\theta})$). La démarche retenue ici est comparable à celle présentée dans Planchet et al. [2006].

On considère désormais que le flux aléatoire de prestation à payer en t est $F_x(t) = \begin{cases} C, & T_x = t < d \\ 0, & \text{sinon} \end{cases}$, en supposant $E[F_x(t)] = F_x(t)$. Le montant aléatoire de l'engagement en 0 s'écrit alors (par souci de comparabilité avec la provision déterministe, il est supposé que les décès sont observés en milieu d'année) :

$$\Lambda = F_x(t) \times \left(1 + r_{T_x+1}\right)^{-T_x - 1/2},$$

et on suppose que $E[\Lambda] = L_0$. 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 calculées sur ces bases (on a ainsi $L_0 = \bar{\lambda} = \frac{1}{G} \sum_{i=1}^G \lambda_i$).

4.2.3. Provisions stochastiques et taux de décès simulés

Ici, on considère deux sources d'aléas : une première liée aux taux de décès stochastiques Π (Π représentant les taux de décès $q_x(\hat{\theta}^k)$ pour une réalisation k) et une seconde liée à la loi conditionnelle des engagements stochastiques $\Lambda | \Pi$. Ici également, la démarche retenue est comparable à celle présentée dans Planchet et al. [2006].

Commençons par présenter la loi conditionnelle des engagements aléatoires. Les probabilités de survie utilisées pour simuler les durées de vie restantes, notées T_x^k pour une réalisation k des taux de décès, sont

$p_x(\hat{\theta}^k) = 1 - q_x(\hat{\theta}^k)$. En pratique, le flux aléatoire de prestation à payer en t

pour une simulation k est donc $F_x^k(t) = \begin{cases} C, & T_x^k = t < d \\ 0, & \text{sinon} \end{cases}$, en considérant

$E[F_x^k(t)] = F_x^k(t)$. Pour une simulation k donnée des taux de décès ajustés,

le montant aléatoire de l'engagement en 0 s'écrit alors :

$$\Lambda^k = F_x^k(t) \times \left(1 + r_{T_x^k+1}\right)^{-T_x^k - 1/2},$$

en supposant $E[\Lambda^k] = L_0^k$. Sur la base de g ($g \in [1, G]$) réalisations de T_x^k ,

on obtient alors des réalisations $\lambda_1^k, \lambda_2^k, \dots, \lambda_G^k$ de Λ^k . En outre, pour une

simulation k donnée des taux de décès, on note que la moyenne empirique de la provision sur l'ensemble des réalisations de T_x^k est $L_0^k = \bar{\lambda}^k = \frac{1}{G} \sum_{i=1}^G \lambda_i^k$.

En intégrant désormais la première source d'aléa, qui pour mémoire correspond aux fluctuations d'échantillonnage sur les taux de décès (et qui représente ainsi le risque d'estimation, non mutualisable), la provision stochastique moyenne issue des taux de décès simulés s'écrit

$$\bar{L}_0 = \frac{1}{K} \sum_{j=1}^K L_0^j = \frac{1}{K} \sum_{j=1}^K \bar{\lambda}^j = \frac{1}{K} \sum_{j=1}^K \left(\frac{1}{G} \sum_{i=1}^G \lambda_i^j \right).$$

4.3. Mesure sur les provisions : illustrations

Les applications numériques relatives aux mesures de risques d'estimation sur les provisions seront réalisées en considérant un contrat temporaire décès dans deux scénarios : un premier dans lequel la durée du contrat est $d = 5$, et un second dans lequel la durée est $d = 20$. Dans chaque scénario, on compare les résultats obtenus avec les provisions déterministes et stochastiques.

Dans les deux scénarios on considère un assuré d'âge $x = 31$, un capital assuré égal à $C = 1$ et la courbe d'actualisation retenue est la courbe de l'Institut des Actuaire au 31/12/2009².

