

Université Catholique de Louvain  
Institut d'Administration et de Gestion  
Institut des Sciences Actuarielles



# Modèle log-bilinéaire pour l'élaboration de tables de mortalité prospectives

Directeur: Professeur M. Denuit

Mémoire présenté par Antoine Delwarde  
en vue de l'obtention du Diplôme d'Études  
Spécialisées en Sciences Actuarielles

Année académique 2003-2004

# Remerciements

Nous tenons à remercier Monsieur le Professeur Michel Denuit pour nous avoir éclairé tout au long de l'élaboration de ce mémoire et pour sa disponibilité de tous les instants. Notre gratitude s'adresse également à toutes les personnes qui, de près ou de loin, ont contribué à l'écriture de ces lignes. Enfin, nos remerciements vont à nos parents pour leur soutien et leur présence, à nos proches, à nos amis et à Delphine, sans qui rien de tout cela n'aurait été possible.

# Table des matières

<b>1</b>	<b>Le diagramme de Lexis</b>	<b>4</b>
1.1	La mesure du temps en démographie . . . . .	4
1.2	La représentation de Lexis . . . . .	5
1.3	Générationns . . . . .	6
1.4	Effectifs à un moment donné et effectifs atteignant un âge donné	6
1.5	Nombre de décès . . . . .	6
<b>2</b>	<b>Estimation brute des statistiques de mortalité</b>	<b>10</b>
2.1	Notations . . . . .	10
2.1.1	Nombre de décès . . . . .	10
2.1.2	Nombre de survivants . . . . .	11
2.2	Méthode de Vallin . . . . .	11
2.3	Méthode des générations éteintes . . . . .	13
2.4	Répartition des décès dans l'année . . . . .	14
2.4.1	Répartition uniforme des décès sur l'année . . . . .	14
2.4.2	Taux de mortalité constants par morceaux . . . . .	16
2.4.3	Comparaison des hypothèses . . . . .	17
2.4.4	Généralisation . . . . .	18
<b>3</b>	<b>Modèle log-bilinéaire pour l'élaboration de tables de mortalité prospectives</b>	<b>20</b>
3.1	Introduction . . . . .	20
3.2	Modèle de Lee et Carter . . . . .	21
3.2.1	Modèle . . . . .	21
3.2.2	Estimation des paramètres par moindres carrés: analyse en composantes principales . . . . .	23
3.2.3	Estimation des paramètres par moindres carrés: méthode itérative . . . . .	30
3.2.4	Estimation des paramètres par moindres carrés pondérés	32
3.2.5	Estimation des paramètres par moindres carrés pénalisés	35
3.2.6	Mesure de la qualité de l'ajustement . . . . .	39
3.2.7	Analyse des résidus . . . . .	41
3.2.8	Extrapolation de la tendance temporelle . . . . .	43
3.2.9	Extension de Renshaw-Haberman . . . . .	50

3.2.10	Modèle avec tendance temporelle linéaire . . . . .	55
3.2.11	Modèle avec table limite . . . . .	59
3.3	Modèle de Poisson . . . . .	63
3.3.1	Modèle . . . . .	64
3.3.2	Estimation des paramètres par maximum de vraisemblance	66
3.3.3	Estimation des paramètres par maximum de vraisemblance pénalisée . . . . .	70
3.3.4	Mesure de la qualité de l'ajustement . . . . .	73
3.3.5	Analyse des résidus . . . . .	75
3.3.6	Extrapolation de la tendance temporelle . . . . .	78
3.3.7	Extension de Renshaw-Haberman . . . . .	80
3.3.8	Modèle avec tendance temporelle linéaire . . . . .	86
3.3.9	Modèle avec table limite . . . . .	89
<b>4</b>	<b>Intervalles de confiance pour les indicateurs de mortalité</b>	<b>92</b>
4.1	Approche par simulation Monte Carlo . . . . .	92
4.1.1	Principe . . . . .	92
4.1.2	Calcul de la matrice d'information de Fisher . . . . .	93
4.2	Méthodes de rééchantillonnage des décès . . . . .	98
4.2.1	Bootstrap paramétrique . . . . .	98
4.2.2	Bootstrap non paramétrique . . . . .	101
4.3	Méthode de rééchantillonnage des résidus . . . . .	104

# Introduction

## La difficile tarification des rentes viagères

Les bases techniques des contrats d'assurance sur la vie sont au nombre de trois: taux d'intérêt, table de mortalité et chargements. Le taux d'intérêt permet d'actualiser les flux financiers générés par le contrat. La table de mortalité décrit la loi de survenance des décès, laquelle permet d'évaluer le coût moyen des contrats souscrits par la compagnie. Les chargements viennent grever la prime pure, et couvrent les frais de gestion de la police.

La détermination du montant de la prime pure relative à un contrat d'assurance sur la vie se fait sur base du principe d'équivalence, en vertu duquel l'espérance de la valeur actuelle des prestations de l'assureur doit être égale à l'espérance de la valeur actuelle des primes pures payées par l'assuré.

Traditionnellement, les actuaires fondent le calcul des primes d'assurance sur des tables de mortalité transversales: les individus sont observés pendant une période variant de un à cinq ans, et fournissent les quotients de mortalité qui seront utilisés pour générer la table d'expérience. Principal avantage de cette approche, elle ne requiert que l'observation d'un groupe d'individus pendant une durée relativement courte. Revers de la médaille, la table de mortalité ainsi obtenue mélangera de nombreuses générations, et ne décrira la mortalité d'aucune d'entre elles. En particulier, les indicateurs démographiques qui en découlent (comme l'espérance de vie, par exemple) n'ont aucun sens concret, ne correspondant à aucune génération d'individus.

Baser une tarification sur une table transversale revient implicitement à considérer que la mortalité est figée. Or, l'analyse des statistiques de mortalité publiées par les instituts nationaux de statistique montre que l'espérance de vie est en constante progression dans les pays industrialisés. Cette tendance n'est pas sans préoccuper les assureurs commercialisant des rentes, qui doivent désormais recourir à des tables de mortalité longitudinales ou dynamiques, anticipant l'évolution probable de la mortalité des individus.

En marge du risque technique inhérent aux rentes viagères et maîtrisé par les mécanismes de compensation, et du risque financier généré par la garantie du taux d'intérêt technique durant toute la durée du contrat, naît un nouveau risque résultant de l'évolution inconnue de la mortalité dans le futur. Ce nouveau risque est appelé risque de longévité et sera d'autant plus important que le rythme d'allongement de la durée de la vie humaine est élevé.

Le risque de longévité rentre dans la catégorie des risques de modèle: il trouve son origine dans l'inadéquation possible entre la table de mortalité et la survie réelle des rentiers. Un tel risque se retrouve bien entendu dans tous les produits d'assurance: si les bases techniques utilisées par l'assureur ne reflètent pas la sinistralité réelle qu'il s'est engagé à couvrir, l'assureur est confronté à des pertes techniques qui peuvent être conséquentes, et auxquelles il répondra généralement par une adaptation de la prime et/ou une diminution de la garantie. Mais l'architecture même des rentes exacerbe le risque de modèle. En effet, ces contrats prévoient une prime unique, ce qui rend impossible une correction de la tarification en cours de contrat. De plus, les prestations sont fixées lors de la conclusion, ce qui ne permet plus de les adapter en cours de contrat. Enfin, la durée de l'engagement de l'assureur est considérable, ce qui augmente l'incertitude quant aux tables de mortalité utilisées par la compagnie. Il est clair que modifier ces caractéristiques, en reportant les risques de mortalité et de longévité sur les assurés, reviendrait à vider les rentes de leur intérêt social.

Un autre écueil menace les assureurs: l'antisélection souvent importante existant sur le marché des rentes viagères. Lors de l'établissement des tables de mortalité pour rentes viagères, il faut garder à l'esprit que la mortalité des rentiers est sans doute différente de celle de la population générale. Il semble évident que la plupart des personnes qui souscrivent à ce genre de contrats estiment être en excellente santé. Il est donc légitime de s'attendre à une sous-mortalité importante de la part des rentiers par rapport à la population dans son ensemble. La non-prise en compte de ce phénomène engendrerait pour l'assureur de lourdes pertes techniques, ce qui oblige l'actuaire à adapter les tables dressées sur base des statistiques de mortalité relatives à la population générale à la mortalité spécifique aux rentiers. Le phénomène d'antisélection mérite certainement un examen attentif. Il ne sera néanmoins pas développé dans ces lignes.

## Objectifs et plan

L'approche transversale de la mortalité, classique aux yeux des actuaires, ne fait ici l'objet d'aucun développement. Nous préférons traiter directement des méthodes de projection de la mortalité.

Les objectifs de ce travail sont au nombre de trois. Tout d'abord, nous présentons une batterie de modèles utiles à la construction de tables de mortalité prospectives. Ils permettent ainsi de passer des données communément utilisées pour ce type de travail aux tables proprement dites. Ensuite, certaines clés sont fournies pour l'utilisation de ces modèles, et notamment les algorithmes d'estimation des paramètres. Enfin, les méthodes sont illustrées par des exemples numériques basés sur des données françaises. Ainsi, il est possible d'analyser et de comparer les diverses techniques proposées.

L'analyse dynamique de la mortalité a pour cadre le diagramme de Lexis qui

permet de rendre facilement compte des trois dimensions que sont l'âge, l'année et la génération. Ce système de représentation est développé au Chapitre 1.

Ensuite, le Chapitre 2 présente les méthodes d'estimation non-paramétriques des indicateurs de mortalité, tels que les quotients et taux instantanés de mortalité.

Des tables de mortalité prospectives ont été dressées dans bon nombre de pays européens à l'aide de techniques diverses et souvent très empiriques<sup>1</sup>. Le Chapitre 3 décrit en profondeur la méthode désormais classique de LEE & CARTER (1992), ainsi que certaines de ses variantes, notamment le modèle de Poisson log-bilinéaire. Ces dernières diffèrent du modèle initial tantôt par la variable dépendante, tantôt par la méthode d'estimation des paramètres.

Finalement, l'incertitude inhérente à la projection peut s'apprécier grâce à des techniques de simulation de Monte Carlo ainsi qu'à des méthodes de rééchantillonnage (ou bootstrap). Ces techniques font l'objet du Chapitre 4.

## Applications

Les données utilisées dans ce travail sont fournies conjointement par l'INSEE (Institut National de la Statistique et des Études Économiques) et l'INED (Institut National d'Études Démographiques), et couvrent la population masculine française sur la période 1950-2000. Nous disposons de deux tables contenant respectivement l'effectif de la population au premier janvier et le nombre de décès recensés lors de chacune des années. Ces tableaux sont ventilés par âge entre 0 et 99 ans. Notons cependant que l'effectif au premier janvier n'est disponible que jusque 89 ans entre 1950 et 1953, et jusque 94 ans entre 1954 et 1974. Certaines des méthodes développées plus loin, notamment le modèle de Lee et Carter présenté à la section 3.2, nécessitent des tableaux dépourvus de valeurs manquantes. En ce sens, dans un but de comparaison, nous avons pris le parti d'homogénéiser les données utilisées pour les illustrations des différentes méthodes, et donc de ne considérer les tableaux que jusque 89 ans.

Enfin, l'INED fournit les chiffres entre 1950 et 1997, tandis que les données des trois dernières années 1998 à 2000 sont apportées par l'INSEE. Les sources étant différentes, il n'est pas exclu que les analyses mettent en évidence une transition brusque entre 1997 et 1998.

