TABLES DE MORTALITE D'EXPERIENCE POUR DES PORTEFEUILLES DE RENTIERS

Note méthodologique

Frédéric PLANCHET

Version 1.0

Octobre 2005

SOMMAIRE

1. Introduction

Cette note a pour objet de présenter les méthodes utilisées par Daniel SERANT pour construire les tables de mortalité d'expérience à partir des données de marché collectées par l'Institut des Actuaires. Elle s'appuie pour l'essentiel sur le document de travail SERANT [2005] rédigé dans le cadre de la construction de tables prospectives sur des données INSEE et marginalement sur les autres références citées en bibliographie.

Afin d'alléger autant que possible la présentation, seules les formules essentielles à la compréhension des modèles sont reprises ici. Le détail des modèles et des algorithmes de résolution numérique est accessible dans les documents cités en référence (voir par exemple SERANT [2005] ou PLANCHET [2005] pour ces questions).

On rappelle que l'objectif de l'étude conduite par Daniel SERANT est de construire des tables de mortalité prospectives pour des portefeuilles de rentiers à partir des données de place transmises par les différents intervenants du marché. Les modèles proposés se doivent donc d'intégrer non seulement l'ajustement des taux de mortalité passés, mais aussi la projection de leur évolution dans les années à venir. Cette projection est abordée d'un point de vue inférentiel, au sens où le modèle doit détecter une éventuelle tendance passée pour l'extrapoler dans le futur. Des approches plus « subjectives » tentant d'intégrer explicitement des facteurs tels que les progrès de la médecine par exemple existent dans la littérature (voir XXX []), elles ne seront pas utilisées ici.

On peut également noter que le besoin de projection des taux de mortalité impose de fait l'utilisation de modèles paramétriques ou semi-paramétriques.

En pratique plusieurs méthodes de construction sont proposées et confrontées pour obtenir les tables prospectives d'expérience. On peut distinguer deux classes de modèles :

- ✓ *Les modèles de construction « intrinsèque »* : la démarche est ici d'exploiter l'information contenue dans les taux bruts de mortalité pour ajuster un modèle représentant correctement les données et permettant une projection « réaliste ». Les effectifs relativement restreints à disposition peuvent toutefois nuire à l'identification de la dérive de mortalité dans quelle proportion délicate à quantifier (sur les effets induits par des échantillons de taille réduite dans des modèles classiques de tables prospectives on pourra consulter LELIEUR [2005]).

- ✓ *Les modèles utilisant une référence externe de mortalité* : Une solution pour pallier les difficultés associées à des échantillons de taille réduite est de positionner la mortalité du groupe considéré par rapport à une référence externe. Disposant d'un ensemble de tables de moments INSEE (féminine et masculine) historiques et prospectives l'idée est alors de « référencer » les tables du moment d'expérience dans cet ensemble de tables.

2. La nature des données utilisées

Ces éléments sont repris de FFSA [2005].

2.1. Le périmètre concerné

L'étude porte sur les personnes assurées pour un contrat de rentes viagères. Deux sous populations sont à distinguer afin de pouvoir évaluer l'anti-sélection :

- ✓ les contrats collectifs et individuels dont la sortie en rente est obligatoire (ex : article 83, Madelin, PERP,….), y compris les rentes d'éducation et les rentes de conjoint survivant ;
- ✓ les contrats d'assurance vie dont la sortie en rente viagère est facultative y compris les contrats de rentes viagères immédiates achetées au moyen d'une prime unique.

2.2. Nature des données disponibles

Les informations demandées aux assureurs sont les suivantes :

- ✓ Le sexe (1 : masculin, 2 : féminin) ;
- ✓ La date de naissance (MM, AAAA) ;
- ✓ La date d'adhésion au contrat (JJ, MM, AAAA) (donnée facultative) ;
- ✓ La date de liquidation de la rente (JJ, MM, AAAA);
- ✓ La date de sortie du risque (décès ou autre motif de sortie) (JJ, MM, AAAA) ;
- ✓ Motif de sortie (1 : décès, 0 : autre motif) ;
- ✓ L'option de sortie (1 : sortie en rente obligatoire, 0 : sortie en rente facultative) ;
- ✓ La nature du contrat (1 : individuel, 0 : collectif) ;
- ✓ Le montant annuel de la rente (donnée facultative).

