


28th INTERNATIONAL CONGRESS OF ACTUARIES

The international meeting of the actuarial profession

28e CONGRÈS INTERNATIONAL DES ACTUAIRES

Le rendez-vous international de la profession actuarielle

Utilisation des méthodes de Lee-Carter et Log-Poisson pour l'ajustement de tables de mortalité dans le cas de petits échantillons

Frédéric PLANCHET / Vincent LELIEUR

WINTER & ASSOCIÉS
I.S.F.A.
Sécurité sociale CNAV la Retraite solidaire — DEPUIS LONGTEMPS ET POUR LONGTEMPS

SOMMAIRE/ SUMMARY

Introduction

A - Contexte

B – Données disponibles

Construction de tables prospectives

A – Modèles de Lee-Carter et log-Poisson

B – Reparamétrisation

Provisionnement

A – Résultats numérique

B - Conclusion

Contexte

Dans le cadre de la mesure de l'engagement d'un régime de retraite professionnel, la mortalité du groupe concerné et son évolution dans les années à venir constituent des hypothèses déterminantes.

L'objectif de la présente étude est de construire des tables prospectives de mortalité dans ce contexte.

A l'échelle d'un tel régime les effectifs sont relativement importants, mais toutefois faibles en regard des contextes usuels de construction de tables prospectives.

Les modèles prospectifs usuels, dont le modèle de Lee-Carter constitue la référence, sont initialement construits pour de très grands groupes, qui sont souvent à l'échelle d'un pays.

Leur utilisation sur des groupes de taille plus modeste nécessite donc des adaptations.

Des versions ajustées de ces modèles, ainsi qu'un modèle alternatif, sont proposés ici.

Contexte

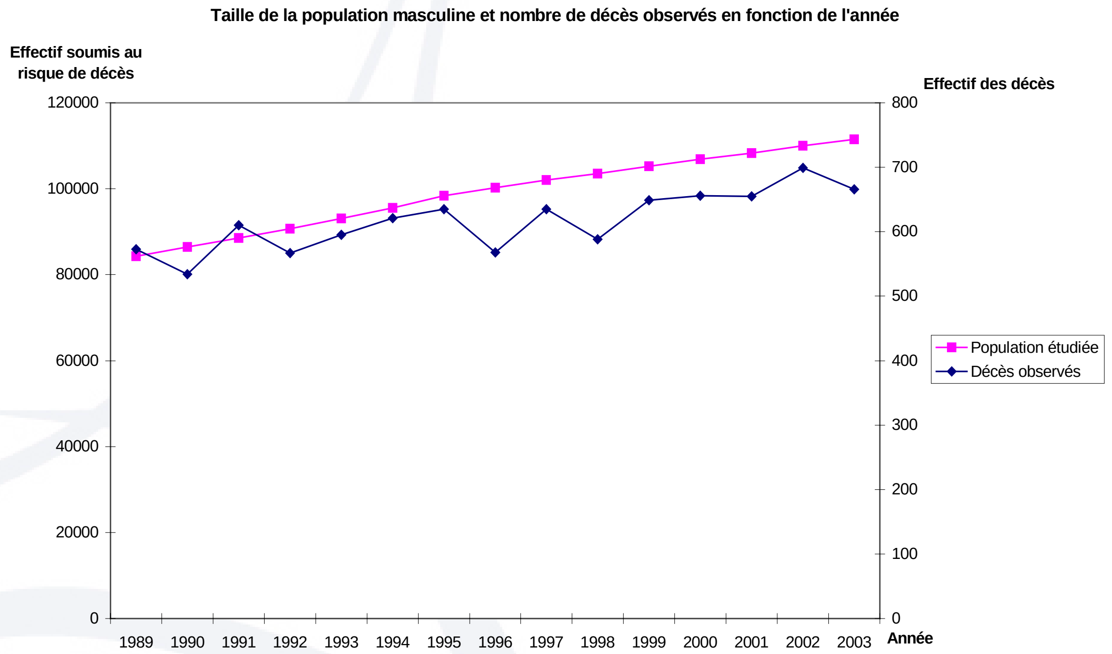
Deux approches sont envisageables, qui utilisent les données des années antérieures sur la mortalité du groupe pour construire un ajustement des taux de décès bruts passés, puis des projections de la mortalité future :

Les modèles de construction « intrinsèque » (ou endogène) : la démarche consiste ici à 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 une proportion délicate à quantifier (sur ces effets dans des modèles classiques de tables prospectives on pourra consulter Lelieur [2005]).