---

1. Pour une description de celles-ci, voyez McDONALD (1997) et McDONALD ET AL. (1998).

# Chapitre 1

## Le diagramme de Lexis

### 1.1 La mesure du temps en démographie

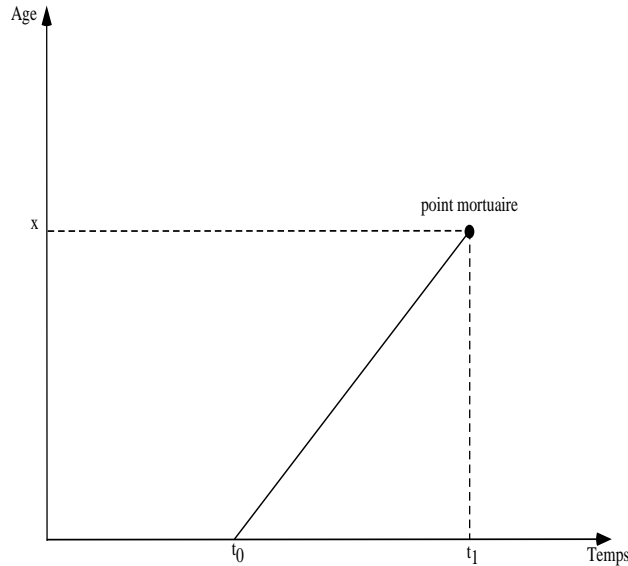
Le temps intervient sous trois formes différentes dans l'étude de la mortalité: l'âge des individus, l'instant d'observation et la génération dont les individus font partie. Chacun de ces trois aspects influence la mortalité. Tout d'abord, il est évident que l'âge d'un individu influence le risque décès. Ensuite, le risque de mourir à un âge donné pourra varier d'une année à l'autre au gré de circonstances telles qu'une épidémie, une guerre, un hiver particulièrement rude, ou un été torride. On parle d'effet de période. L'instant d'observation est donc décisif. Enfin, le risque de mourir à un âge donné peut aussi se modifier indépendamment d'un effet de moment, selon le passé des individus concernés. Supposons par exemple que nous désirions comparer la mortalité à l'âge  $x$  durant deux années distinctes  $t$  et  $t'$ . Les individus concernés sont donc nés lors de deux années différentes  $g$  et  $g'$ . Supposons que l'année  $g$ , une grave épidémie ait frappé les nourrissons et décimé la génération. À l'inverse, la génération  $g'$  n'a pas eu à souffrir d'une telle calamité. Supposons en outre qu'entre les années  $t$  et  $t'$ , les conditions de vie (niveau de vie, moyens sanitaires, hygiène, etc.) soient restées identiques. Il est réaliste de penser que le risque de mourir à l'âge  $x$  sera plus faible l'année  $t$  que l'année  $t'$ . En effet, l'épidémie subie par la génération  $g$  pourrait avoir opéré une sélection importante, avec pour conséquence que seuls les enfants les plus robustes auront survécu. Le risque de décès diffère donc suite à un effet de génération.

On dispose ainsi de trois coordonnées pour repérer un événement tel que le décès d'un individu:

1. la date à laquelle il se produit
2. l'âge de l'individu concerné par l'événement
3. le moment de naissance de l'individu concerné par l'événement.

Ces trois données sont évidemment redondantes: seules deux d'entre elles sont en fait nécessaires pour repérer l'événement. Néanmoins, les trois deviennent indispensables lorsqu'on s'intéresse à un groupe d'individus, comme nous le verrons plus loin. Le diagramme de Lexis permet le repérage de faits démogra-



FIG. 1.1 – *Parcours d'un individu dans le diagramme de Lexis.*

phiques en fonction des trois coordonnées date-âge-génération. Il privilégie en fait les coordonnées "âge" et "date", qui se voient chacune attribuer un axe.

## 1.2 La représentation de Lexis

Le diagramme de Lexis est un système d'axes rectangulaires permettant de préciser comment se combinent les mesures de temps, selon le calendrier civil et selon la durée écoulée depuis un événement antérieur. Les dates du calendrier sont portées en abscisse, les durées écoulées figurent en ordonnée, l'échelle de mesure du temps étant la même. S'agissant de la mortalité, la durée prise en compte est l'âge, durée écoulée entre le décès et l'événement antérieur qu'est la naissance. Dans un tel système d'axes, le déroulement de l'existence d'un individu est figuré par une ligne de vie, parallèle à la première bissectrice (l'âge de l'individu augmente d'une année pour chaque année qui passe) interrompue lors du décès en un point, appelé le point mortuaire. Nous avons représenté à la Figure 1.1 le parcours d'un individu dans le diagramme de Lexis: celui-ci est né à la date  $t_0$ , et décédé à la date  $t_1$  à l'âge  $x$ . Clairement,  $x = t_1 - t_0$ .

L'âge peut faire l'objet d'une mesure exacte, ou être exprimé plus approximativement, mais plus commodément en années révolues. L'âge en années révolues est exprimé par la valeur du nombre entier d'années au dernier anniversaire. L'âge exact au 31 juillet 2002 d'un individu né le 30 juin 2001 est d'un an et un mois; son âge en années révolues, à la même date, est d'un an.

### 1.3 Générations

En règle générale, nous ne nous intéressons pas à un individu en particulier, mais plutôt à des cohortes d'individus. Le diagramme de Lexis comporte, outre le quadrillage formé par les droites parallèles aux axes rectangulaires, les diagonales limitant les lignes de vie des personnes nées au cours d'une même année. Ces personnes font donc partie d'une même génération. Dorénavant, nous désignerons une génération par l'année de naissance des individus qui la composent; ainsi, la génération  $g$  est constituée de tous les individus de la population nés au cours de l'année calendrier  $g$ .

Le diagramme de Lexis nous permet de repérer facilement

1. l'effectif d'une génération, aux intersections avec les axes verticaux (effectifs aux premiers janvier successifs, date à laquelle tous les individus ont le même âge en années révolues) ou horizontaux (effectifs aux anniversaires successifs);
2. le nombre de décès localisés dans les surfaces délimitées par le système.

### 1.4 Effectifs à un moment donné et effectifs atteignant un âge donné

Lors d'un recensement, on note l'âge des individus à un moment précis. Les résultats publiés nous apprennent l'effectif correspondant à chaque âge révolu, i.e. le nombre d'individus ayant atteint l'âge considéré à l'instant du recensement. La localisation de cet effectif s'obtient en faisant l'intersection entre la verticale correspondant à la date du recensement et le couloir horizontal correspondant à l'âge révolu. Ainsi, si à l'instant  $t_0$  du recensement,  $S_x$  individus avaient  $x$  années révolues, on note ce nombre verticalement sur le segment en question, comme décrit à la Figure 1.2. Cette information est donc traduite par le fait que  $S_x$  lignes de vie ont au moins traversé ce segment.

Une autre information est de savoir qu'au cours de l'année  $t$ ,  $S_x$  individus ont fêté leur  $x^{\text{ème}}$  anniversaire. Ce nombre  $S_x$  est à placer sur le segment à l'intersection entre l'horizontale correspondant à l'âge  $x$  et le couloir vertical correspondant à l'année  $t$ . Cela signifie que  $S_x$  lignes de vie ont au moins atteint ce segment. Voyez la Figure 1.3. Notez que tous ces individus appartiennent à la génération  $t - x$ .

### 1.5 Nombre de décès

Pendant l'année  $t$ , on a relevé  $D_{x,t}$  décès parmi les individus de  $x$  années révolues. Ces décès sont à localiser dans un carré, intersection entre un couloir vertical (celui de l'année  $t$ ) et un couloir horizontal (celui de l'âge  $x$ ). Cela signifie que  $D_{x,t}$  lignes de vie se sont terminées par un décès dans ce carré. Voyez la Figure 1.4. Notez que ces  $D_{x,t}$  individus décédés appartenaient aux deux générations  $t - x$  et  $t - x - 1$ .

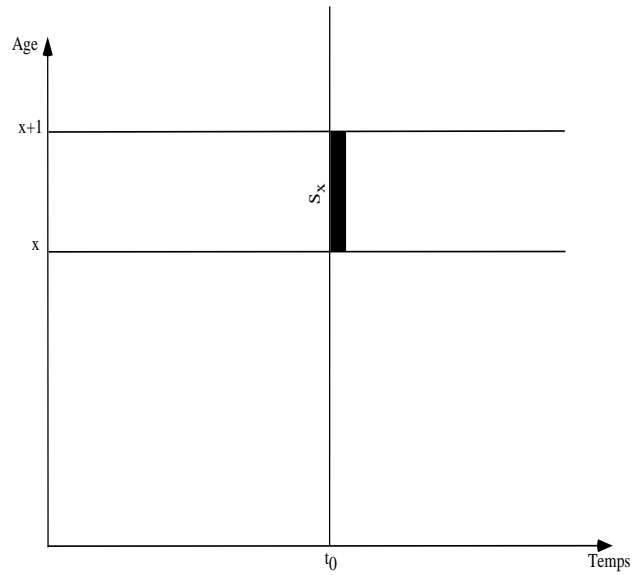


FIG. 1.2 – Nombre de survivants d'âge  $x$  à une date donnée  $t_0$ .

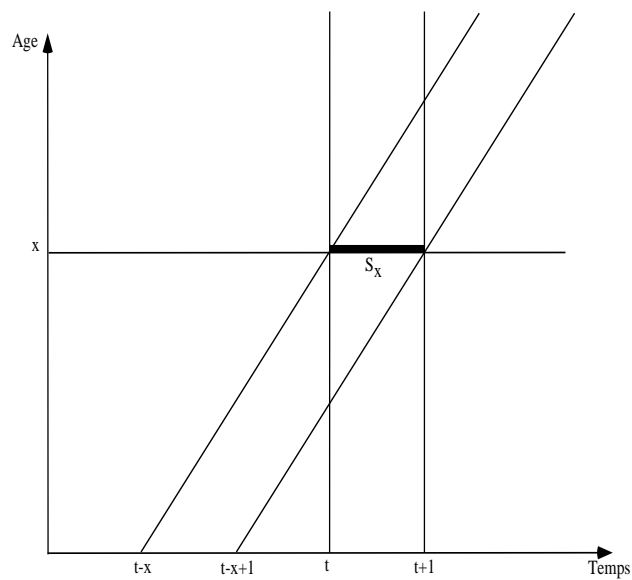
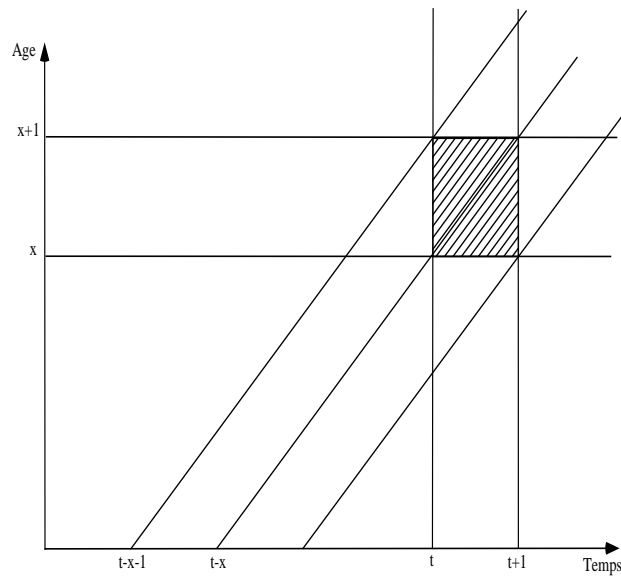


FIG. 1.3 – Nombre de survivants à l'âge  $x$  au sein d'une génération donnée.

FIG. 1.4 – Nombre de décès à l'âge  $x$  durant l'année  $t$ .

Il est également possible de relever le nombre de décès à l'âge  $x$  dans une génération  $g$ . Ces décès sont à localiser dans un parallélogramme, intersection entre un couloir oblique (correspondant à la génération  $g$ ) et un couloir horizontal (correspondant à l'âge  $x$ ). Voyez la Figure 1.5. Notez que ces décès se sont produits au cours des deux années  $g+x$  et  $g+x+1$ .

Enfin, si on s'intéresse au nombre de décès d'une année au sein d'une génération, ces décès sont à localiser dans un parallélogramme, intersection entre un couloir oblique (la génération  $g$ ) et un couloir vertical (l'année  $t$ ). Voyez la Figure 1.6. Ces décès se sont produits aux âges  $t-g$  et  $t-g-1$ .

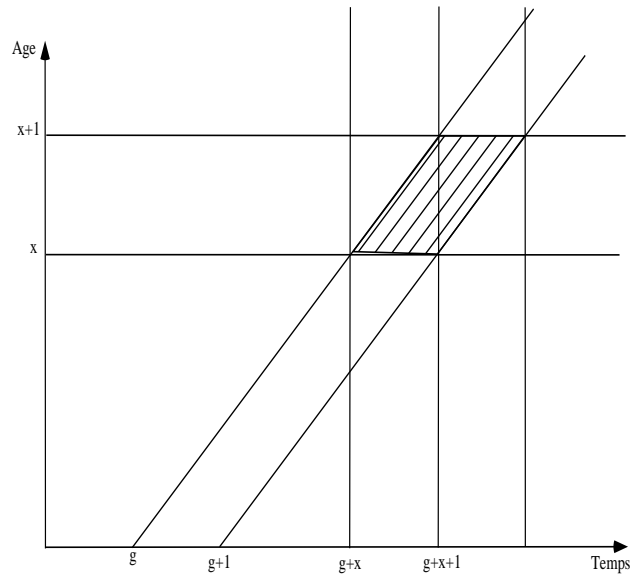


FIG. 1.5 – Nombre de décès à l'âge  $x$  parmi les individus de la génération  $g$ .