Il est également demandé de préciser :

- ✓ la date de début d'observation (date de constitution de la base de gestion ou de la dernière « expurge », par exemple 1/1/1994) ;
- ✓ la date de fin d'observation (par exemple le 31/12/2004 ou date ultérieure).

Enfin, ces données sont demandées sur une période d'observation suffisamment longue pour permettre l'établissement de résultats significatifs, c'est-à-dire sur une dizaine d'années d'observations.

2.3. Le calcul des taux bruts de mortalité

Le calcul des taux bruts de mortalité constitue une étape déterminante pour la qualité des études d'ajustement. Il est considéré dans la présente note que les taux bruts sont communiqués directement et sont donc une donnée dans le cadre de l'étude dont la méthodologie est évoquée ici.

Le passage des quotients de mortalité bruts au taux instantané de mortalité, qui est parfois la variable modélisée, s'effectue via une hypothèse sur la répartition des décès dans l'année; dans le cas où l'on fait l'hypothèse de constance du taux instantané dans chaque carré du diagramme de Lexis[1], on obtient l'estimateur suivant :

$$\mu_{xt}^{*} = -\mathbf{ln}\left(1 - \hat{q}_{xt}\right)$$

[1] Voir PLANCHET [2005] pour la définition.

3. Les modèles étudiés

3.1. Les modèles intrinsèques

3.1.1. Le modèle de Lee-Carter

Ce modèle proposé initialement par LEE et CARTER [1992] dans le cadre de la détermination de tables de mortalité prospectives à l'échelle d'un pays a connu un vif succès et de nombreuses variantes (on pourra par exemple consulter BROUHNS et DENUIT [2002a] et [2002b], BROUHNS et al. [2002], SITHOLE et al. [2000] et CZADO et al. [2004]).

Il s'agit d'une méthode d'extrapolation des tendances passées initialement utilisée sur des données américaines, qui est devenue rapidement un standard. La modélisation retenue pour le taux instantané de mortalité est la suivante :

$$\mathbf{ln}\,\mu_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt},$$

avec les variables aléatoires ε_{xt} iid distribuées selon une loi $N(0,\sigma^2)$; l'idée du modèle est donc d'ajuster à la série (doublement indicée par *x* et *t*) des logarithmes des taux instantanés de décès une structure paramétrique (déterministe) à laquelle s'ajoute un phénomène aléatoire ; le critère d'optimisation retenu va consister à maximiser la variance expliquée par le modèle, ce qui revient à minimiser la variance des erreurs.

Les coefficients du modèle peuvent s'interpréter de la manière suivante : α_x s'interprète comme la valeur moyenne des $\mathbf{ln}(\mu_{xt})$ au cours du temps. On vérifie que $\frac{d\,\mathbf{ln}(\mu_{xt})}{dt} = \beta_x \frac{dk_t}{dt}$ et on en déduit que le coefficient β_x traduit la sensibilité de la mortalité instantanée à l'âge *x* par rapport à l'évolution générale k_t, au sens où $\frac{d\,\mathbf{ln}(\mu_{xt})}{dk_t} = \beta_x$. En particulier, le modèle de Lee-Carter suppose la constance au cours du temps de cette sensibilité. Cette contrainte du modèle peut apparaître relativement forte :

- ✓ Pour tout age *x* les quotients des variations relatives des taux de mortalité à des dates différentes ne dépendent pas de l'age *x*. Si la variation relative du taux de mortalité à 50 ans en 2000 était 80% de ce quelle était en 1990 ce coefficient de 80% est retenu pour tous les ages ;
- ✓ Pour une même date *t* les quotients des variations relatives des taux de mortalité à des âges différents ne dépendent pas de la date *t*. Si en 2000 la variation relative du taux de mortalité à 20 ans est 50% de la variation relative du taux à 50 ans ce coefficient de 50% s'appliquera à toute date future ou passée.

Enfin, on peut remarquer que la forme du modèle implique l'homoscédasticité des taux de mortalité, ce qui est manifestement faux en pratique.