Concernant le nombre de simulations, pour la provision stochastique calculée à partir des taux ajustés, on retient $G = 37,5 \cdot 10^6$ simulations des durées de vie restantes. Pour la provision stochastique calculée à partir des taux simulés (au titre du risque d'estimation), pour chaque simulation k des taux de décès on retient $G = 2\,500$ simulations des durées de vie restantes (pour mémoire, le nombre de simulations des taux de décès est $K = 15\,000$, ce qui conduit donc à un total de 37,5 millions de simulations).

4.3.1. Illustrations et comparaisons sur une garantie temporaire décès de 5 ans

Pour quantifier le risque d'estimation dans les différentes approches de calcul des provisions, on présente dans le tableau ci-dessous les résultats synthétiques des estimations et des simulations dans quatre scénarios : calcul de provisions déterministe avec les taux ajustés, calcul de provisions stochastiques avec les taux ajustés, calcul de provisions déterministes avec les taux simulés et calcul de provisions stochastiques avec les taux simulés.

² Courbe disponible sur le site de l'Institut des Actuaire : http://www.institutdesactuaire.com/gene/link.php?doc_link=../docs/2010003190313_I A20091231.xls.

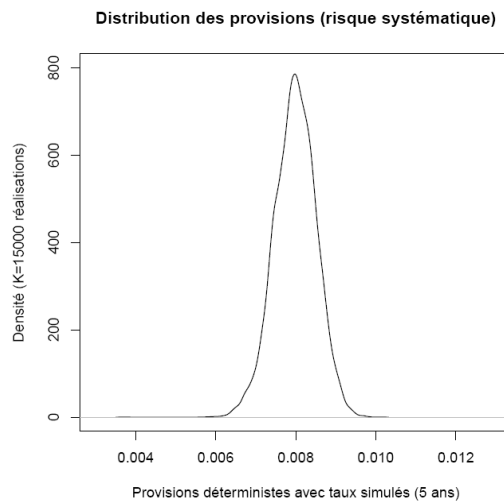
Tableau 3 - Risque d'estimation sur les provisions (5 ans) (UEMOA)

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 $c(Y) = 6,64 \%$.

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.

Figure 7 – Distribution des provisions (5 ans) (UEMOA)

Cette illustration permet de disposer d'éléments additionnels sur la mesure du risque d'estimation (risque systématique) sur les provisions.

4.3.2. Illustrations et comparaisons sur une garantie temporaire décès de 20 ans

On présente ci-après les résultats obtenus dans le cadre d'une provision d'une garantie temporaire décès de 20 ans (et non plus 5 ans comme dans le paragraphe précédent). Nous obtenons les résultats synthétiques présentés dans les tableaux suivants.

Tableau 4 - Risque d'estimation sur les provisions (20 ans) (UEMOA)

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	$L_0 = 3,83.10^{-2}$	$L_0 = 3,84.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$

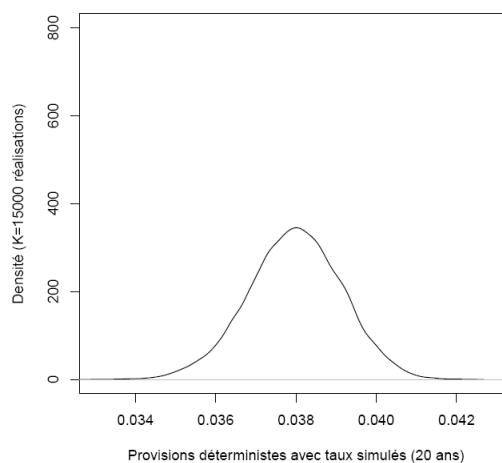
Il ressort lorsque la durée de l'engagement est de 20 ans que la prise en compte du risque systématique a un impact limité sur la provision calculée (baisse comprise entre 0,8 et 1,0 % lors de la prise en compte du risque systématique), et plus faible que celui observé lorsque la durée de l'engagement est de 5 ans.

Concernant la dispersion de la distribution de la provision déterministe calculée à partir des taux simulés, il apparaît que le coefficient est égal $c(\gamma) = 3,18\%$. 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.