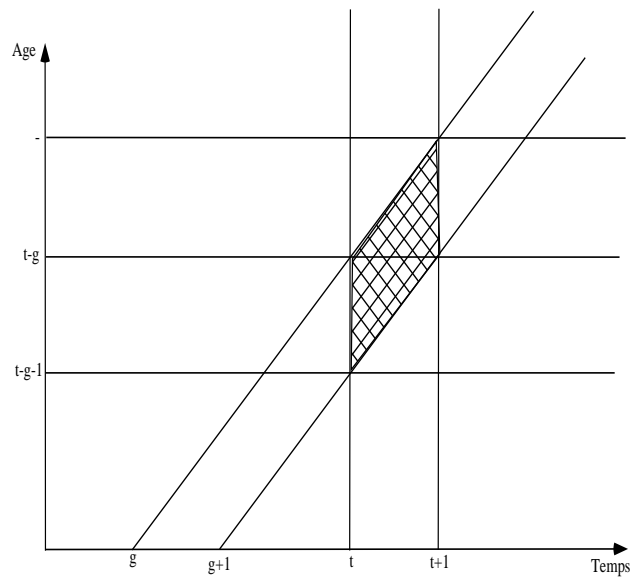


FIG. 1.6 – Nombre de décès parmi les individus de la génération  $g$  au cours de l'année  $t$ .

## Chapitre 2

# Estimation brute des statistiques de mortalité

### 2.1 Notations

#### 2.1.1 Nombre de décès

Notons  $g = t - x$  la plus grande des deux années de naissance des individus âgés de  $x$  années révolues lors de l'année  $t$ . Intéressons-nous à l'estimation des quotients de mortalité au sein d'un carré  $(x, t)$  du diagramme de Lexis tel que représenté à la Figure 1.4. Une telle surface peut être décomposée en deux triangles: le triangle supérieur regroupe le nombre de décès  $D_{x,t,g-1}$  survenus l'année  $t$  parmi les individus d'âge  $x$  et issus de la génération  $g - 1 = t - x - 1$ ; le triangle inférieur représente quant à lui les décès  $D_{x,t,g}$  survenus la même année parmi les personnes de même âge mais nés lors de l'année  $g = t - x$ . Définissons donc  $D_{x,t,\bullet}$  le nombre total de décès observés dans chacun de ces carrés:

$$D_{x,t,\bullet} = D_{x,t,g} + D_{x,t,g-1}.$$

À chacun de ces nombres de décès correspond un quotient  $q_{x,t,\bullet}$  régissant la mortalité dans chacun des carrés.

De la même manière,  $D_{x,\bullet,g}$  représente le nombre de décès survenus à l'âge  $x$  pour la génération  $g$ . La surface du diagramme de Lexis qui nous intéresse est alors constituée par le parallélogramme de la Figure 1.5. Il vient donc

$$D_{x,\bullet,g} = D_{x,t,g} + D_{x,t+1,g}$$

et le quotient de mortalité  $y$  correspondant est noté  $q_{x,\bullet,g}$ .

Enfin, le nombre  $D_{\bullet,t,g}$  de décès survenus l'année  $t$  parmi les individus de la génération  $g$ , donc dans le parallélogramme de la Figure 1.6, vaut

$$D_{\bullet,t,g} = D_{x,t,g} + D_{x-1,t,g}$$

et le quotient de mortalité qui lui est relatif est noté  $q_{\bullet,t,g}$ .

Dans la plupart des applications ultérieures, nous ne nous intéressons qu'à des carrés du diagramme de Lexis, sans différencier les deux triangles qui les composent. À ce titre, nous utiliserons les notations simplifiées

$$D_{x,t} = D_{x,t,\bullet}$$

et

$$q_x(t) = q_{x,t,\bullet}$$

### 2.1.2 Nombre de survivants

Notons  $L_{x,t}$  le nombre d'individus âgés de  $x$  années révolues le premier janvier de l'année  $t$ . Il s'agit en fait du cas particulier de la section 1.4 où on choisit pour dates  $t_0$  des moments équidistants d'un an. Ces nombres sont donc représentés par des segments verticaux tels que celui de la Figure 1.2 pris à intervalles réguliers.

Il est intéressant de constater qu'en cas d'absence de migration, nous disposons de la relation

$$\begin{aligned} L_{x,t} &= \sum_{i=0}^{\omega-x} D_{\bullet,t+i,g-1} \\ &= \sum_{i=0}^{\omega-x} D_{x+i,t+i,g-1} + D_{x+1+i,t+i,g-1} \end{aligned}$$

où  $\omega$  représente l'âge ultime qu'aucun membre de la génération ne dépasse. Cette formule, qui retranscrit tout simplement le fait que chacun des  $L_{x,t}$  individus d'âge  $x$  au premier janvier de l'année  $t$  mourra un jour, sera à la base de la méthode des générations éteintes vue à la section 2.3. Elle permet également d'éprouver l'hypothèse d'absence de migration au sein de la population, voire d'analyser la qualité des données disponibles.

## 2.2 Méthode de Vallin

La méthode de VALLIN (1973) permet, à partir des  $D_{x,t,g}$  et  $L_{x,t}$ , d'estimer les quotients de mortalité  $q_x(t)$  correspondant aux carrés du diagramme de Lexis tels que celui de la Figure 1.4.

Cette technique consiste à estimer des quotients de mortalité partiels pour chacun des deux triangles  $(x,t,g)$  et  $(x,t,g-1)$  (soit respectivement les triangles inférieur et supérieur) de la Figure 1.4. En l'absence de migration, il est possible d'estimer des quotients de mortalité partiels pour chacun de ces deux types d'individus. Ces quotients sont définis par le rapport entre le nombre de décès recensés et la population qui doit y faire face, donc qui entre dans la surface considérée. Dans le triangle inférieur, nous trouvons

$$q_{x,t,g} = \frac{D_{x,t,g}}{L_{x,t+1} + D_{x,t,g}} \quad (2.1)$$

tandis que le quotient partiel du triangle supérieur vaut

$$q_{x,t,g-1} = \frac{D_{x,t,g-1}}{L_{x,t}}. \quad (2.2)$$

Les migrations peuvent être prises en considération en posant l'hypothèse qu'elles sont uniformément réparties sur un intervalle d'un an. Dès lors, pour le triangle inférieur, il vient une immigration nette de

$$i_{x,t,g} = \frac{1}{2}(L_{x,t+1} - L_{x-1,t} + D_{x,t,g} + D_{x-1,t,g}).$$

Si on suppose que les migrants n'ont pas une mortalité différente de l'ensemble de la population, il est possible d'adapter le quotient partiel  $q_{x,t,g}$ :

$$q_{x,t,g} = \frac{D_{x,t,g} - \frac{1}{2}q_{x,t,g}i_{x,t,g}}{L_{x,t+1} - i_{x,t,g} + D_{x,t,g} - \frac{1}{2}q_{x,t,g}i_{x,t,g}}.$$

En négligeant  $\frac{1}{2}q_{x,t,g}^2 i_{x,t,g}$ , il vient alors

$$\begin{aligned} q_{x,t,g} &\approx \frac{D_{x,t,g}}{L_{x,t+1} + D_{x,t,g} - \frac{1}{2}i_{x,t,g}} \\ &\approx \frac{D_{x,t,g}}{\frac{3}{4}(L_{x,t+1} + D_{x,t,g}) + \frac{1}{4}(L_{x-1,t} - D_{x-1,t,g})}. \end{aligned} \quad (2.3)$$

De la même manière, l'immigration nette pour le triangle supérieur vaut

$$i_{x,t,g-1} = \frac{1}{2}(L_{x+1,t+1} - L_{x,t} + D_{x+1,t,g-1} + D_{x,t,g-1})$$

et le quotient partiel  $q_{x,t,g-1}$  devient alors

$$\begin{aligned} q_{x,t,g-1} &= \frac{D_{x,t,g-1} - \frac{1}{2}q_{x,t,g-1}i_{x,t,g-1}}{L_{x,t}} \\ &= \frac{D_{x,t,g-1}}{L_{x,t} + \frac{1}{2}i_{x,t,g-1}} \\ &= \frac{D_{x,t,g-1}}{\frac{3}{4}L_{x,t} + \frac{1}{4}(L_{x+1,t+1} + D_{x+1,t,g-1} + D_{x,t,g-1})}. \end{aligned} \quad (2.4)$$

Dès lors, la probabilité de ne pas mourir à l'âge  $x$  pour la génération  $g$  peut être estimée par

$$\widehat{p}_{x,\bullet,g} = p_{x,t,g} \cdot p_{x,t+1,g}$$

avec  $p_{x,t,g} = 1 - q_{x,t,g}$  et  $p_{x,t,g-1} = 1 - q_{x,t,g-1}$ . Il vient donc une estimation du quotient de mortalité correspondant:

$$\begin{aligned} \widehat{q}_{x,\bullet,g} &= 1 - (1 - q_{x,t,g})(1 - q_{x,t+1,g}) \\ &= q_{x,t,g} + q_{x,t+1,g} - q_{x,t,g}q_{x,t+1,g}. \end{aligned} \quad (2.5)$$



Ces quotients représentent la mortalité survenue dans un parallélogramme du diagramme de Lexis tel que représenté à la Figure 1.5. Ils permettent donc de construire des tables de mortalité par générations.

Il est également possible d'estimer la probabilité  $q_{\bullet,t,g}$  pour un individu issu de la génération  $g$  de mourir lors de l'année  $t$ :

$$\begin{aligned}\widehat{q}_{\bullet,t,g} &= 1 - (1 - q_{x,t,g})(1 - q_{x-1,t,g}) \\ &= q_{x,t,g} + q_{x-1,t,g} - q_{x,t,g}q_{x-1,t,g}.\end{aligned}\tag{2.6}$$

Ces quotients sont donc relatifs aux parallélogrammes du diagramme de Lexis de la Figure 1.6.

L'estimation de  $q_{x,t,\bullet}$ , quotient relatif au carré de la Figure 1.4, est plus délicate. En effet, ce quotient combine des individus issus de deux générations différentes. VALLIN (1973) et MESLÉ & VALLIN (2002) proposent de calculer l'estimation  $\widehat{q}_{x,t,\bullet}$  à la manière de (2.5) et (2.6), i.e.

$$\widehat{q}_x(t) = \widehat{q}_{x,t,\bullet} = 1 - (1 - q_{x,t,g})(1 - q_{x,t,g-1}).\tag{2.7}$$

Appelons  $A$  l'événement selon lequel un individu arrivé aux portes du triangle inférieur  $y$  décède. De la même manière, l'événement  $B$  indique le décès d'une personne dans le triangle supérieur. La probabilité qui nous intéresse est donc

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

qui peut être estimée si on considère que les événements  $A$  et  $B$  sont indépendants:

$$\widehat{q}_x(t) = \widehat{q}_{x,t,\bullet} = q_{x,t,g} + q_{x,t,g-1} - q_{x,t,g}q_{x,t,g-1}.\tag{2.8}$$

Nous retrouvons donc la formule proposée par les auteurs sus-nommés.

## 2.3 Méthode des générations éteintes

Dans de nombreux cas, alors que les statistiques sur les décès  $D_{x,t,g}$  sont suffisamment fournies, il est impossible d'obtenir des chiffres complets sur les effectifs de survivants de la population  $L_{x,t}$ . Ceux-ci sont souvent tronqués à un âge relativement faible (95 ou 100 ans), ou sont regroupés en catégories plus importantes (par groupes de 5 ans par exemple). Néanmoins, avec l'allongement de la vie humaine, il devient nécessaire d'accorder une importance croissante aux âges avancés. VINCENT (1951) propose donc de reconstituer les effectifs de survivants pour une génération en additionnant les nombres de décès observés ultérieurement en son sein. Le dénominateur de (2.1) devient donc

$$\sum_{i=0}^{\omega-x} D_{x+i,\bullet,g} = \sum_{i=0}^{\omega-x} D_{x+i,t+i,g} + D_{x+i,t+1+i,g}$$

où  $\omega$  est l'âge limite auquel toute la génération  $g$  s'éteint. De la même manière, nous trouvons le nouveau dénominateur de (2.2):

$$\sum_{i=0}^{\omega-x} D_{\bullet, t+i, g-1} = \sum_{i=0}^{\omega-x} D_{x+i, t+i, g-1} + D_{x+1+i, t+i, g-1}.$$

Ces formules supposent évidemment une migration nulle au delà de l'âge  $x$ , hypothèse facilement critiquable chez les jeunes. Néanmoins, selon KANNISTO (1988), elle devient vraisemblable dès lors qu'on s'approche des très grands âges, où la mobilité internationale peut être considérée comme quasiment nulle. La méthode des générations éteintes s'adressant prioritairement aux âges avancés, l'hypothèse de migration nulle est donc plausible.