Après l'ajout de contraintes d'identifiabilité au modèle, l'estimation des paramètres nécessite plusieurs étapes qui, si elles sont assez fastidieuses, ne posent pas de problème majeur. Le nombre de paramètres à estimer, égal à $2(x_1-x_0+1)+(a_1-a_0+1)-2$ est important.

Il reste alors pour extrapoler les tendances à modéliser la série (k_t) ; pour cela on utilise en général un modèle ARIMA[2], mais toute modélisation de série temporelle peut être utilisée. La modélisation la plus simple que l'on puisse imaginer est par exemple une régression linéaire de ces coefficients :

$$k_t = at + b + \varepsilon_t$$

avec (ε_t) un bruit blanc gaussien. Empiriquement, on constate que la tendance sur les k_t est effectivement linéaire.

Dans le choix d'un modèle susceptible de structurer un jeu de données historiques, la « flexibilité » du modèle et par la même sa fidélité aux données est directement liée aux nombres de paramètres introduits. Le choix d'un modèle très flexible se fait le plus souvent au détriment des qualités prédictives de celui-ci (un modèle totalement non paramétrique n'autorise aucune prédiction).

Le modèle de Lee Carter peut de ce fait paraître très paramétré. Au surplus, dans le contexte de données de portefeuilles, dont le volume est sensiblement inférieur à ce que l'on peut obtenir comme taille de population sous risque à l'échelle d'un pays, le nombre élevé de paramètres du modèle peut conduire à des irrégularités conséquences de fluctuations d'échantillonnage. Ce phénomène est mis en évidence dans LELIEUR [2005].

Dans ce contexte il a semblé utile de confronter ce modèle à des modèles alternatifs moins paramétrés mettant en jeu des expressions analytiques portant sur les âges ou sur les années (ou les deux).

Par ailleurs les influences de l'age *x* et de l'année *t* sur les taux de mortalité $q_x(t)$ sont exprimés via l'introduction du « logit » $\mathbf{lg}_x(t) = \mathbf{ln}(q_{xt}/(1 - q_{xt}))$. Le logit pour des taux de mortalités faibles est peu différent de la variable $\mathbf{ln}(\mu_{xt})$ du modèle de Lee-Carter mais il peut être sensiblement différent pour des ages élevés. Il présente l'avantage de varier dans $]-\infty,+\infty[$, ce qui simplifie la mise en œuvre de modèle de régression.

On est ainsi conduit à introduire les modèles log-linéaires.

3.1.2. Les modèles log-linéaires

Cette famille de modèle est proposée dans SERANT [2005].

Le modèle de base de cette famille (modèle *npl*) impose une tendance linéaire en fonction du temps :

$$\mathbf{lg}_x(t) = \alpha_x + \beta_x \times t + \varepsilon_{xt}$$

[2] En suivant la démarche de Box et Jenkins.

On suppose les résidus iid et homoscédastiques, ce qui permet d'utiliser les résultats standards du modèle linéaire ordinaire. On dispose ainsi d'expressions explicites pour les paramètres. Ce modèle est en particulier utilisé pour la construction des tables TPG 1993.

On constate empiriquement une très forte corrélation entre les séries (α_x) et (β_x), ce qui conduit à proposer une variante du modèle dans laquelle ces deux coefficients sont liés par une relation affine ; cela conduit au modèle suivant (modèle *pl*) :

$$\mathbf{lg}_x(t) = a\beta_x + b + \beta_x \times t + \varepsilon_{xt} = (a+t)\,\beta_x + b + \varepsilon_{xt}$$

Le nombre de paramètres à estimer diminue sensiblement pour s'établir à $2+(x_1-x_0+1)$ au lieu de $2(x_1-x_0+1)$ dans le modèle précédent et de $2(x_1-x_0+1)+(a_1-a_0+1)-2$ dans le modèle de Lee-Carter. Cependant le problème de moindres carrés devient non linéaire, ce qui complique un peu l'estimation des paramètres. En pratique on doit avoir recours à des méthodes numériques alors que dans la première version du modèle on dispose d'une expression explicite directe des paramètres.