Les modèles utilisant une référence externe de mortalité (ou exogène) : 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, par exemple au moyen d'une régression logistique. Disposant d'un ensemble de tables de moments INSEE (féminine et masculine) historiques et prospectives l'idée est alors de « positionner » les tables du moment d'expérience par rapport à cet ensemble de tables.

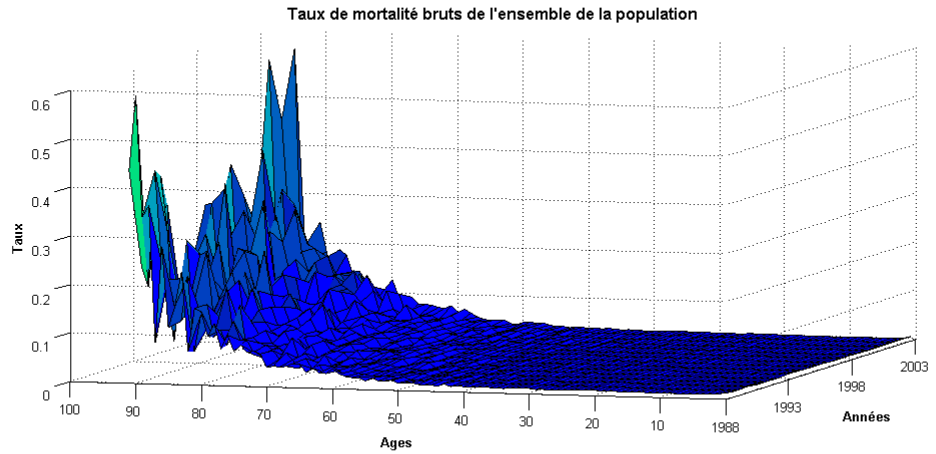
Notre attention se porte dans ce travail pour l'essentiel sur la première classe de modèles, même si une alternative relevant de la seconde approche est présentée.

Les données utilisées



On peut noter que la taille de la population étudiée est relativement restreinte, puisqu'elle avoisine les 100 000 personnes (avec une variation de plus ou moins 15%) sur chaque année de l'étude. Comme cet effectif est réparti en plus de 100 classes d'âges, l'effectif étudié pour une année et un âge considérés est en moyenne inférieur à 1000 personnes. Ce nombre relativement faible sera à l'origine de taux bruts de mortalité assez volatiles.

Taux de décès bruts par année



Les taux bruts sont erratiques, notamment aux grands âges

On observe une amélioration de la mortalité au cours de la période d'étude (1989-2003), ainsi par exemple à 52 ans, le taux brut de la mortalité annuelle correspondant à 1989 était approximativement de 1.1%, il n'était plus que de 0,8 % en 1996 et réduit à environ 0,5 % en 2003. Cela signifie en conséquence une diminution de plus de 50 % sur la période.

Pour d'autres âges, comme entre 25 et 30 ans, l'amélioration est moins nette, mais cela est à mettre en relation avec des données d'origine peu interprétables (en effet, l'échantillon étant petit, de l'ordre de quelques milliers, le nombre de décès à cet âge est de quelques unités, d'où des taux bruts 'peu représentatifs').

Modèles de Lee-Carter et log-Poisson

Ces modèles proposent une modélisation « bidimensionnel » du taux instantané de décès ;

- Lee-Carter : $\ln \mu_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt}$ $\qquad \left(\hat{\alpha}_x, \hat{\beta}_x, \hat{k}_t\right) = \arg\min \sum_{x,t} \left(\ln \mu^*_{xt} - \alpha_x - \beta_x k_t\right)^2$

- log-Poisson : $\mu_{xt} = \exp(\alpha_x + \beta_x k_t)$ $\qquad P(D_{xt} = d) = \frac{(L_{xt}\mu_{xt})^d}{d!} \exp(-L_{xt}\mu_{xt})$

La variante log-Poisson corrige certains défauts de l'approche de Lee-Carter :