Cette technique se heurte cependant à un obstacle majeur. En effet, elle ne permet de traiter que les générations totalement éteintes, ce qui restreint considérablement son domaine d'application. Il est ainsi illusoire de recourir à une telle technique en cas d'intérêt prononcé pour les âges faibles ou les années récentes.

Pour plus de précisions quant à cette méthode, voyez MESLÉ & VALLIN (2002). À l'aide de données françaises très anciennes, ces auteurs proposent notamment une comparaison de cette technique avec celle de Vallin.

## 2.4 Répartition des décès dans l'année

### 2.4.1 Répartition uniforme des décès sur l'année

La plupart du temps, les données dont dispose l'actuaire sont le fruit d'un groupement par période. Ainsi, il connaît le nombre de décès  $D_x$  recensés sur une période parmi les  $L_x$  individus d'un certain âge  $x$  en début de période. En l'absence d'information quant à la répartition des décès durant l'année, il faut recourir à des hypothèses de travail.

Lors des travaux empiriques, il est souvent supposé que les décès sont uniformément distribués entre deux âges entiers  $x$  et  $x+1$  (en moyenne, les décès se produisent donc en milieu d'année). Cette hypothèse est acceptable pour autant que  $D_x$  est petit par rapport à  $L_x$ ; il faut donc parfois s'en méfier, surtout aux âges très faibles et très élevés. Ceci revient à considérer que les fonctions  $t \mapsto L_{x+t}$ ,  $x = 0, 1, 2, \dots$ , sont linéaires sur  $[0, 1]$ , i.e.

$$L_{x+\xi} = L_x + \xi(L_{x+1} - L_x), \quad 0 \leq \xi \leq 1.$$

Dès lors, on obtient sans peine une mesure d'exposition au risque  $E_x$ , qui représente le nombre total d'années vécues par les individus à l'âge  $x$ :

$$\begin{aligned} E_x &= \frac{1}{2}(L_x + L_{x+1}) \\ &= L_x - \frac{1}{2}(L_x - L_{x+1}) \\ &= L_x - \frac{D_x}{2} \end{aligned}$$

Nous trouvons également la durée de vie  $T_x$  d'un individu d'âge  $x$ . En notant  $\omega$  l'âge ultime auquel toute la population est éteinte, il vient

$$\begin{aligned} T_x &= \sum_{k=0}^{\omega-x-1} \int_{\xi=x+k}^{x+k+1} L_\xi d\xi \\ &= \sum_{k=0}^{\omega-x-1} \frac{L_{x+k} + L_{x+k+1}}{2} \\ &= \frac{L_x}{2} + \sum_{k=1}^{\omega-x-1} L_{x+k}. \end{aligned}$$

Ensuite, nous trouvons le taux de mortalité  $m_x$  qui représente le quotient entre le nombre de décès  $D_x$  et l'exposition au risque  $E_x$ :

$$m_x = \frac{D_x}{E_x} = \frac{D_x}{L_x - \frac{D_x}{2}} = \frac{D_x}{\frac{1}{2}(L_x + L_{x+1})}.$$

Enfin, le quotient de mortalité, division du nombre de décès  $D_x$  par l'effectif initial  $L_x$ , est fourni par

$$q_x = \frac{D_x}{L_x} = \frac{2m_x}{2 + m_x}.$$

En notant  ${}_\tau q_x$  la probabilité pour un individu d'âge  $x$  de mourir endéans  $\tau$  années, l'hypothèse de distribution uniforme des décès sur l'année est en fait équivalente à

$${}_\tau q_x = \tau q_x, \quad 0 \leq \tau \leq 1, \quad x \in \mathbb{N}. \quad (2.9)$$

De la même manière, notons  ${}_\tau D_x$  le nombre de décès subis pendant  $\tau$  années par les  $L_x$  individus d'âge  $x$ , et  ${}_\tau E_x$  et  ${}_\tau m_x$  l'exposition au risque et le taux de mortalité y correspondants.

Si on suppose  $\xi \mapsto L_\xi$  linéaire sur  $(x, x + \tau)$ ,  $\tau > 1$ , hypothèse acceptable tant que  ${}_\tau D_x$  est petit par rapport à  $L_x$ , on a

$$\begin{aligned} {}_\tau E_x &= \frac{t}{2}(L_x + L_{x+\tau}) \\ {}_\tau m_x &= \frac{2{}_\tau D_x}{t(L_x + L_{x+\tau})}. \end{aligned}$$

Comme  $L_{x+\tau} = L_x(1 - {}_\tau q_x)$ ,  ${}_\tau m_x$  et  ${}_\tau q_x$  sont approximativement liés par les formules

$$\begin{aligned} {}_\tau m_x &\approx \frac{2{}_\tau q_x}{\tau(2 - {}_\tau q_x)} \\ {}_\tau q_x &\approx \frac{2{}_\tau m_x}{2 + {}_\tau m_x}. \end{aligned}$$

Enfin, les taux instantanés de mortalité  $\mu_x$  à l'âge  $x$  sont définis par

$$\mu_{x+\tau} = -\frac{d}{d\tau} \ln {}_\tau p_x.$$

Dès lors, sous l'hypothèse (2.9), ils sont donnés par

$$\mu_{x+\tau} = \frac{q_x}{1 - \tau q_x}, \quad 0 \leq \tau \leq 1.$$

Examinons maintenant brièvement les conséquences techniques de l'hypothèse (2.9). Pour ce faire, introduisons la variable aléatoire  $K_x$  mesurant le nombre d'années entières de survie d'une personne ayant atteint l'âge  $x$ , i.e.

$$K_x = \lfloor T_x \rfloor = \max\{n \in \mathbb{N} | T_x \geq n\}.$$

Définissons également  $U_x$  comme la fraction d'année que vivra cette personne au-delà de ces  $K_x$  années, i.e.

$$U_x = T_x - K_x.$$

La distribution de  $K_x$  se déduit de la table de mortalité de la manière suivante

$$\mathbb{P}[K_x = k] = \mathbb{P}[k \leq T_x < k + 1] = {}_k p_x - {}_{k+1} p_x.$$

Comme pour tous entiers  $x$  et  $k$ , et  $t \in [0,1]$  on a

$$\mathbb{P}[U_x \leq t | K_x = k] = \frac{\mathbb{P}[k < T_x \leq k + t]}{\mathbb{P}[k < T_x \leq k + 1]} = \frac{{}_k p_x (1 - t p_{x+k})}{{}_k p_x (1 - p_{x+k})} = \frac{t q_{x+k}}{q_{x+k}} = t,$$

on voit que (2.9) revient à supposer que les variables  $K_x$  et  $U_x$  sont indépendantes et que  $U_x$  est distribuée uniformément sur  $[0,1]$ , puisque

$$P[U_x \leq t | K_x = k] = P[U_x \leq t] = t, \quad 0 \leq t \leq 1.$$

### 2.4.2 Taux de mortalité constants par morceaux

Une alternative à (2.9) consiste à supposer que les taux instantanés de mortalité sont constants par morceaux, i.e.

$$\mu_{x+\tau} = \mu_x \text{ pour } 0 \leq \tau < 1 \text{ et } x \text{ entier.} \quad (2.10)$$

On vérifie facilement que dans ce cas

$$\begin{aligned} \mu_x &= -\ln(1 - q_x) \\ {}_\tau p_x &= \exp\left(-\int_{\zeta=0}^{\tau} \mu_{x+\zeta} d\zeta\right) \\ &= \exp(-\tau \mu_x) = \{p_x\}^\tau \text{ pour } 0 \leq \tau < 1 \\ E_x &= \int_{\tau=0}^1 L_{x+\tau} d\tau = L_x \int_{\tau=0}^1 {}_\tau p_x d\tau \\ &= L_x \int_{\tau=0}^1 \{p_x\}^\tau d\tau = \frac{-L_x q_x}{\ln(1 - q_x)} \\ m_x &= \frac{1}{L_x} \int_{\tau=0}^1 L_{x+\tau} \mu_{x+\tau} d\tau = \mu_x. \end{aligned}$$

Notez l'égalité entre taux et taux instantané de mortalité sous l'hypothèse (2.10).

Sous (2.10), les variables  $U_x$  et  $K_x$  ne sont pas indépendantes. En effet,

$$\mathbb{P}[U_x \leq t | K_x = k] = \frac{1 - p_{x+t}^k}{1 - p_{x+k}}.$$

Etant donné  $K_x = k$ ,  $U_x$  a une loi exponentielle tronquée en  $t = 1$  dont le paramètre  $p_{x+k}$  dépend de  $k$ . L'hypothèse (2.10) permet d'estimer sans difficulté les probabilités de survie. En effet, si les taux instantanés de mortalité sont constants par morceaux, i.e.

$$\mu_x = \begin{cases} c_1 \text{ pour } x \in [0, x_1[ \\ c_2 \text{ pour } x \in [x_1, x_2[ \\ \vdots \\ c_{k-1} \text{ pour } x \in [x_{k-2}, x_{k-1}[ \\ c_2 \text{ pour } x \geq x_{k-1} \end{cases}$$

on obtient alors

$${}_x p_0 = \begin{cases} \exp\{-c_1 x\} \text{ pour } x \in [0, x_1[ \\ \exp\{-c_1 x_1 - c_2(x - x_1)\} \text{ pour } x \in [x_1, x_2[ \\ \vdots \\ \exp\{-c_1 x_1 - c_2(x_2 - x_1) - \dots - c_k(x - x_{k-1})\} \text{ pour } x \geq x_{k-1}. \end{cases}$$

Il est intéressant de noter que lorsque  ${}_k q_x \rightarrow 0$ , (2.10) redonne (2.9). Sous (2.10),  $U_x$  est approximativement distribuée selon la loi uniforme sur  $[0, 1]$  et indépendante de  $K_x$  lorsque  ${}_k q_x$  est petit.

### 2.4.3 Comparaison des hypothèses

Le choix d'une des hypothèses (2.9) et (2.10) n'est pas innocent, comme le montre le résultat suivant. Notons  $T_x^{(1)}$  et  $T_x^{(2)}$  la durée de vie restante d'un assuré d'âge  $x$  sous les hypothèses (2.9) et (2.10), respectivement. Si  $k = \lfloor t \rfloor$  et  $\epsilon = t - k$ , on a quel que soit l'âge  $x$  entier

$$\begin{aligned} \mathbb{P}[T_x^{(1)} \leq t] &= 1 - \mathbb{P}[T_x^{(1)} > k + \epsilon] \\ &= 1 - \mathbb{P}[K_x = k] \mathbb{P}[U_x > \epsilon | K_x = k] - \mathbb{P}[K_x \geq k + 1] \\ &= 1 - ({}_k p_x - {}_{k+1} p_x)(1 - \epsilon) - {}_{k+1} p_x \\ &= 1 - {}_k p_x(1 - \epsilon) + {}_{k+1} p_x(1 - \epsilon) - {}_{k+1} p_x \\ &= 1 - {}_k p_x(1 - \epsilon) - \epsilon {}_{k+1} p_x \\ &= 1 - {}_k p_x(1 - \epsilon + \epsilon p_{x+k}) \\ &= 1 - {}_k p_x(1 - \epsilon q_{x+k}). \end{aligned}$$

De la même façon, quel que soit  $x$  entier,

$$\mathbb{P}[T_x^{(2)} \leq t] = 1 - {}_k p_x (p_{x+k})^\epsilon.$$

Par ailleurs, HÜRLIMANN (1990) a également prouvé que

$$\mathbb{P}[T_x^{(2)} > t] \leq \mathbb{P}[T_x^{(1)} > t] \quad \forall t \in \mathbb{R}^+.$$

Interprétons à présent ces résultats, en nous basant sur HÜRLIMANN (1990), corrigé par KLING & WOLTHUIS (1992) et KLING (1993). Pour cela distinguons les opérations en cas de vie et en cas de décès. L'hypothèse de constance des taux de mortalité est celle qui donne la plus petite durée de survie. Pour des opérations en cas de vie, c'est donc celle qui engendrera la prime unique pure la plus faible et elle représentera la démarche la moins conservatrice. Ensuite vient l'interpolation linéaire qui est donc plus sûre (puisqu'elle fournit une prime unique plus élevée). Pour des opérations en cas de décès par contre, c'est l'hypothèse de constance des taux qui engendre la prime unique pure la plus élevée (car les assurés ont tendance à vivre moins longtemps sous cette hypothèse). C'est donc elle qui représente la démarche la plus conservatrice.