La dérive linéaire peut apparaître irréaliste sur le long terme, et on constate par exemple sur des données américaines un ralentissement de la tendance. On peut alors chercher des modélisations permettant d'introduire au niveau des prévisions de très long terme des informations exogènes traduisant un ralentissement prévisible de la dérive. Ceci peut être réalisé avec les modèles suivants :

$$\mathbf{lg}_x(t) = \alpha_x + \beta_x \times t + \gamma_x \times t^{\delta} + \varepsilon_{xt}$$

(modèle *np2* si $\delta = 2$ et *nprac* si $\delta = 1/2$). Dans ces modèles les estimations font également apparaître une très forte corrélation entre les estimations des paramètres (α_x), (β_x) et (γ_x), ce qui incite à proposer deux nouveaux modèles en posant $\alpha_x = a\gamma_x + b$ et $\beta_x = c\gamma_x + d$ et conduit à la spécification :

$$\mathbf{lg}_x(t) = b + d\ t + \gamma_x \left(t^{\delta} + c\ t + a\right) + \varepsilon_{xt}$$

(on désignera ces modèles par *p2* si $\delta = 2$ et *prac* si $\delta = 1/2$)

La résolution numérique du critère de moindres carrés associé n'appelle pas de commentaire particulier.

3.1.3. Utilisation des séries chronologiques

Le modèle de Lee-Carter, après avoir ajusté sur les données historiques les paramètres α, β et *k*, propose de considérer la suite des k_t comme une série chronologique pour obtenir les valeurs prospectives des taux. On est ainsi amené à poser :

$$k_t = at + b + \varepsilon_t$$

Cette approche peut être transposée dans le cadre des modèles logistiques, dans le but de réduire le nombre de paramètres. On cherche alors à paramétrer la fonction $x \rightarrow \mathbf{lg}_x(t)$ pour prendre en compte l'influence de l'année *t* de manière non paramétrique, puis à modéliser dans un second temps les séries chronologiques introduites. On considère ainsi un modèle de la forme :

$$\mathbf{lg}_x(t) = f(x, \theta_t) + \varepsilon_{xt}$$

où la fonction $f(x, \theta_t)$ est choisie, pour des arguments de simplicité de mise en œuvre, linéaire par rapport au paramètre (vectoriel) θ_t. Dans une deuxième phase la série (θ_t) est modélisée.

La forme retenue pour *f* est celle d'une spline cubique avec des nœuds aux ages 20 ans, 28 ans, 40 ans , 80 ans et 90 ans. La forme de la fonction *f* avec *p* nœuds est alors la suivante (modèle *splnp*) :

$$f(x, a, b, c, d, e_1, ..e_p) = a_t + b_t x + c_t x^2 + d_t x^3 + \sum_{i=1}^{p} e_{i,t}\left[(x - x_i)^+\right]^3$$

En pratique, une version simplifiée de ce modèle dans laquelle seul le paramètre a_t dépend du temps fournit des résultats fiables. En observant que la modélisation de a_t au travers d'une régression linéaire analogue à celle menée pour k_t, on peut construire une version entièrement paramétrique du modèle en proposant (modèle *splp*) :

$$f(x, a, b, c, d, e_1, ..e_p) = a + \lambda t + bx + cx^2 + dx^3 + \sum_{i=1}^{p} e_i\left[(x - x_i)^+\right]^3$$

3.2. Les modèles à référence externe

Si on ne dispose pas de données suffisantes pour structurer correctement la table complète, on peut imaginer d'utiliser la structure d'une table de référence existante et de simplement positionner la mortalité du groupe considéré par rapport à cette référence.

3.2.1. Régression logistique

Deux approches sont envisageables. Lorsque l'on souhaite positionner une table par rapport à une autre, il peut apparaître naturel d'effectuer la régression des logits des taux bruts sur les logits de la table de référence, ce qui conduit au modèle suivant :

$$\mathbf{ln}\left(q_{xt} / (1 - q_{xt})\right) = a_t \,\mathbf{ln}\left(q_{xt}^{ref} / \left(1 - q_{xt}^{ref}\right)\right) + b_t + \varepsilon_{xt}$$

ou encore :

$$\mathbf{lg}_x(t) = a_t\, \mathbf{lg}_x^{ref}(t) + b_t + \varepsilon_{xt}$$

La mise en œuvre de cette approche si l'on retient un critère de type « moindres carrés » est très simple, puisqu'il s'agit d'une régression linéaire dans le cadre d'un modèle linéaire ordinaire. On dispose donc d'une expression explicite des paramètres *a* et *b*.