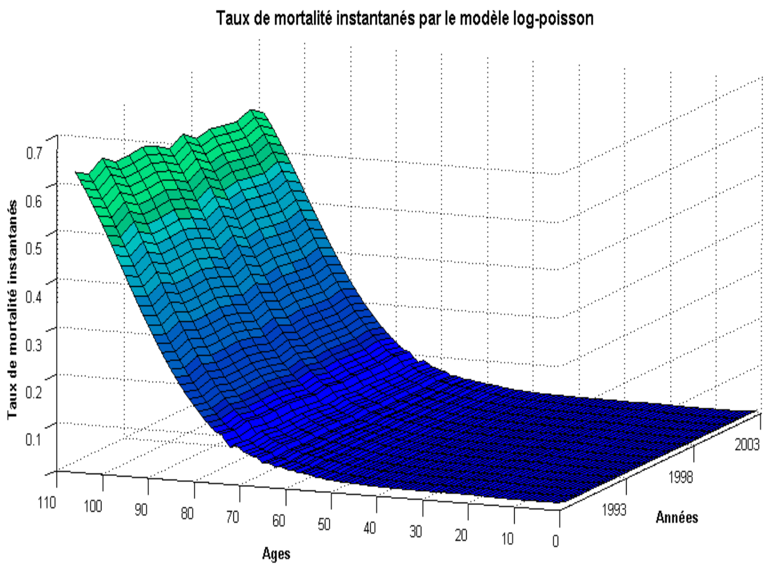
- homoscédasdicité
- critère de détermination des paramètres

Ces deux modèles nécessitent l'estimation d'un nombre important de paramètres :

$$2 \times (x_M - x_m + 1) + t_M - t_m - 1$$

Modèles de Lee-Carter et log-Poisson

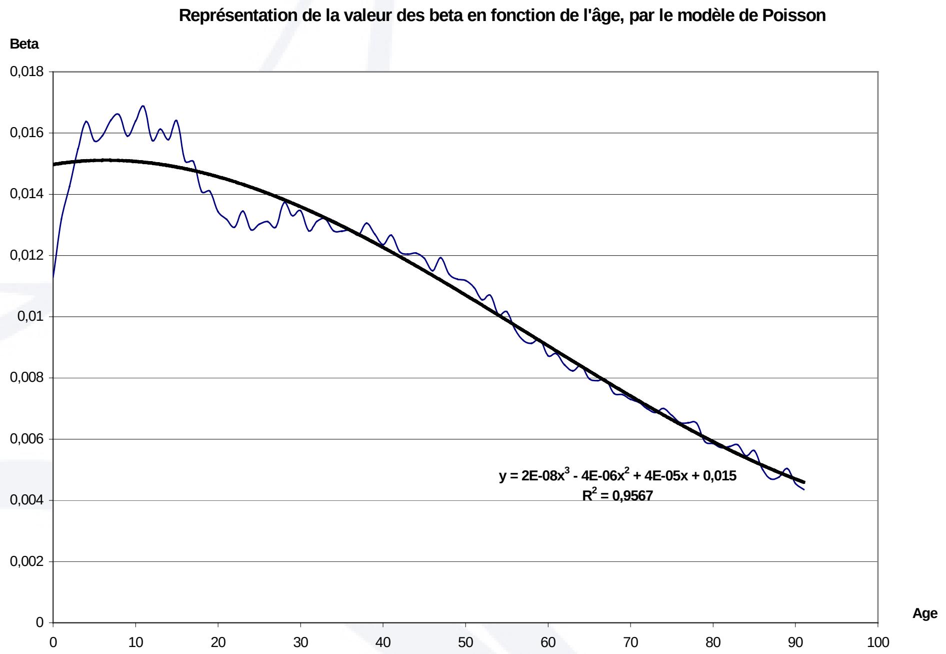
Dans le contexte de l'étude les effectifs sont relativement faibles et il subsiste des fluctuations d'échatillonnage, qui conduisent à des irrégularités sur les résultats :



La surface ci-dessus est irrégulière dans la dimension temporelle et dans la dimension âge.

Reparamétrisation

Cette irrégularité est la conséquence du caractère erratique des paramètres estimés ; l'idée est alors de diminuer le nombre de paramètres en reparamétrisant le modèle :

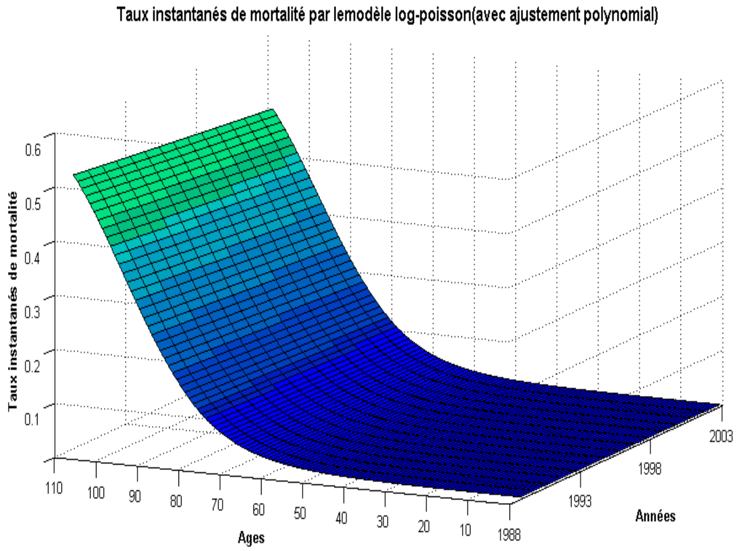


On retient ici une expression polynômiale d'ordre 3 :