Bien que nous nous intéressons en priorité aux rentes viagères, nous avons retenu l'hypothèse de constance par morceaux des taux de mortalité dans les applications numériques. Ce choix est lié aux techniques de projection de la mortalité utilisées par la suite. En pratique, l'impact réel du passage d'une hypothèse à l'autre sur les primes uniques pures des rentes est très limité.

#### 2.4.4 Généralisation

Dans le cadre de l'élaboration de tables de mortalité prospectives, nous avons besoin d'une généralisation à deux indices de l'hypothèse (2.10) de constance des taux instantanés de mortalité par âge. Il est naturel de considérer les taux  $\mu_x(t)$  constants dans chaque carré du diagramme de Lexis, représenté à la Figure 1.4. Formellement, cela revient à supposer

$$\mu_{x+\xi}(t+\tau) = \mu_x(t) \quad \text{pour } 0 \leq \xi, \tau < 1 \text{ et } x \text{ et } t \text{ entiers.} \quad (2.11)$$

Le taux  $\mu_x(t)$  représente la mortalité subie à chaque instant par un individu d'âge  $x$  au temps  $t$ . Ce taux est donc supposé constant pendant chaque période successive d'un an.

Indexons  $p_x$  et  $q_x$  par le temps calendaire de la même manière que  $\mu_x$ . L'hypothèse (2.11) fournit alors

$$p_x(t) = 1 - q_x(t) = \exp(-\mu_x(t)).$$

De plus, en notant  ${}_x p_x(t)$  la probabilité pour une personne d'âge  $x$  au temps  $t$  de survivre pendant au moins  $\xi$  années, nous trouvons une mesure d'exposition

au risque pour les individus d'âge  $x$  durant l'année  $t$ :

$$\begin{aligned} E_{xt} &= \int_{\xi=0}^1 L_{x+\xi, t+\xi} d\xi \\ &= L_{xt} \int_{\xi=0}^1 \xi p_x(t) d\xi \\ &= L_{xt} \int_{\xi=0}^1 (p_x(t))^\xi d\xi \\ &= \frac{-L_{xt} q_x(t)}{\ln(1 - q_x(t))}. \end{aligned}$$

## Chapitre 3

# Modèle log-bilinéaire pour l'élaboration de tables de mortalité prospectives

### 3.1 Introduction

La première étape de toute projection de mortalité consiste à réduire la dimensionnalité des données. Il est en effet impossible de traiter simultanément autant de séries chronologiques décrivant l'évolution au cours du temps des quotients de mortalité ou des taux de mortalité aux différents âges.

Une première approche, très appréciée par les actuaires, consiste à ajuster les observations de chaque année à l'aide d'un modèle paramétrique (Makeham, par exemple); voyez par exemple BENJAMIN & POLLARD (1993) ou BENJAMIN & SOLIMAN (1993). Ceci permet de condenser l'information annuelle dans un petit nombre de paramètres (3 dans notre exemple). Ensuite, l'évolution au cours du temps de ces paramètres est à son tour modélisée, afin de fournir des projections de la mortalité dans le futur. Il ne faut néanmoins pas perdre de vue que, chaque année, ces paramètres seront vraisemblablement fort corrélés, ayant été estimés sur base des mêmes données. On ne peut donc pas se contenter d'une modélisation univariée mais on doit recourir à un modèle de série temporelle multivariée. Ce point de vue est malheureusement trop souvent ignoré dans la littérature. De plus, la pertinence de cette approche dépend bien évidemment de la bonne spécification du modèle paramétrique. Si ce dernier s'avère erroné, cela compromet gravement les projections qui en découlent.

Dans ce chapitre, nous présenterons la méthode de LEE & CARTER (1992), qui a fait ses preuves en démographie et a été présentée aux actuaires par LEE (2000). L'idée est de passer par une décomposition en valeurs singulières de la matrice des taux de mortalité (doublement indexés, par l'âge et le temps calendaire). La matrice initiale sera ainsi approximée au rang 1 par un produit de deux vecteurs propres: l'un d'entre eux traduira l'effet de l'âge, et l'autre l'effet du temps calendaire. Il suffira alors d'extrapoler dans le futur le vecteur



décrivant l'évolution temporelle pour en déduire des tables de mortalité prospectives. Cette projection se fera à l'aide d'un modèle ARIMA, en suivant la méthodologie désormais classique de Box et Jenkins. Nous serons alors à même de calculer des indicateurs de mortalité tels que l'espérance de vie ou la valeur actuelle de rentes viagères. Cette dernière permet notamment d'apprécier le coût réel des promesses de pension dans le cadre de la sécurité sociale ou de tarifier des produits d'assurance.

## 3.2 Modèle de Lee et Carter

### 3.2.1 Modèle

La méthode de Lee et Carter a déjà prouvé son efficacité à de maintes reprises, et sur diverses populations (USA cf. LEE & CARTER (1992), Canada cf. LEE & NAULT (1993), Chili cf. LEE & ROFMAN (1994), Japon cf. WILMOTH (1996), Belgique cf. LOUIS (2001) et BROUHNS & DENUIT (2001b)). Elle consiste à décomposer la mortalité en deux composantes, l'une propre à l'âge et l'autre tendancielle, et ensuite à extrapoler celle relative au temps. Il est bon de noter d'emblée que la méthode de Lee et Carter possède les avantages et les inconvénients de l'objectivité: elle n'incorpore pas d'avis d'expert sur l'évolution présumée de la mortalité, sur les progrès de la médecine, l'apparition de nouvelles maladies ou encore l'évolution du style de vie. Vu les avis parfois diamétralement opposés des experts, nous considérons cela comme un avantage. La méthode se borne donc à extrapoler dans le futur les tendances constatées dans le passé.

Nous présentons en fait une variante de la méthode originale de LEE & CARTER (1992) suggérée par LEE (2000), laquelle a été légèrement modifiée et appliquée avec succès par RENSHAW & HABERMAN (2001). L'idée est ici de décomposer l'estimation brute  $\hat{\mu}_x(t)$  de  $\mu_x(t)$  comme suit sur l'échelle logarithmique:

$$\ln \hat{\mu}_x(t) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt} \quad (3.1)$$

où les erreurs  $\epsilon_{xt}$  sont supposées centrées, indépendantes et de même variance  $\sigma^2$  (hypothèse d'homoscédasticité).

Rappelons qu'un modèle est identifiable lorsqu'il n'est pas possible de trouver deux jeux de paramètres distincts conduisant au même modèle. Ainsi, (3.1) n'est a priori pas identifiable puisqu'il est insensible à certaines modifications des paramètres, comme par exemple

$$\begin{aligned} \beta_x &\leftarrow c\beta_x \\ \kappa_t &\leftarrow \frac{\kappa_t}{c} \quad \forall c \in \mathbb{R}, c \neq 0 \end{aligned}$$

ou

$$\begin{aligned} \alpha_x &\leftarrow \alpha_x - c\beta_x \\ \kappa_t &\leftarrow \kappa_t + c \quad \forall c \in \mathbb{R}. \end{aligned}$$

Des contraintes sur les paramètres doivent donc venir compléter le modèle (3.1). Lee et Carter proposent de fixer la valeur des sommes des  $\beta_x$  et des  $\kappa_t$ :

$$\sum_{x=x_{\min}}^{x_{\max}} \beta_x = 1 \quad (3.2)$$

$$\sum_{t=t_{\min}}^{t_{\max}} \kappa_t = 0. \quad (3.3)$$

Ce choix n'est pas totalement innocent. Nous verrons notamment que (3.3) permet une interprétation facile des  $\alpha_x$ .

Néanmoins, (3.2) n'est pas toujours adoptée dans la littérature. Pour des raisons de calcul, on lui préfère parfois

$$\sum_{x=x_{\min}}^{x_{\max}} \beta_x^2 = 1$$

ou

$$\beta_{x_{\min}} = 1.$$

En vertu de (3.1), le taux de mortalité à l'âge  $x$  pour l'année  $t$  est donc décomposé sur l'échelle logarithmique, à un terme d'erreur près, en la somme d'une composante spécifique à l'âge  $x$  et d'un produit entre un paramètre temporel décrivant l'évolution générale de la mortalité et un paramètre propre à l'âge décrivant l'évolution du taux à l'âge  $x$  par rapport à ceux relatifs aux autres âges. On espère bien entendu que la variance des erreurs  $\epsilon_{xt}$  sera aussi petite que possible. Ainsi, la majeure partie de la variance des  $\mu_x(t)$  à chaque âge  $x$  sera expliquée par le paramètre  $\kappa_t$ , les  $\epsilon_{xt}$  n'étant plus qu'un bruit blanc.

Il est bon de saisir précisément la signification des différentes composantes de la décomposition (3.1):

$\alpha_x$ : décrit le comportement moyen des taux instantanés de mortalité au cours du temps. Plus précisément,  $\exp(\alpha_x)$  est la moyenne géométrique des  $\mu_x(t)$  puisque

$$\begin{aligned} \prod_{t=t_{\min}}^{t_{\max}} \mu_x(t) &= \exp\left(\alpha_x(t_{\max} - t_{\min} + 1) + \beta_x \sum_{t=t_{\min}}^{t_{\max}} \kappa_t\right) \\ &= \exp(\alpha_x(t_{\max} - t_{\min} + 1)) \end{aligned}$$

en vertu de (3.3) si bien que

$$\alpha_x = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t=t_{\min}}^{t_{\max}} \ln \mu_x(t).$$

$\beta_x$ : décrit (toujours sur l'échelle logarithmique) l'écart des  $\mu_x(t)$  par rapport au comportement moyen  $\alpha_x$  puisque

$$\frac{d}{dt} \ln \mu_x(t) = \beta_x \frac{d}{dt} \kappa_t.$$

Si les  $\kappa_t$  sont (approximativement) linéaires,  $d\kappa_t/dt$  est constant et les taux de mortalité  $\mu_x(t)$  présentent une décroissance exponentielle à un taux constant.

$\kappa_t$ : décrit l'évolution de la mortalité au cours du temps.

L'avantage de cette méthode est de réduire le problème à l'étude de la série chronologique univariée  $\{\kappa_t, t = 1, 2, \dots\}$ .

On se référera à LEE & CARTER (1992), BELL (1997), LEE (2000) et LEE & MILLER (2000) pour une description complète de la méthode. Il est utile de souligner que cette méthode qui a été appliquée avec succès à des situations très variées, a aussi ses détracteurs parmi les actuaires; voyez par exemple GUTTERMAN & VANDERHOOF (1999) pour un regard critique à ce sujet.

### 3.2.2 Estimation des paramètres par moindres carrés: analyse en composantes principales

Il est clair que le modèle (3.1) ne peut pas être ajusté par une simple régression linéaire, puisqu'il n'y a pas de variable observable dans le membre de droite. L'estimation des paramètres s'effectue par la méthode des moindres carrés ordinaires, c'est-à-dire en résolvant le problème d'optimisation

$$(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t) = \arg \min_{\alpha_x, \beta_x, \kappa_t} \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \left( \ln \hat{\mu}_x(t) - \alpha_x - \beta_x \kappa_t \right)^2. \quad (3.4)$$

Une solution de (3.4) existe et est fournie par la méthode de décomposition aux valeurs singulières. Cette solution est unique grâce aux contraintes (3.2) et (3.3).