En complément, on présente ci-après une estimation de la fonction de densité de la provision déterministe calculée à partir des taux simulés.

Figure 8 - Distribution des provisions (20 ans)

Distribution des provisions (risque systématique)



L'échelle en ordonnée de ce graphique est identique à celle de la Figure 7, et on note toutefois que le montant moyen de la provision est plus de quatre fois supérieur à celui associé à la Figure 7.

5. Impact de la réduction de l'échantillon sur le risque d'estimation

Dans cette section, on se replace dans le cadre d'une problématique de segmentation et on cherche à présenter l'évolution du risque d'estimation lorsqu'on réduit la taille de l'échantillon (au titre de la segmentation de la population). Ainsi, dans un premier temps on rappelle le contexte de l'étude (*cf.* sous-section 5.1), puis on présente l'évolution du risque d'estimation au niveau des taux de décès et au niveau des provisions (*cf.* respectivement sous-sections 5.2 et 5.3).

5.1. Rappel du contexte (hétérogénéité)

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

Tableau 5 - Statistiques ventilés par pays (UEMOA)

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%

Dans le cadre d'une modélisation de la mortalité d'expérience, il est nécessaire de tenir compte de l'hétérogénéité afin de limiter le biais associé lorsque la structure de la population évolue. En pratique ici, il s'agit ainsi de tenir compte des différences de mortalité entre pays. À cet effet, dans le cadre de cette étude on se limite à une démarche simple qui consiste à modéliser le comportement de chaque pays de manière indépendante.

L'objectif de cette section 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.

Dans les illustrations présentées ci-après, les méthodes statistiques utilisées sont celles présentées et utilisées dans les sections précédentes (en particulier, la méthode de simulation des taux bruts retenue est la simulation directe). En outre, toutes ces illustrations présentent une comparaison des résultats obtenus à partir des données du Togo et à partir celles de l'UEMOA (dans ce dernier cas, les illustrations sont alors identiques à celles présentés dans les sections précédentes).

5.2. Évolution du risque d'estimation sur les taux de décès ajustés

En pratique, il s'agit ici d'étudier l'évolution du risque d'estimation, c'est-à-dire du risque systématique lié aux fluctuations d'échantillonnage, sur les taux de décès et sur les espérances de vie partielles (entre 30 et 55 ans).

5.2.1. Comparaison des taux de décès uniquement

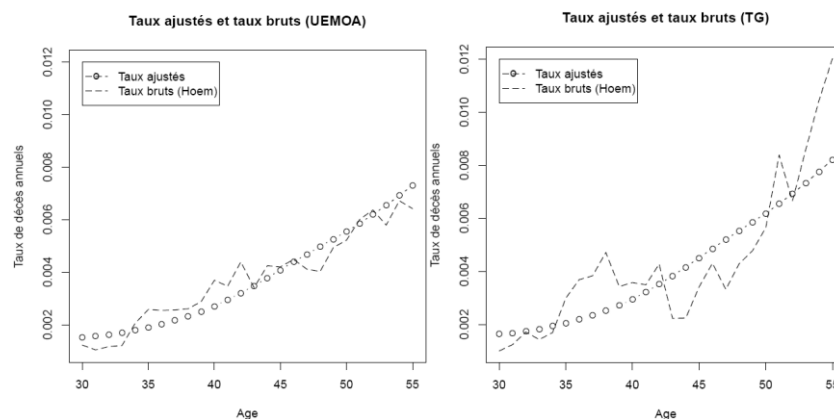
Dans un premier temps, on décrit les statistiques associées à l'ajustement des taux de décès.

Dans le cas des données de l'UEMOA, les statistiques sont présentées dans la section 3.3.