$$\beta_x = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$

Reparamétrisation

On obtient ainsi une surface régulière (par construction) :



Propriétés

L'extrapolation des taux futurs est alors directe, *via* le prolongement de la tendance (linéaire) *k*. On notera incidemment que la poursuite d'une tendance linéaire sur le long terme peut apparaître irréaliste, mais on ne la remet pas en cause ici car elle est de nature à conduire à une majoration des engagements du régime, et est donc *a priori* une hypothèse prudente.

On obtient ainsi un modèle robuste, reposant sur une approche « maximum de vraisemblance » et avec un petit nombre de paramètres (10 = 4 + 4 + 2), ce qui constitue une situation *a priori* favorable pour conduire la démarche prospective (voir sur ce sujet Serant [2005] qui discute le lien entre paramétrisation et pouvoir prédictif d'un modèle). De plus l'approche par maximum de vraisemblance sur un nombre réduit de paramètres permet d'exploiter les propriétés asymptotiques de ces estimateurs (efficacité, convergence, tests de Chi-deux associés, etc.).

Enfin, on peut observer que l'on obtient ainsi une forme paramétrique simple du taux instantané de mortalité en fonction de *x* et de *t*, ce qui conduit à des formules fermées pour le calculs des provisions à chaque âge (et à des calculs valables aux âges non entiers).

$$a_{xt} = \frac{1}{L_{xt}} \int_0^{+\infty} L_{xt}(h) \mathbf{exp}(-rh)\, dh$$

$$L_{xt}(h) = \mathbf{exp}\left(-\int_0^h \mu(x+\theta, t+\theta)\, d\theta \right) \qquad \mu_{xt} = \mathbf{exp}(\alpha_x + \beta_x k_t)$$

Fermeture des tables

Les taux bruts au-delà de 90 ans sont inexploitables sur nos données ; il convient donc d'extrapoler les taux de décès au delà de cet âge. Deux méthodes ont été testées:

- la méthode de Coale-Kisker ;
- l'extrapolation des coefficients du modèle log-Poisson polynômiale estimé sur la plage d'âges restreinte.

La méthode de Coale-Kisker consiste à poser :

$$\hat{\mu}_x = \hat{\mu}_{65} \cdot e^{g_x (x-65)}$$

Puis à estimer le coefficient g avec :

$$g_x = g_{80} + s(x - 80)$$

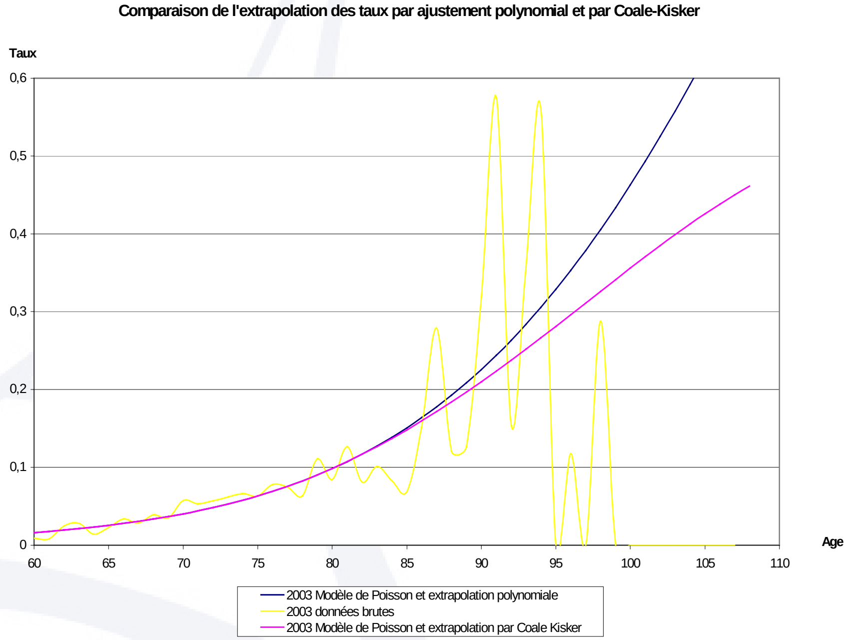
Ce qui conduit finalement à la forme suivante des taux de décès aux grands âges :