#### Etape 1: Estimation des $\alpha_x$

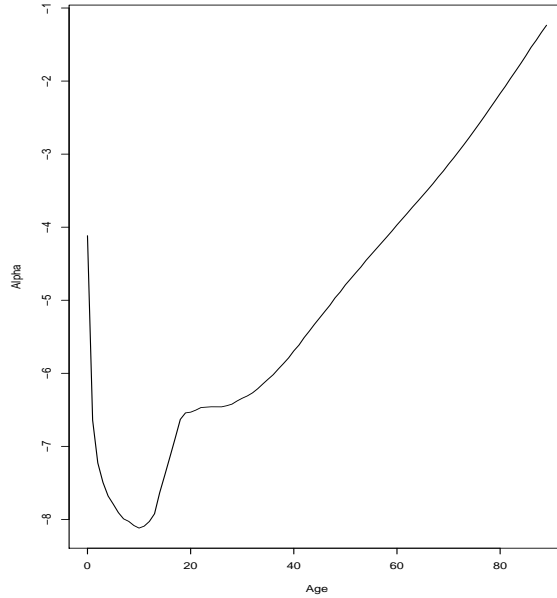
Les  $\alpha_x$  sont estimés par les moyennes des  $\ln \hat{\mu}_x(t)$  au cours du temps, i.e.

$$\hat{\alpha}_x = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t=t_{\min}}^{t_{\max}} \ln \hat{\mu}_x(t)$$

puisque l'annulation de la dérivée de (3.4) par rapport à  $\alpha_x$  donne

$$\sum_{t=t_{\min}}^{t_{\max}} \ln \hat{\mu}_x(t) = (t_{\max} - t_{\min} + 1) \alpha_x.$$

La Figure 3.1 représente cette estimation à partir des données masculines françaises. S'agissant de l'allure générale de la mortalité suivant l'âge, on y


 FIG. 3.1 – Estimation des  $\alpha_x$  du modèle (3.1).

retrouve les phénomènes habituellement observables. La courbe, relativement élevée chez les nouveaux-nés et les nourrissons, décroît rapidement avec l'âge pour atteindre son minimum absolu vers l'âge de dix ans. Survient alors un pic de mortalité pudiquement appelé bosse-accident. Cette bosse, qui touche les jeunes d'une vingtaine d'années, est en fait essentiellement composée de suicides et d'accidents. Ensuite, les logarithmes moyens des taux instantanés de mortalité augmentent quasiment linéairement avec l'âge.

### Étape 2: Estimation des $\beta_x$ et $\kappa_t$

Désignons par  $\mathbf{Z}$  la matrice de dimensions  $(x_{\max} - x_{\min} + 1) \times (t_{\max} - t_{\min} + 1)$  dont l'élément  $(x, t)$ , noté  $z_{xt}$ , est donné par

$$z_{xt} = \ln \hat{\mu}_x(t) - \hat{\alpha}_x.$$

Il s'agit donc d'un centrage des  $\ln \hat{\mu}_x(t)$  par rapport à leur moyenne temporelle.

L'idée est d'approximer  $\mathbf{Z}$  par le produit d'une matrice colonne et d'une matrice ligne

$$\mathbf{Z} \approx \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\kappa}}' \quad (3.5)$$

avec  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{x_{\min}}, \dots, \hat{\beta}_{x_{\max}})'$  et  $\hat{\boldsymbol{\kappa}} = (\hat{\kappa}_{t_{\min}}, \dots, \hat{\kappa}_{t_{\max}})'$  de manière optimale au sens des moindres carrés ordinaires, i.e. en minimisant

$$\sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} (z_{xt} - \beta_x \kappa_t)^2.$$

La résolution de ce problème est directe en passant par la décomposition en valeurs singulières de  $\mathbf{Z}$ .

Notons  $\mathbf{u}_i$  le  $i^{\text{ème}}$  vecteur propre normé de la matrice carrée  $\mathbf{Z}'\mathbf{Z}$  de dimensions  $(t_{\max} - t_{\min} + 1) \times (t_{\max} - t_{\min} + 1)$  correspondant à la valeur propre  $\lambda_i$ ; ceci signifie que  $\mathbf{u}_i$  satisfait les conditions

$$\begin{aligned}\mathbf{Z}'\mathbf{Z}\mathbf{u}_i &= \lambda_i\mathbf{u}_i \\ \mathbf{u}_i'\mathbf{u}_i &= 1.\end{aligned}$$

En prémultipliant les deux membres de la première égalité par  $\mathbf{Z}$ , on obtient

$$(\mathbf{Z}\mathbf{Z}')\mathbf{Z}\mathbf{u}_i = \lambda_i(\mathbf{Z}\mathbf{u}_i).$$

Cette dernière relation montre qu'à tout vecteur propre  $\mathbf{u}_i$  de  $\mathbf{Z}'\mathbf{Z}$  relatif à une valeur propre  $\lambda_i$  non nulle correspond un vecteur propre  $\mathbf{Z}\mathbf{u}_i$  de  $\mathbf{Z}\mathbf{Z}'$ , relatif à la même valeur propre  $\lambda_i$ . Ainsi, toutes les valeurs propres de  $\mathbf{Z}'\mathbf{Z}$  et  $\mathbf{Z}\mathbf{Z}'$  sont égales (avec le même ordre de multiplicité, le cas échéant). Si on note  $\mathbf{v}_i$  le  $i^{\text{ème}}$  vecteur propre de la matrice carrée  $\mathbf{Z}\mathbf{Z}'$  de dimensions  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ , correspondant à la valeur propre  $\lambda_i$ , on a alors pour  $\lambda_i \neq 0$ ,

$$\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{Z}\mathbf{u}_i$$

ou encore

$$\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{Z}'\mathbf{v}_i.$$

Partons à présent de la relation

$$\mathbf{Z}\mathbf{u}_i = \mathbf{v}_i\sqrt{\lambda_i}$$

et postmultiplions les deux membres de cette égalité par  $\mathbf{u}_i'$ , avant de sommer sur toutes les valeurs propres de  $\mathbf{Z}'\mathbf{Z}$

$$\mathbf{Z} \left( \sum_{i \geq 1} \mathbf{u}_i\mathbf{u}_i' \right) = \sum_{i \geq 1} \sqrt{\lambda_i}\mathbf{v}_i\mathbf{u}_i'.$$

Les vecteurs propres  $\mathbf{u}_i$  étant orthogonaux et de norme 1,

$$\sum_{i \geq 1} \mathbf{u}_i\mathbf{u}_i' = \mathcal{I}_{x_{\max} - x_{\min} + 1}$$

avec  $\mathcal{I}_n$  la matrice identité de dimensions  $n \times n$ , de sorte qu'on aboutit à la décomposition suivante de la matrice  $\mathbf{Z}$ :

$$\mathbf{Z} = \sum_{i \geq 1} \sqrt{\lambda_i}\mathbf{v}_i\mathbf{u}_i'.$$

Cette dernière formule est appelée décomposition aux valeurs singulières. Elle garantit que, sous des conditions assez générales, une matrice rectangulaire

peut être écrite de façon unique comme une somme optimale de matrices de rang 1 (c'est-à-dire de produits d'une matrice ligne par une matrice colonne). L'optimalité dont il est question signifie que la première matrice de rang 1 constitue la meilleure approximation de rang 1 de la matrice initiale (au sens des moindres carrés), que la somme des deux premières constituent la meilleure approximation de rang 2, etc.

Si la valeur propre  $\lambda_1$  domine nettement les autres, on obtient alors l'approximation

$$\mathbf{Z} \approx \sqrt{\lambda_1} \mathbf{v}_1 \mathbf{u}'_1. \quad (3.6)$$

On mesure la qualité de l'approximation (3.6) par le taux d'inertie  $\tau_1$  (ou pourcentage de variance expliqué), donné par

$$\tau_1 = \frac{\lambda_1}{\sum_{i \geq 1} \lambda_i}. \quad (3.7)$$

En confrontant (3.5) à (3.6), on voit qu'il suffit de prendre

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \frac{\mathbf{v}_1}{\sum_j v_{1j}} \\ \hat{\boldsymbol{\kappa}} &= \sqrt{\lambda_1} \left( \sum_j v_{1j} \right) \mathbf{u}_1 \end{aligned}$$

en supposant  $\sum_j v_{1j} \neq 0$ . Il est évident que la contrainte (3.2) est satisfaite par les  $\hat{\beta}_x$ . De plus, les  $\hat{\kappa}_t$  satisfont eux aussi la contrainte (3.3) puisque  $\sum_j u_{1j} = 0$  du fait que  $\sum_t z_{xt} = 0 \forall x \in \{x_{\min}, \dots, x_{\max}\}$ .

Signalons que la technique de la décomposition en valeurs singulières n'est applicable qu'à des tableaux de données exactement rectangulaires et dépourvus de la moindre valeur manquante. Cette méthode ne peut donc être appliquée à des données dont l'âge maximum traité n'est pas constant sur la période d'observation.

La Figure 3.2 représente les courbes des  $\hat{\beta}_x$  et  $\hat{\kappa}_t$  pour les données françaises. On y remarque une mortalité décroissante avec le temps. Si cette tendance est chaotique entre 1950 et 1970, une courbe fortement linéaire se distingue par après. Comme le montre la courbe des  $\hat{\beta}_x$ , le déclin de la mortalité n'est pas identique pour tous les âges et se fait surtout ressentir chez les très jeunes enfants. Au contraire, vers vingt ans,  $\hat{\beta}_x$  avoisine zéro et la chute temporelle de la mortalité est donc beaucoup moins sensible.

### Etape 3: Réajustement des $\hat{\kappa}_t$

À ce stade du développement, nous possédons déjà des estimateurs  $\hat{\alpha}_x$ ,  $\hat{\beta}_x$  et  $\hat{\kappa}_t$ . Néanmoins, en traitant directement les estimations brutes des taux instantanés de mortalité  $\hat{\mu}_x(t)$ , nous ne faisons que peu de cas de la qualité

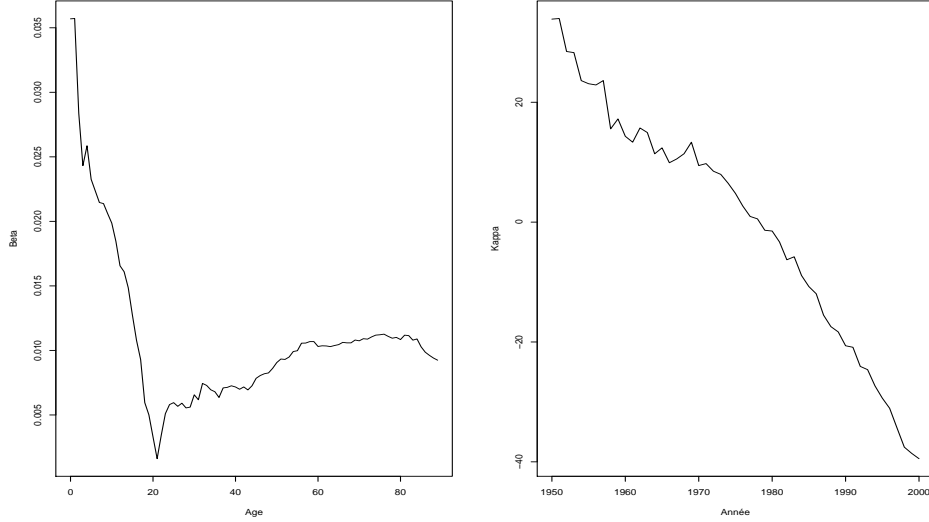


FIG. 3.2 – Estimation des  $\beta_x$  et  $\kappa_t$  du modèle (3.1) avec contraintes (3.2)-(3.3) par la décomposition en valeurs singulières.

de l'ajustement du point de vue des nombres de décès. Il est donc conseillé d'effectuer une étape supplémentaire en exigeant par exemple que

$$\sum_{x=x_{\min}}^{x_{\max}} D_{xt} = \sum_{x=x_{\min}}^{x_{\max}} E_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t). \quad (3.8)$$

Le modèle reproduirait donc fidèlement le nombre total de décès observés lors d'une année donnée. Il convient donc de réestimer le paramètre temporel  $\kappa_t$ .

Ce réajustement présente également un second avantage. Si nous ne disposons pas de données par âges pour certaines années  $t$ , mais seulement du nombre total de décès  $\sum_x D_{xt}$ , nous pouvons quand même en déduire une valeur de l'indice  $\kappa_t$  grâce aux  $\alpha_x$  et  $\beta_x$  estimés sur base des autres années calendrier.