Elle permet au surplus une extrapolation aisée des logits des taux d'expérience dans les plages d'âge pour lesquelles les données d'expérience seraient insuffisantes.

On peut adapter le critère d'optimisation utilisé pour tenir compte du contexte d'utilisation des tables en retenant plutôt :

$$(\hat{a}, \hat{b}) = \mathbf{arg\,min}[e_{60}^{liss\acute{e}}(a,b) - e_{60}^{nonliss\acute{e}}]$$

sous la contrainte suivante :

$$e_{60}^{liss\acute{e}}(a,b) - e_{60}^{nonliss\acute{e}} > 0$$

où $e_{60}^{liss\acute{e}}(a,b)$ désigne l'espérance de vie résiduelle à 60 ans, fonction des paramètres *a* et *b*, calculée à partir de la régression sur les logits ; $e_{60}^{nonliss\acute{e}}$ désignant l'espérance de vie résiduelle à 60 ans calculée à partir des données brutes.

Le détail de l'approche est présenté dans LELIEUR [2005].

3.2.2. Positionnement par rapport à une référence externe

On peut également rechercher, dans un ensemble de tables prospectives exogènes disponibles la période des tables de référence [an, an+h] la plus « proche » de la période [anexp, anexp+h] issu des données d'expériences. Cela conduit à utiliser comme tables d'expérience les tables exogènes décalées.

Remarque : la notion de « la plus proche » suppose l'utilisation d'une distance entre deux tables. Différentes approches sont possibles à ce niveau : Chi2 sur les $q_x(t)$, distance déduite des espérances résiduelles ou de leurs intégrale (qui représente à une unité monétaire près l'engagement d'un portefeuille de rentes ou tous les âges sont équi-représentés et à taux technique nul) l'avantage de cet indicateur est le « gommage » des fluctuations.

4. Les critères de validation du modèle

Les critères de validation de modèle fournissent des aides à la décision dans le cadre de la sélection du modèle le plus pertinent. La pertinence est ici appréciée en regard du contexte d'utilisation des tables proposées : l'évaluation des engagements au titre de rentes viagères.

Cela conduit notamment à porter une attention particulière à la représentation des espérances de vie résiduelles.

4.1. La fidélité aux données

La première des exigences que doit satisfaire un modèle est d'être fidèle aux données qui ont servi à le calibrer. Cette fidélité peut être examinée *a priori* de deux manières :

- ✓ Au travers des taux de mortalité $q_x(t)$, $(x,t) \in [x_0, x_1] \times [a_0, a_1]$;
- ✓ Au travers de l'espérance de survie résiduelle dans la plage $[x_0, x_1]$, définie par : $e_t(x, x_1) = E\left[Min(X - x, x_1 - x) \mid X > x\right]$

Ce second critère est motivé par le fait que l'utilisation des tables prospectives est principalement orientée vers les calculs des engagements des rentes viagères. Les espérances de survie résiduelles représentent les engagements associés au calcul des rentes avec un taux d'actualisation nul . L'audit des espérances conditionnelles est donc incontournable.

Les modèles les moins paramétrés sont en principe et en général les plus fidèles. Néanmoins cette logique statistique n'est pas toujours respectée le calibrage se faisant sur $\mathbf{lg}_x(t)$ ou sur $\mathbf{ln}(\mu_{xt})$ et non sur les éléments retenus pour apprécier la fidélité du modèle (taux de mortalité $q_x(t)$ et espérance résiduelle $e_t(x, x_1)$). On peut toutefois, et avant la mise en œuvre proprement dite, faire les remarques suivantes :

- ✓ Le modèle Lee-Carter peut conduire à sous-évaluer notablement les taux de mortalité des ages élevés (à partir de 85-90 ans). En effet, l'algorithme de référence construit sur une approche maximum de vraisemblance favorise les premiers âges (les plus « jeunes) et par ailleurs la relation $\mu_{xt} = -\mathbf{ln}(1 - q_{xt})$ repose que l'hypothèse de constance du taux instantané de décès entre deux âges entiers, hypothèse discutable aux âges élevés.
- ✓ On peut s'attendre à ce que les modèles les moins paramétrés épousent mieux les « irrégularités résiduelles » des données brutes ce qui constitue un handicap d'autant plus important que le volume de données est restreint.
- ✓ Au niveau des espérances résiduelles les irrégularités des tables brutes et des tables ajustées sont classiquement écrasées et ne ressortent que les dérives éventuelles et systématiques des modèles sur les q_x.