$$\hat{\mu}_x = \hat{\mu}_{x-1} \cdot e^{g_{80} + s(x-80)}$$

$$s = -\frac{\ln(\hat{\mu}_{79} + 31 . g_{80})}{465} \qquad g_{80} = \frac{\ln(\frac{\hat{\mu}_{80}}{\hat{\mu}_{65}})}{15}$$

Fermeture des tables

On obtient:



Il n'est pas aisé de déterminer la méthode d'extrapolation qui correspond le mieux à la réalité, puisqu'on a eu recours à l'extrapolation en raison du faible échantillonnage (impliquant des taux très variables) aux grands âges. Cependant, par soucis de prudence pour le calcul des rentes viagères, on adoptera la méthode d'extrapolation de Coale-Kisker; celle-ci présentant des taux de mortalité plus faibles aux grands âges (sur nos données).

Fermeture des tables

La question de la fermeture de la table est importante dans le cas de la construction d'une table pour des provisionnements de rentes viagères. On pourra toutefois noter que cette importance doit être relativisée si les rentiers d'âge très élevé sont en proportion modeste dans le portefeuille.

En effet, considérons l'exemple simple dans lequel on évalue un capital constitutif d'une rente viagère sur une tête avec la table TF00-02 ; on compare le calcul réalisé avec la table complète et celui réalisé avec la même table fermée de manière prudente en figeant le taux de décès à 95 ans. Ainsi, si pour évaluer le capital constitutif d'une rente viagère sur une tête à 75 ans au taux de 2,5 % (et avec la TH00-02), on considère que le taux de décès est stable à partir de 95 ans (et que les survivants sortent brutalement à 120 ans), on ne majore la provision que de 0,7 % (et environ 2,5 % à 85 ans).

L'écart entre deux méthodes de fermeture en terme de provisionnement n'est véritablement significatif qu'à des âges très élevés (voir par exemple Delwarde et Denuit [2006]).

Limites

Ainsi, le modèle log-Poisson contraint présente un certain nombre de « bonnes propriétés » et améliore notablement, dans le cas des petits échantillons étudié ici, les résultats obtenus par rapport à l'application directe du modèle de Lee-Carter. Toutefois, l'extrapolation des taux de décès aux grands âges, qui définit la structure de la table après 80 ans, peut apparaître relativement arbitraire. L'utilisation d'un modèle « endogène » est ici délicate compte tenu de la faible quantité d'information apportée par les données.

De plus, le modèle Log-Poisson, même contraint comme nous l'avons proposé *supra*, 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). Enfin, la relation

$$\mu_{xt} = -\ln(1 - q_{xt})$$

repose sur l'hypothèse de constance du taux instantané de décès entre deux âges entiers, hypothèse discutable aux âges élevés.

Dans ce contexte il peut être utile de se tourner vers des modèles alternatifs « naturellement » moins paramétrés et ne nécessitant pas l'hypothèse de constance du taux instantané sur chaque carré du diagramme de Lexis.

Il peut alors apparaître naturel de chercher à s'appuyer sur une structure exogène existante ; l'INSEE fournit ainsi une série de table du moment pour l'ensemble du 20e siècle, avec une projection jusqu'en 2050. Cet ensemble de table va alors nous servir à positionner les données d'expérience, selon une logique semblable à ce que propose par exemple le modèle de Cox.

Un modèle alternatif

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 :

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

Avec :