Dans le cas des données du Togo, l'ajustement conduit aux estimations suivantes : $a = 0,824$ et $\hat{b} = -0,841$ (avec des probabilités de rejeter à tort l'hypothèse H_0 de nullité du paramètre selon le test de Student respectivement égales à $8,6 \cdot 10^{-8}$ et à 19,9 %). Aussi, il ressort que le R^2 ajusté est égal à 69,2 %. Enfin, on note que la probabilité associée au test Shapiro-Wilk sur les résidus est égale à 71,7 % ; en d'autres termes, on accepte l'hypothèse H_0 de normalité des résidus observés. De manière générale, il apparaît que ces statistiques sont moins satisfaisantes que celles issues de l'ajustement avec les données de l'UEMOA.

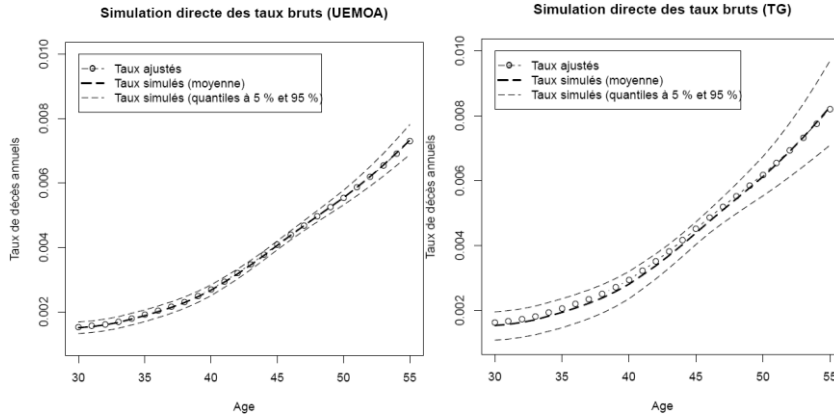
Sur ces bases, les graphiques ci-dessous présentent les taux ajustés et les taux bruts pour les données de l'UEMOA et du Togo.

Figure 9 - Taux de décès ajustés et taux de décès bruts (UEMOA et TG)



Il apparaît sur ces graphiques que les fluctuations d'échantillonnage sont logiquement plus importantes avec les données du Togo.

Pour confirmer ce premier résultat, les graphiques suivants présentent les taux de décès ajustés issus des estimations $\hat{\theta} = (a, \hat{b})$, ainsi que, pour les deux populations présentées, les taux de décès simulés au titre des fluctuations d'échantillonnage (dans ce cas, on présente les moyennes et les quantiles à 5 % et à 95 %).

Figure 10 - Taux de décès ajustés et simulés (UEMOA et TG)

Il ressort que la dispersion des taux de décès simulés à partir des données du Togo est plus importante que celle des taux simulés à partir des données de l'UEMOA.

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

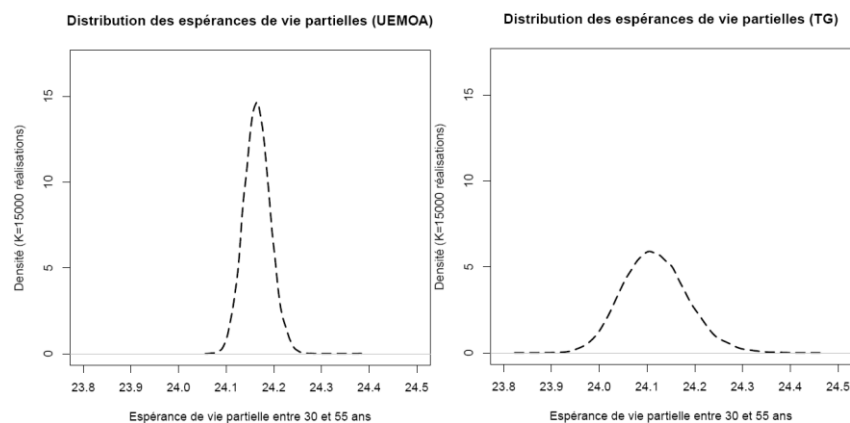
Tableau 6 - Risque d'estimation sur les taux (moyenne) (UEMOA et TG)

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

La moyenne du coefficient $c(\psi_x)$, qui pour mémoire représente la dispersion en pourcentage des taux simulés (au titre du risque d'estimation dû aux fluctuations d'échantillonnage) autour des taux ajustés, augmente de près de 150 %.