4.2. La stabilité des estimations

Le choix de la plage d'ages et la plage d'années à partir desquelles on doit générer les prévisions est important dans la mesure où les estimations des paramètres dépendent sensiblement de ce choix. En effet il est possible que ces différences, si elles existent, engendrent des prévisions différentes

En ce qui concerne les estimations concernant les âges le choix de la plage d'âges ne doit pas avoir d'incidence notable sur les estimations. Ainsi si l'on retient par exemple la plage [50 ans-70 ans] et la plage [60 ans-80 ans] les âges communs (de 60 à 70 ans) doivent avoir des estimations voisines.

Pour les modèles où les estimations se font âge par âge la nature même des modèles assure l'égalité des estimations quand on fait varier la plage utilisée. Pour les autres modèles qui tiennent compte conjointement de tous les âges constituant une plage on peut craindre que les estimations des paramètres dépendent (plus ou moins fortement) de la plage d'étude choisie. Toutefois, on peut espérer que cette instabilité ne perturbe pas les estimations des logits, taux de décès et espérances résiduelles compte tenu des qualités de fidélité des modèles.

Pour le modèle de Lee-Carter (le moins paramétré en âges) la contrainte d'identification $\sum \beta_x = 1$ a pour conséquence mécanique des différences entre les estimations limitées cependant à des translations.

4.3. La capacité prospective

On peut noter que, d'une manière générale, la capacité d'un modèle à une utilisation prospective est d'autant plus importante que le modèle est fortement paramétrique. Cette remarque conduit à privilégier les approches paramétriques.

5. Conclusion

Les modèles présentés dans cette note couvrent un large spectre de possibilité (caractère paramétrique plus ou moins affirmé, capacité prospective plus ou moins importante, etc.) ; un point décisif est toutefois éludé à ce stade, et devra être traité au moment de la construction, il s’agit de la question de la fermeture de la table : en pratique, sur des données de portefeuilles, les effectifs deviennent insuffisants à partir d’un certain âge et il convient de définir une méthode d’extrapolation aux grands âges. La méthode retenue a un impact important sur les résultats obtenus dans une problématique de rentes.

Ce point devra donc être traité avec soin, et une seconde version de la présente note décrira les méthodes de fermetures usuelles.

Bibliographie

BROUHNS N., DENUIT M. [2002a], « Risque de longévité et rentes viagères : II. Tables de mortalité prospectives pour la population belge », *Belgian Actuarial Bulletin*.

BROUHNS N., DENUIT M. [2002b], « Risque de longévité et rentes viagères : I. Evolution de la mortalité en Belgique de 1880 à nos jours », *Belgian Actuarial Bulletin*.

BROUHNS N., DENUIT M., VERMUNT J.K. [2002], « A Poisson log-bilinear regression approach to the construction of projected lifetables », *Insurance : Mathematics and Economics*.

CZADO C., DELWARDE A., DENUIT M. [2004], « Bayesian Poisson log-bilinear mortality projections »

DENUIT M., MAGIS C., WALHIN F. [2004], « La mortalité, un phénomène en pleine mutation: quelle solution pour le marché des rentes ? »

FFSA [2005] *Demande de données relatives aux populations d'assurés,* Document de travail FFSA.

LEE, R.D., CARTER, L. [1992] « Modelling and forecasting the time series of US mortality ». *Journal of the American Statistical Association* 87,659-671.

LELIEUR V. [2005] « Construction de tables de mortalité prospectives : le cas des petites populations ». ISFA, Mémoire d'actuariat.

PLANCHET F. [2005] *Modèles de durée : applications actuarielles,* Support de cours de l'ISFA.

SERANT D. [2005] « Construction de tables prospectives de mortalité », *Document interne FFSA.*

SITHOLE T., HABERMAN S., VERRALL R.J. [2000] « *An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners* », Insurance: Mathematics and Economics, vol. 27, pp. 285–312.