$$\mathbf{lg}_x(t) = \mathbf{ln}\left(q_{xt} / (1 - q_{xt})\right)$$

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

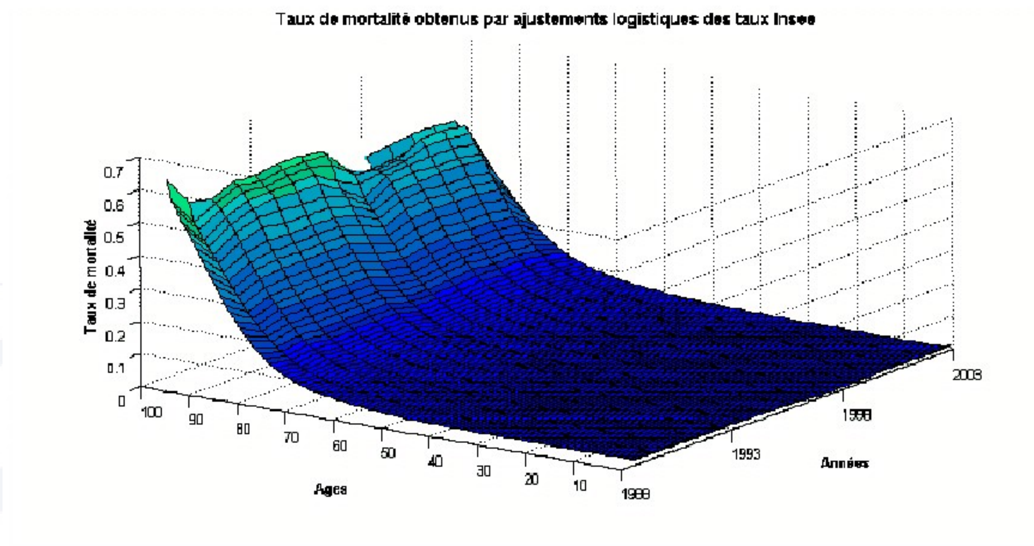
$$(\hat{a}, \hat{b}) = \mathbf{arg\,min}[e_{60}^{liss\acute{e}}(a,b) - e_{60}^{nonliss\acute{e}}]$$

sous la contrainte suivante :

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

Un modèle alternatif

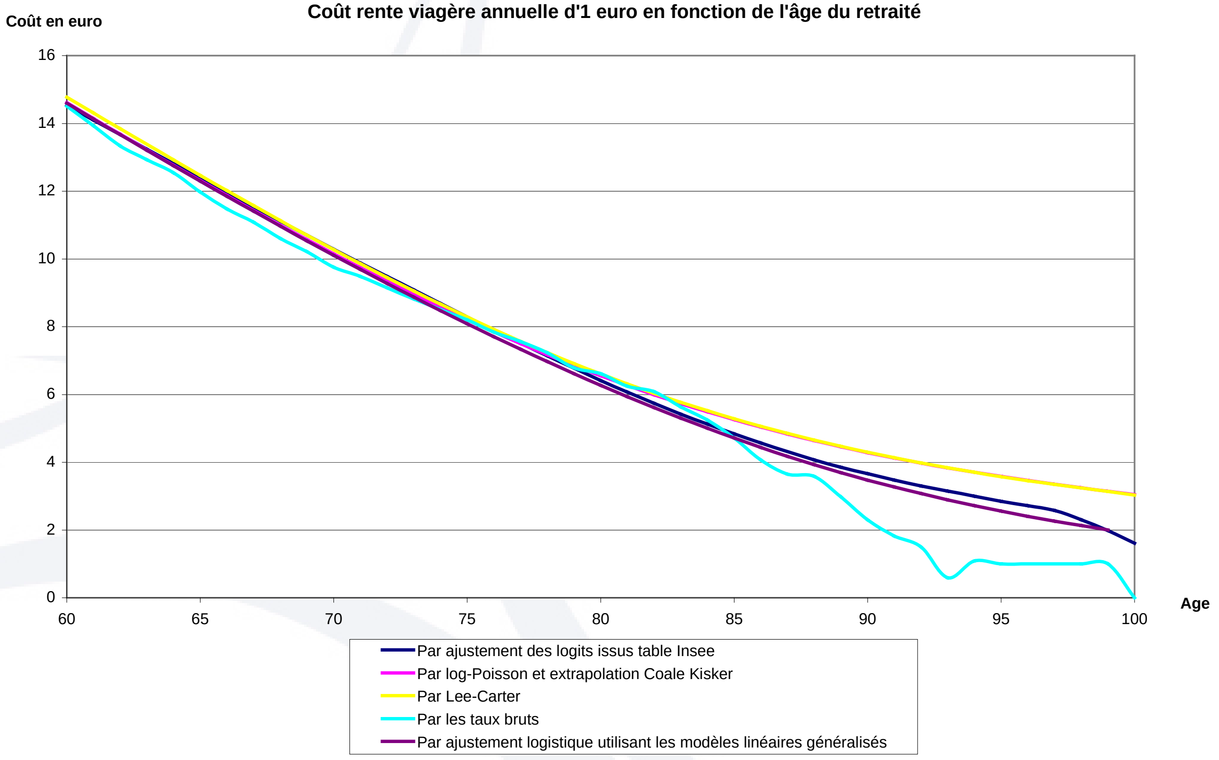
On obtient la surface suivante :



On constate qu'au-delà de 70 ans, les taux de mortalité lissés obtenus sont assez volatiles d'une année sur l'autre, contrairement aux résultats constatés avec les modèles de Lee-Carter et de Poisson, munis de l'ajustement polynomial des paramètres.