Les nouveaux estimateurs  $\hat{\kappa}_t$  sont donc solutions des équations (3.8), qui peuvent être réécrites

$$\begin{aligned} 0 &= \sum_{x=x_{\min}}^{x_{\max}} \left[ E_{xt} \exp(\hat{\alpha}_x) \exp(\hat{\beta}_x \hat{\kappa}_t) - D_{xt} \right] \\ &= \sum_{x=x_{\min}}^{x_{\max}} \left[ K_{xt} \exp(\hat{\beta}_x \hat{\kappa}_t) \right] - K_t \end{aligned}$$

où  $K_{xt}$  et  $K_t$  sont des constantes qui valent

$$\begin{aligned} K_{xt} &= E_{xt} \exp(\widehat{\alpha}_x) \\ K_t &= \sum_{x=x_{\min}}^{x_{\max}} D_{xt}. \end{aligned}$$

Sous l'hypothèse (souvent observée en pratique) que tous les  $\widehat{\beta}_x$  sont de même signe, l'unicité de la racine de chaque fonction est évidente. Dès lors, une méthode itérative telle que Newton-Raphson donnera de bons résultats en peu d'itérations. Pour rappel, cette technique se base sur l'itération suivante :

$$\kappa_t^{(\nu+1)} = \kappa_t^{(\nu)} - \frac{F(\kappa_t^{(\nu)})}{\left. \frac{\partial F}{\partial \kappa_t} \right|_{\kappa_t = \kappa_t^{(\nu)}}} \quad (3.9)$$

où  $\kappa_t \mapsto F(\kappa_t)$  est la fonction à annuler. Dans notre situation, l'initialisation de la récurrence s'impose d'elle-même : nous prenons  $\kappa_t^{(0)} = \widehat{\kappa}_t$ .

Néanmoins, dans les cas critiques où on observe des valeurs de  $\widehat{\beta}_x$  de signes différents, les équations (3.8) peuvent présenter 0, 1 ou 2 solutions. Le réajustement est alors impossible en pratique. Cette situation peut se présenter lorsque l'on traite les très grands âges. Dans ces cas, il n'est pas rare de voir apparaître des  $\widehat{\beta}_x$  négatifs au-delà de 90 ou 100 ans.

Bien entendu, la contrainte (3.3) est violée par les nouveaux estimateurs  $\widehat{\kappa}_t$  issus de l'algorithme itératif; une transformation s'impose donc:

$$\begin{aligned} \widehat{\kappa}_t &\leftarrow \widehat{\kappa}_t - \bar{\kappa} \\ \widehat{\alpha}_x &\leftarrow \widehat{\alpha}_x + \widehat{\beta}_x \bar{\kappa} \end{aligned}$$

où  $\bar{\kappa}$  est la moyenne des  $\widehat{\kappa}_t$  de sorte que les taux instantanés de mortalité projetés restent indépendants de la transformation.

L'impact de cette réestimation est visible à la Figure 3.3.



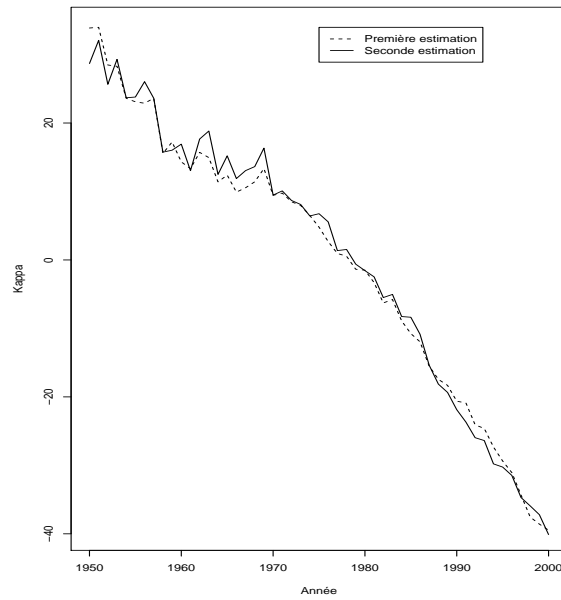


FIG. 3.3 – Réestimation des  $\kappa_t$  de manière à satisfaire (3.8).

### 3.2.3 Estimation des paramètres par moindres carrés: méthode itérative

L'estimation des paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  peut également être réalisée par un algorithme itératif du type Newton-Raphson (3.9). Cette méthode se base sur les itérations suivantes:

$$\begin{aligned}\widehat{\alpha}_x^{(\nu+1)} &= \widehat{\alpha}_x^{(\nu)} + \frac{\sum_{t=t_{\min}}^{t_{\max}} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)})}{t_{\max} - t_{\min} + 1} \\ \widehat{\kappa}_t^{(\nu+1)} &= \widehat{\kappa}_t^{(\nu)} + \frac{\sum_{x=x_{\min}}^{x_{\max}} \widehat{\beta}_x^{(\nu)} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu+1)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)})}{\sum_{x=x_{\min}}^{x_{\max}} (\widehat{\beta}_x^{(\nu)})^2} \\ \widehat{\beta}_x^{(\nu+1)} &= \widehat{\beta}_x^{(\nu)} + \frac{\sum_{x=x_{\min}}^{x_{\max}} \widehat{\kappa}_t^{(\nu+1)} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu+1)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu+1)})}{\sum_{t=t_{\min}}^{t_{\max}} (\widehat{\kappa}_t^{(\nu+1)})^2}.\end{aligned}\quad (3.10)$$

Cet algorithme nécessite une initialisation. L'attribution d'une valeur nulle à  $\widehat{\beta}_x^{(0)}$  engendrerait une division par 0 dans le calcul de  $\widehat{\kappa}_t^{(1)}$ . Ormis ce cas particulier, le choix des paramètres initiaux est libre. Il ne reste alors plus qu'à fixer un critère d'arrêt, par exemple une borne inférieure pour la diminution relative à chaque itération de la somme des carrés des erreurs. Formellement, cela revient à arrêter l'algorithme à l'itération  $\nu$  lorsque

$$\frac{e^{(\nu-1)} - e^{(\nu)}}{e^{(\nu-1)}} < \varepsilon$$

avec

$$e^{(\nu)} = \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)})$$

et  $\varepsilon$  la borne inférieure pour la diminution relative de l'erreur  $e^{(\nu)}$ .

Les itérations (3.10) peuvent être réécrites sous forme matricielle:

$$\begin{aligned}(\mathbf{C}_\alpha^{(\nu)}) \widehat{\boldsymbol{\alpha}}^{(\nu+1)} &= (\mathbf{C}_\alpha^{(\nu)}) \widehat{\boldsymbol{\alpha}}^{(\nu)} + \mathbf{r}_\alpha^{(\nu)} \\ (\mathbf{C}_\kappa^{(\nu)}) \widehat{\boldsymbol{\kappa}}^{(\nu+1)} &= (\mathbf{C}_\kappa^{(\nu)}) \widehat{\boldsymbol{\kappa}}^{(\nu)} + \mathbf{r}_\kappa^{(\nu)} \\ (\mathbf{C}_\beta^{(\nu)}) \widehat{\boldsymbol{\beta}}^{(\nu+1)} &= (\mathbf{C}_\beta^{(\nu)}) \widehat{\boldsymbol{\beta}}^{(\nu)} + \mathbf{r}_\beta^{(\nu)}\end{aligned}\quad (3.11)$$

avec  $\mathbf{C}_\alpha^{(\nu)}$ ,  $\mathbf{C}_\kappa^{(\nu)}$  et  $\mathbf{C}_\beta^{(\nu)}$  des matrices diagonales respectivement de dimensions  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ ,  $(t_{\max} - t_{\min} + 1) \times (t_{\max} - t_{\min} + 1)$  et  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ , et dont les éléments valent

$$\begin{aligned}(\mathbf{C}_\alpha^{(\nu)})_{xx} &= t_{\max} - t_{\min} + 1 \\ (\mathbf{C}_\kappa^{(\nu)})_{tt} &= \sum_{x=x_{\min}}^{x_{\max}} (\widehat{\beta}_x^{(\nu)})^2 \\ (\mathbf{C}_\beta^{(\nu)})_{xx} &= \sum_{t=t_{\min}}^{t_{\max}} (\widehat{\kappa}_t^{(\nu+1)})^2\end{aligned}\quad (3.12)$$

et  $\mathbf{r}_\alpha^{(\nu)}$ ,  $\mathbf{r}_\kappa^{(\nu)}$  et  $\mathbf{r}_\beta^{(\nu)}$  des vecteurs-colonnes respectivement de dimensions  $(x_{\max} - x_{\min} + 1)$ ,  $(t_{\max} - t_{\min} + 1)$  et  $(x_{\max} - x_{\min} + 1)$ , et dont les éléments valent

$$\begin{aligned} (\mathbf{r}_\alpha^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} (\ln \hat{\mu}_x(t) - \hat{\alpha}_x^{(\nu)} - \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu)}) \\ (\mathbf{r}_\kappa^{(\nu)})_t &= \sum_{x=x_{\min}}^{x_{\max}} \hat{\beta}_x^{(\nu)} (\ln \hat{\mu}_x(t) - \hat{\alpha}_x^{(\nu+1)} - \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu)}) \\ (\mathbf{r}_\beta^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} \hat{\kappa}_t^{(\nu+1)} (\ln \hat{\mu}_x(t) - \hat{\alpha}_x^{(\nu+1)} - \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu+1)}). \end{aligned} \quad (3.13)$$

Les estimateurs issus de cet algorithme ne satisfont naturellement pas les contraintes (3.2)-(3.3). Ils nécessitent donc la transformation suivante, qui laisse inchangées les valeurs de  $\exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$ :

$$\begin{aligned} \hat{\alpha}_x &\leftarrow \hat{\alpha}_x + \hat{\beta}_x \bar{\kappa} \\ \hat{\kappa}_t &\leftarrow (\hat{\kappa}_t - \bar{\kappa}) \hat{\beta}_\bullet \\ \hat{\beta}_x &\leftarrow \frac{\hat{\beta}_x}{\hat{\beta}_\bullet} \end{aligned}$$

avec  $\hat{\beta}_\bullet$  la somme des  $\hat{\beta}_x$  initiaux et  $\bar{\kappa}$  la moyenne des  $\hat{\kappa}_t$  initiaux.

L'un des avantages de cette méthode est de ne pas nécessiter de tableaux de données exactement rectangulaires, comme c'était le cas pour la décomposition en valeurs singulières. Néanmoins, il faut garder à l'esprit que ce schéma itératif ne remplace que les étapes 1 et 2 de la section précédente. La troisième étape, qui consistait en la réestimation de la tendance temporelle pour reproduire fidèlement le nombre total de décès observés chaque année, peut donc être maintenue.

### 3.2.4 Estimation des paramètres par moindres carrés pondérés

La méthode des moindres carrés ordinaires décrite précédemment minimise l'erreur quadratique moyenne sur les logarithmes des taux de mortalité. Ce faisant, un poids identique est apporté à chacun des couples  $(x, t)$ , indépendamment de leur contribution au nombre total de décès. Pour tenir compte de l'importance relative de chaque couple, WILMOTH (1993) propose d'avoir recours au principe des moindres carrés pondérés par les nombres de décès pour l'estimation des paramètres. Ainsi, le critère des moindres carrés pondérés accordera plus d'importance aux couples dont le taux brut instantané est peu variable tandis qu'il discréditera quelque peu les couples dont le taux est plus volatile. On minimisera alors

$$\sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} w_{xt} \left( \ln \hat{\mu}_x(t) - \alpha_x - \beta_x \kappa_t \right)^2$$

où les poids  $w_{xt}$  sont les nombres de décès observés aux différents âges au cours de la période d'observation.

Le choix de ces poids nécessite une explication. Supposons que les nombres de décès  $D_{xt}$  sont distribués selon la loi normale

$$D_{xt} \sim \mathcal{N}(\mu_x(t)E_{xt}, \mu_x(t)E_{xt}),$$

approximation du modèle  $D_{xt} \sim \mathcal{Poisson}(\mu_x(t)E_{xt})$  qui sera utilisé et justifié à la Section 3.3. Nous trouvons alors

$$\hat{\mu}_x(t) = \frac{D_{xt}}{E_{xt}} \sim \mathcal{N}\left(\mu_x(t), \frac{\mu_x(t)}{E_{xt}}\right)$$

et l'application de la méthode delta fournit

$$\log \hat{\mu}_x(t) = \log \frac{D_{xt}}{E_{xt}} \sim \mathcal{N}\left(\log \mu_x(t), \frac{1}{E_{xt}\mu_x(t)}\right).$$

Dès lors, la variance de  $\log \hat{\mu}_x(t)$  peut être estimée par

$$\text{Var}[\log \hat{\mu}_x(t)] \approx \frac{1}{E_{xt}\hat{\mu}_x(t)}.$$

Comme le poids de chaque couple  $(x, t)$  doit être inversement proportionnel à la variance de l'estimation brute  $\log \hat{\mu}_x(t)$ , nous pouvons prendre  $w_{xt} = E_{xt}\hat{\mu}_x(t)$ , i.e. le nombre observé de décès à l'âge  $x$  durant l'année  $t$ .