5.2.2. Comparaison des 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 on utilise les espérances de vie partielles. Dans ce contexte, le graphique suivant présente, pour chaque population, les estimations des fonctions de densité des espérances de vie résiduelles établies à partir des taux de décès simulés.

Figure 11 - Distribution des esp. de vie partielles (UEMOA et TG)

Dans la continuité des observations et des conclusions 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. En d'autres termes, les tables de mortalité issues des taux simulés à partir des données UEMOA présentent une volatilité plus faible que celles issues des taux simulés à partir des données du Togo.

Le tableau ci-dessous confirme et quantifie ces écarts.

Tableau 7 - Espérances de vie partielles (UEMOA et TG)

Statistique	UEMOA		TG	
	Avec taux ajustés	Avec taux simulés	Avec taux ajustés	Avec taux simulés
Moyenne	24,15668	24,16488	24,08086	24,11716
Quantile à 0,5 %	Sans objet	24,09700	Sans objet	23,95855
Quantile à 5 %	Sans objet	24,12027	Sans objet	24,01182
Quantile à 95 %	Sans objet	24,21085	Sans objet	24,23265
Quantile à 99,5 %	Sans objet	24,23849	Sans objet	24,31461

Il apparaît effectivement que les distributions de l'espérance de vie partielles sont légèrement plus dispersées lorsque les taux sont simulés à partir des données du Togo.

Par ailleurs, au niveau des moyennes des espérances de vie partielles, il apparaît que l'écart par rapport au calcul à partir des taux ajustés est égal à 0,15 % avec les données du Togo, contre un écart égal à 0,03 % avec les données de l'UEMOA.

5.3. Évolution du risque d'estimation sur les provisions

Cette sous-section s'attache à quantifier le risque d'estimation dans les provisions. À ce titre, on présente dans le tableau ci-dessous les résultats synthétiques des estimations et des simulations dans un scénario de calcul de provisions déterministes avec les taux ajustés d'une part, et avec les taux simulés d'autre part (le calcul de provisions stochastiques n'est donc pas repris dans ce tableau).

Tableau 8 - Risque d'estimation sur les provisions (20 ans) (UEMOA)

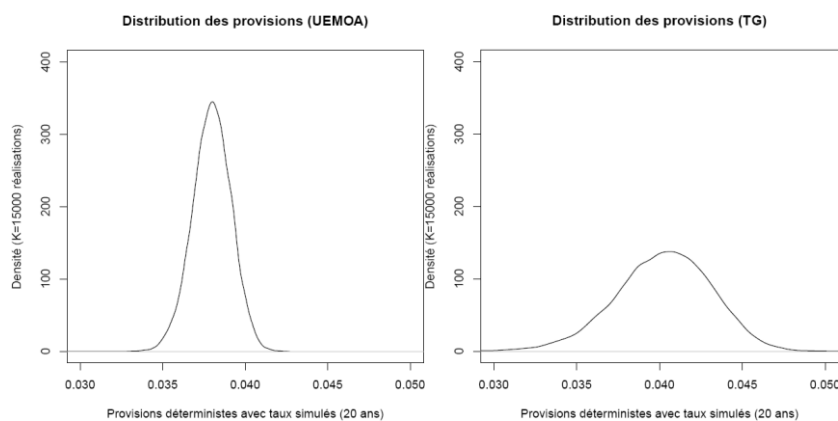
Statistique	UEMOA		TG	
	Provision déterministe et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)	Provision déterministe et taux ajustés	Provision déterministe et taux simulés (risque d'estimation)
Moyenne	$L_0 = 3,83.10^{-2}$	$\bar{L}_0 = 3,80.10^{-2}$	$L_0 = 4,18.10^{-2}$	$\bar{L}_0 = 4,02.10^{-2}$
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	$c(Y) = 3,18 \%$	Sans objet	$c(Y) = 7,91 \%$

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

Concernant les quantiles, il ressort qu'à partir des données du Togo les quantiles à 0,5 % et à 99,5 % présentent respectivement un écart de -21 % et 17 % par rapport à la provision moyenne, contre un écart respectif de -8 % et 7 % avec les données UEMOA.