Provisionnement

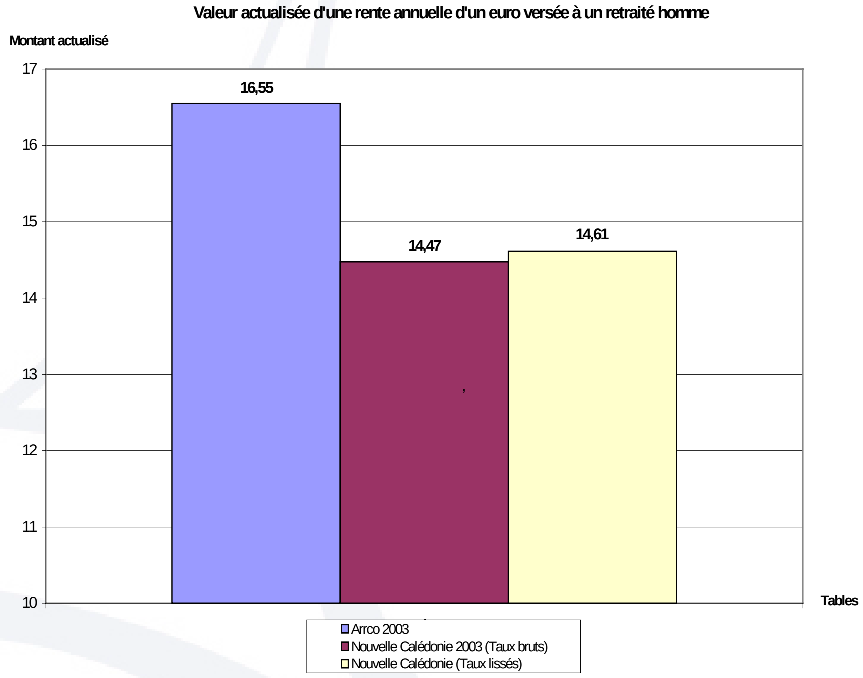
Dans un contexte de provisionnement de rentes viagères on a :



Ce graphique illustre le caractère prudent, par construction, du modèle log-Poisson (et Lee-Carter), notamment par rapport à l'approche logistique. On y lit également le caractère inexploitable des données brutes.

Provisionnement

On peut ainsi justifier le coût d'une rente viagère décalé par rapport aux tables de référence :



Si on effectue les calculs à partir de la table lissée par le modèle log-poisson sur la mortalité en Nouvelle Calédonie, 1€ de retraite versé à un homme en Nouvelle-Calédonie annuellement à partir de ses 60 ans coûte environ 11,8 % moins cher qu'en métropole, en considérant un taux d'actualisation de 2,25% (le coût étant actualisé à la date d'aujourd'hui).

Il existe par conséquent une faible différence sur le coût d'une rente viagère, selon la méthode de lissage utilisée (Lee-Carter ou log-Poisson). La méthode log-poisson est légèrement moins prudente que celle de Lee-Carter, puisqu'elle diminue le coût d'une rente viagère d'environ 1,15% par rapport au même calcul effectué avec Lee-Carter.

Provisionnement

Les deux méthodes prospectives d'ajustement proposées - par Lee-Carter, log-poisson - impliquent un provisionnement très proche pour le versement de rentes viagères, ceci pour toutes les années et tous les âges. Il n'y a jamais plus de 1% d'écart entre le provisionnement issu des tables produites par ces modèles.

Les tables produites par la méthode log-poisson sont cependant les plus prudentes. Elles impliquent quasiment toujours un provisionnement supérieur à celui issu des taux bruts. L'écart entre les deux provisionnements varie fréquemment entre 3% et 10% pour les âges compris entre 60 ans et 70 ans. Au-delà de cet âge, l'écart entre les deux provisionnements a tendance à devenir plus important (souvent supérieur à 15%).

Le provisionnement issu de l'ajustement logistique utilisant les modèles linéaires généralisés

est souvent proche de ceux obtenus par les modèles de Lee-Carter et log-poisson, pour les âges inférieurs à 75 ans. Après, ce modèle présente parfois un défaut de prudence (notamment entre 75 ans et 85 ans).