Enfin, il apparaît que le coefficient $c(Y)$ relatif à la modélisation à partir des données du Togo majore de près de 150 % celui obtenu à partir des données de l'UEMOA.

En complément au Tableau 8, on présente ci-après une estimation de la fonction de densité de la provision déterministe calculée à partir des taux simulés à partir des données de l'UEMOA et du Togo.

Figure 12 - Distribution des provisions (20 ans) (UEMOA)

Ces graphiques constituent ainsi une forme d'illustration des troisième et cinquième colonnes du Tableau 8, relatives à la provision déterministe calculée à partir des taux simulés de l'UEMOA et du Togo. Ils confirment la présence d'un risque systématique plus important dans le cas d'une modélisation à partir des données du Togo.

6. Synthèse et conclusion

Dans le cadre de travaux sur le traitement de l'hétérogénéité, cette étude permet de présenter des mesures du risque d'estimation, c'est-à-dire du risque systématique associé à l'utilisation d'une loi de durée dont on sait qu'elle n'est pas bonne du fait des erreurs d'échantillonnage affectant sa construction. En particulier, il s'agit ici de présenter l'évolution du risque d'estimation en cas de baisse de l'échantillon d'expérience au titre d'une prise en compte de l'hétérogénéité à partir d'une modélisation indépendante de chaque sous-population. Cette analyse est réalisée en quatre étapes.

Les trois premières étapes introduisent et présentent les outils de mesure de risque d'estimation retenus. Ainsi, dans un premier temps les intervalles et bandes de confiance au titre du risque d'échantillonnage sont présentés pour les taux de décès bruts et pour la fonction de survie. Ces outils, présentés pour les estimateurs de Hoem (taux bruts) et de Kaplan-Meier (fonction de survie), constituent une première mesure du risque lié aux erreurs d'échantillonnage.

Dans un deuxième temps, les analyses portent sur les taux de décès ajustés à partir du modèle de Brass (modèle à référence externe). À ce niveau, deux approches ont été testées pour simuler l'impact des erreurs d'échantillonnage sur les taux de décès ajustés. La première porte sur la simulation directe des taux de décès bruts (desquels on déduit les paramètres du modèle de Brass pour chaque réalisation des taux bruts), et la seconde porte sur la simulation indirecte des taux bruts puisqu'elle consiste à simuler des résidus du modèle ajusté (de laquelle on déduit des réalisations de « taux de décès bruts », puis des paramètres du modèle de Brass pour chaque réalisation de ces « taux bruts »). Il apparaît ainsi que la seconde méthode inclut un risque lié à la spécification au modèle, au-delà du seul risque d'estimation lié aux fluctuations d'échantillonnage, qui se traduit par une mesure de risque plus importante, tant

au niveau des taux de décès qu'au niveau des tables de mortalité (dans ce dernier cas, l'espérance de vie partielle a été utilisée comme fonctionnelle). Cette seconde méthode présente toutefois l'avantage de ne s'appuyer sur aucune loi pour les taux bruts et peut ainsi se révéler particulièrement intéressante lorsque la loi des taux bruts n'est pas connue.

Dans un troisième temps, les analyses portent sur le calcul de provision (fonctionnelle naturelle dans le cas de l'assurance), dans le cadre d'une garantie temporaire décès. Cette partie permet de présenter certains résultats évidents, tels que l'évolution de la moyenne des provisions en cas de prise en compte du risque systématique (pour mémoire, la moyenne des provisions est en revanche inchangée en cas de prise en compte du risque mutualisable) ou la réduction (en pourcentage des provisions) du risque systématique lorsque la durée de couverture augmente.

Les trois étapes ci-dessus permettent de présenter les outils de mesures de risques. Toutes les applications numériques associées à ces étapes ont été réalisées à partir des données d'expérience de la zone UEMOA, qui incluent les données d'expérience de la Côte d'Ivoire, du Togo et du Mali. Dans la quatrième étape, on présente l'évolution du risque d'estimation lorsque l'échantillon diminue au titre d'une prise en compte de l'hétérogénéité à partir d'une modélisation indépendante de chaque pays. À cet effet, on considère les données d'un pays, en l'occurrence le Togo, dont l'exposition au risque représente 20 % de l'exposition de la zone UEMOA. Il apparaît alors que l'impact du risque d'estimation sur le *best estimate* est plus de 4 fois plus important avec les données du Togo qu'avec les données de l'UEMOA (il est égal à 3,8 % pour le Togo, contre 0,8 % pour la zone UEMOA).

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 et 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 donc être utile de construire un modèle intégrant l'hétérogénéité à partir de facteurs observables.

Notations des taux de décès

Les principales notations retenues ici pour les taux de décès annuels à l'âge x sont les suivantes :

- q_x : taux de décès réel théorique ;
- Q_x : variable aléatoire des taux de décès observés (taux bruts) ;
- q_x : réalisation des taux de décès observés (taux bruts) ;
- q_x^{ref} : taux de décès (ajustés) issus de la table de référence externe retenue ;
- $q_x(\hat{\theta})$: taux de décès ajustés à partir des taux observés q_x ;
- q_x^k : réalisation k des taux de décès bruts (réalisation à partir de la simulation directe de la variable aléatoire Q_x ou à partir de la simulation des résidus, *cf.* respectivement 3.1 et 3.2) ;
- $q_x(\hat{\theta}^k)$: taux de décès simulés à partir de la réalisation des taux bruts q_x^k .

Bibliographie

- Alho J. M. [2007] « Méthodes utilisées pour établir les projections relatives à la mortalité – Distributions prédictives de la mortalité future », *ISSA, Quinzième Conférence internationale des actuaires et statisticiens de la sécurité sociale*.
- Bourbonnais R. [2004] *Économétrie*, Dunod, 5^e édition.
- Delwarde A., Denuit M. [2006] « Construction de tables de mortalité périodiques et prospectives », *Economica*.
- Droesbeke J.-J., Fichet B., Tassi P. [1989] (éditeurs) « Analyse statistique des durées de vie », *Economica*.
- Guibert Q. [2010] *Analyse de la solvabilité d'un régime de retraite supplémentaire*, Mémoire d'actuariat, ISFA.
- Klein J. P., Moeschberger M. L. [2005] « Survival Analysis – Techniques for Censored and Truncated Data », *Springer, 2nd edition*.
- Martinussen T., Scheike T. H. [2006] « Dynamic Regression Models for Survival Data », *Springer*.
- Meucci A. [2007] « Risk and Asset Allocation », *Springer, Reprint of the 2007 edition*.
- Planchet F. [2009] « Provisionnement *best estimate* et risque arrêt de travail », *La Tribune de l'Assurance, n. 140*.
- Planchet F., Juillard M., Faucillon L. [2006] « Quantification du risque systématique de mortalité pour un régime de rentes en cours de service », *Assurance et gestion des risques, Vol. 75*.
- Planchet F., Kamaga A., Ziguélé M. [2010], « Confection des tables de mortalité réglementaires pour la zone CIMA », *CIMA/WINTER & Associés (confidentiel)*.
- Planchet F., Leroy G. [2009] « Quel niveau de segmentation pertinent ? », *La Tribune de l'Assurance, n. 142*.
- Planchet F., Théron P. [2006] « Modèles de Durée - Applications actuarielles », *Economica*.
- Rakotomalala R. [2008] « Tests de normalité – Techniques empiriques et tests statistiques », *Université Lumière Lyon 2, Version 2.0*.
- Saporta G. [2006] « Probabilités, analyses des données et statistique », *Technip, 2^{ème} édition*.