La méthode la moins prudente est celle ajustant les taux bruts à l'aide des taux logistiques Insee, hormis pour l'année 2003. Pourtant, le provisionnement ainsi calculé est souvent plus proche de celui issu des taux bruts (par rapport aux provisionnements calculés par les méthodes prospectives) ; mais il subsiste un risque de sous-provisionnement, particulièrement remarqué en 1990 entre 75 ans et 85 ans.

La méthode logistique par les modèles linéaires généralisés implique fréquemment le provisionnement le plus proche de celui issu des taux bruts. Mais, il subsiste comme pour l'autre modèle logistique un risque de sous-provisionnement, remarqué notamment en 1990 pour les âges compris entre 75 ans et 85 ans.

Conclusion

A partir de données brutes, la création de tables de mortalité prospectives prend en compte les deux composantes, âge et année, pour l'estimation des taux de mortalité, et permet la projection de la mortalité future de la population. Deux méthodes différentes, Lee-Carter et log-Poisson, ont été appliquées dans ce document. Le modèle de Lee-Carter ne prend pas en compte l'hétéroscédasticité des résidus alors que sa variante log-Poisson permet de s'affranchir de l'hypothèse d'hétéroscédasticité et met au surplus en œuvre une estimation par maximum de vraisemblance.

La construction de ces tables prospectives présente la particularité dans cette étude de s'effectuer sur des portefeuilles de taille réduite. Cela implique des difficultés non rencontrées sur les précédents travaux, lorsque ces modèles étaient utilisés sur des populations ayant une échelle nationale.

D'une part, la volatilité des taux de mortalité bruts observés, liée aux fluctuations d'échantillonnage, entraîne une instabilité des trois coefficients permettant le calcul des taux de mortalité par les méthodes prospectives. On est ainsi conduit à diminuer le nombre de degrés de liberté de ces paramètres. Le fait d'imposer une structure polynomiale aux paramètres a finalement permis un meilleur lissage de la mortalité observée, en réduisant de manière drastique le nombre de paramètres du modèle.

D'autre part, aux âges les plus élevés, au-delà de 90 ans, on ne dispose plus de données brutes sur une population de taille suffisante. Cette caractéristique explique l'intervention de la méthode d'extrapolation de Coale-Kisker pour obtenir les taux de mortalité aux âges avancés.

Les résultats obtenus par l'utilisation de ces deux modèles montrent qu'en présence de l'ajustement polynomial des paramètres et lorsqu'on applique les méthodes sur les effectifs bruts (sans normalisation à 100 000 personnes à la naissance), les taux de mortalité obtenus sont très proches.

Le modèle log-Poisson, associé à l'extrapolation de Coale-Kisker, apparaît meilleur sur la période 1989-2003 (période d'étude) que le modèle de Lee-Carter. Au surplus, il est plus prudent que celui de Lee-Carter pour l'estimation de rentes viagères sur cette période.

Conclusion

Afin de remédier aux inconvénients associés à la petite taille de l'échantillon, il est également possible de se tourner vers des modèles à référence externe. Le modèle retenu consiste à ajuster les taux logistiques bruts linéairement en fonction des taux logistiques des tables INSEE. L'ajustement est réalisé de telle sorte que les espérances de vie à 60 ans par les taux bruts et lissés soient équivalentes. Ce modèle permet d'obtenir un lissage conduisant par construction à des montants de provisions très proches de ceux issus des données brutes, mais présente parfois l'inconvénient d'être moins prudent que les modèles de Lee-Carter et log-poisson pour l'estimation du coût des rentes viagères.

D'autres techniques de lissage que celles présentées ici auraient pu être appliquées, on pense notamment aux ajustements par splines (De Borr [1978]) ainsi qu'à la méthode non paramétrique de Whittaker-Henderson (Taylor [1992]), qui présente l'intérêt d'appartenir à la famille des lissage bayésiens.

Enfin, on soulignera que l'ensemble des modèles discutés ici se prête aisément à une approche stochastique, pour la mesure du risque systématique de mortalité auquel se trouve confronté un régime de rentiers ; des travaux sont en cours afin de proposer un modèle simple permettant de quantifier ce risque, et notamment d'en comparer l'importance par rapport au risque financier auquel est également soumis le régime.

Ils fournissent une alternative intéressante aux modèles à structure affine d'origine financière souvent proposés pour la modélisation de ce risque.