L'estimation des paramètres par la méthode des moindres carrés pondérés peut être réalisée de la même manière que précédemment, à l'aide du schéma itératif (3.11). Il convient alors de modifier les valeurs de (3.12) et (3.13) de

sorte que les matrices  $\mathbf{C}_\alpha^{(\nu)}$ ,  $\mathbf{C}_\kappa^{(\nu)}$  et  $\mathbf{C}_\beta^{(\nu)}$  deviennent

$$\begin{aligned}
 (\mathbf{C}_\alpha^{(\nu)})_{xx} &= \sum_{t=t_{\min}}^{t_{\max}} D_{xt} \\
 (\mathbf{C}_\kappa^{(\nu)})_{tt} &= \sum_{x=x_{\min}}^{x_{\max}} D_{xt} (\widehat{\beta}_x^{(\nu)})^2 \\
 (\mathbf{C}_\beta^{(\nu)})_{xx} &= \sum_{t=t_{\min}}^{t_{\max}} D_{xt} (\widehat{\kappa}_t^{(\nu+1)})^2.
 \end{aligned} \tag{3.14}$$

De même, les vecteurs de résidus  $\mathbf{r}_\alpha^{(\nu)}$ ,  $\mathbf{r}_\beta^{(\nu)}$  et  $\mathbf{r}_\kappa^{(\nu)}$  prennent les valeurs

$$\begin{aligned}
 (\mathbf{r}_\alpha^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} D_{xt} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \\
 (\mathbf{r}_\kappa^{(\nu)})_t &= \sum_{x=x_{\min}}^{x_{\max}} \widehat{\beta}_x^{(\nu)} D_{xt} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu+1)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \\
 (\mathbf{r}_\beta^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} \widehat{\kappa}_t^{(\nu+1)} D_{xt} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu+1)} - \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu+1)}).
 \end{aligned} \tag{3.15}$$

Les Figures 3.4 et 3.5 comparent les nouveaux estimateurs avec ceux issus de la méthode des moindres carrés ordinaires. Des différences sensibles entre les deux types d'estimateurs y apparaissent.

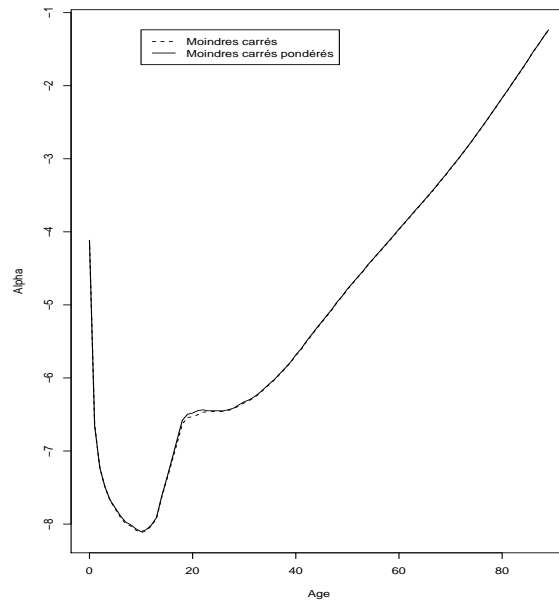


FIG. 3.4 – Estimation des  $\alpha_x$  par la méthode des moindres carrés pondérés.

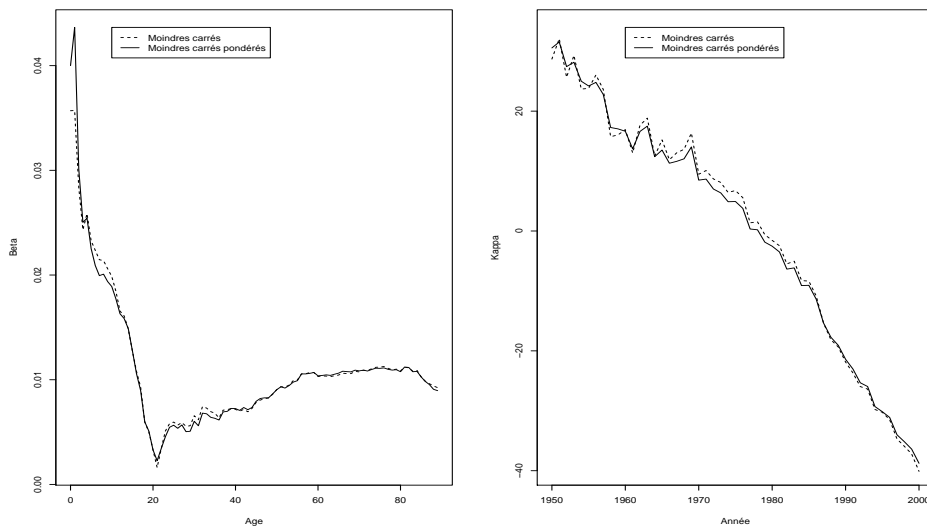


FIG. 3.5 – Estimation des  $\beta_x$  et  $\kappa_t$  par la méthode des moindres carrés pondérés.

### 3.2.5 Estimation des paramètres par moindres carrés pénalisés

Dans de nombreuses situations, les courbes obtenues pour les estimateurs  $\widehat{\alpha}_x$ ,  $\widehat{\beta}_x$  et  $\widehat{\kappa}_t$  ne sont pas lisses. Ce phénomène peut être naturel pour le cas de la tendance temporelle  $\widehat{\kappa}_t$ . Il est en effet raisonnable de penser que certaines années sont plus meurtrières que d'autres (en temps de guerre par exemple). Néanmoins, l'extrapolation du paramètre temporel est généralement initiée par sa dernière valeur estimée  $\widehat{\kappa}_{t_{\max}}$ . Si cette valeur est atypique au regard de l'allure générale de la courbe, l'extrapolation risque d'être biaisée.

Les discontinuités des deux autres séries de paramètres engendrent des problèmes d'un autre type. Tout d'abord, il est difficilement compréhensible que la courbe des  $\widehat{\alpha}_x$ , censée représenter le comportement moyen de la mortalité avec l'âge, présente des aspérités. Tout au plus nous attendons-nous à observer les phénomènes habituels, comme les mortalités infantile et juvénile, la bosse accident, voire un comportement plus curieux chez les vieillards. Toutefois, la pratique nous offre habituellement des courbes de  $\widehat{\alpha}_x$  très lisses. Il n'est donc souvent pas nécessaire de recourir à un lissage de ces paramètres.

Les  $\widehat{\beta}_x$  conditionnent pour leur part toute l'extrapolation des taux instantanés de mortalité dans le futur. Ainsi, une courbe des  $\widehat{\beta}_x$  irrégulière conjuguée à des  $\widehat{\kappa}_t$  de plus en plus importants avec le temps engendre des taux instantanés futurs de mortalité très erratiques. Cette remarque est illustrée à la Figure 3.13, où la série temporelle a déjà été extrapolée dans le futur.

Le problème est donc double. Tout d'abord, de telles discontinuités ne trouvent pas d'origine logique. Ensuite, elles engendrent des conséquences non négligeables.

Il devient donc essentiel d'élaborer une solution alternative pour l'estimation des paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$ . Une solution empirique consisterait en un lissage pur et simple des estimateurs considérés précédemment. Il existe de nombreuses méthodes pour effectuer une telle opération.

Néanmoins, nous proposons de remédier à ce problème de manière plus élégante, en adoptant le critère des moindres carrés pénalisés. Par ce dernier, les nouveaux estimateurs doivent minimiser l'expression

$$\sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} (\ln \widehat{\mu}_x(t) - \alpha_x - \beta_x \kappa_t)^2 + \boldsymbol{\alpha}' \mathbf{P}_\alpha \boldsymbol{\alpha} + \boldsymbol{\beta}' \mathbf{P}_\beta \boldsymbol{\beta} + \boldsymbol{\kappa}' \mathbf{P}_\kappa \boldsymbol{\kappa} \quad (3.16)$$

où  $\mathbf{P}_\alpha = \pi_\alpha \mathbf{D}'_\alpha \mathbf{D}_\alpha$ ,  $\mathbf{P}_\beta = \pi_\beta \mathbf{D}'_\beta \mathbf{D}_\beta$  et  $\mathbf{P}_\kappa = \pi_\kappa \mathbf{D}'_\kappa \mathbf{D}_\kappa$  sont des matrices de pénalité et  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  des paramètres de lissage.

Si les matrices  $\mathbf{D}_\alpha$ ,  $\mathbf{D}_\beta$  et  $\mathbf{D}_\kappa$  peuvent être de structures très diverses, leur détermination n'en reste pas moins essentielle. Une technique régulièrement utilisée, à la fois simple et efficace, consiste à choisir les matrices de différentiation  $\mathbf{D}^{(i)}$  de dimensions  $(n-i) \times n$  où  $n$  est le nombre de paramètres (donc le nombre d'âges dans le modèle si on traite les  $\widehat{\alpha}_x$  ou les  $\widehat{\beta}_x$ , ou le nombre

d'années pour les  $\widehat{\kappa}_t$ ):

$$\mathbf{D}^{(1)} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

$$\mathbf{D}^{(2)} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

La matrice de différentiation du premier ordre  $\mathbf{D}^{(1)}$  pénalisera donc les écarts entre deux paramètres successifs,  $\beta_x$  et  $\beta_{x+1}$  par exemple. Il en découlera donc des courbes de paramètres plus proches d'une constante, et donc plus plates.  $\mathbf{D}^{(2)}$  pénalisera pour sa part les écarts du second ordre, et engendrera donc des séries de paramètres plus lisses. Elle constitue donc pour notre problème une solution intéressante. C'est elle que nous utiliserons dans notre application numérique.

Il ne nous reste donc plus qu'à choisir les paramètres de lissage  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$ . Ceux-ci influencent directement la puissance du lissage. Des valeurs importantes engendreront des courbes de paramètres très lisses, mais plus forcément très pertinentes. Au contraire, l'attribution de valeurs nulles nous ramène au problème original des moindres carrés. Il convient donc de choisir ces paramètres avec soin.

Le procédé de sélection des paramètres de lissage  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  est basé sur la cross-validation. Définissons des erreurs de prédiction  $\epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}$  de la manière suivante:

$$\epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa} = \ln \widehat{\mu}_x(t) - \widehat{\alpha}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) - \widehat{\beta}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) \widehat{\kappa}_{t, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) \quad (3.17)$$

avec  $\widehat{\alpha}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$ ,  $\widehat{\beta}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$  et  $\widehat{\kappa}_{t, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$  les estimateurs au sens des moindres carrés pénalisés par les paramètres  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  du modèle duquel on a supprimé le couple de données  $(x, t)$ . Les paramètres de lissage optimaux  $\widehat{\pi}_\alpha$ ,  $\widehat{\pi}_\beta$  et  $\widehat{\pi}_\kappa$  sont alors solutions du problème d'optimisation

$$(\widehat{\pi}_\alpha, \widehat{\pi}_\beta, \widehat{\pi}_\kappa) = \arg \min_{\pi_\alpha, \pi_\beta, \pi_\kappa} \epsilon_{\bullet\bullet, \pi_\alpha, \pi_\beta, \pi_\kappa}^2 \quad (3.18)$$

avec

$$\epsilon_{\bullet\bullet, \pi_\alpha, \pi_\beta, \pi_\kappa}^2 = \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^2.$$

En pratique, ce problème d'optimisation est très complexe. Nous ne pouvons que calculer la somme du carré des erreurs de prédiction pour un certain nombre de valeurs prédéfinies de  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$ . Les estimateurs  $\widehat{\pi}_\alpha$ ,  $\widehat{\pi}_\beta$  et  $\widehat{\pi}_\kappa$  sont alors les paramètres de la liste qui minimisent cette somme.



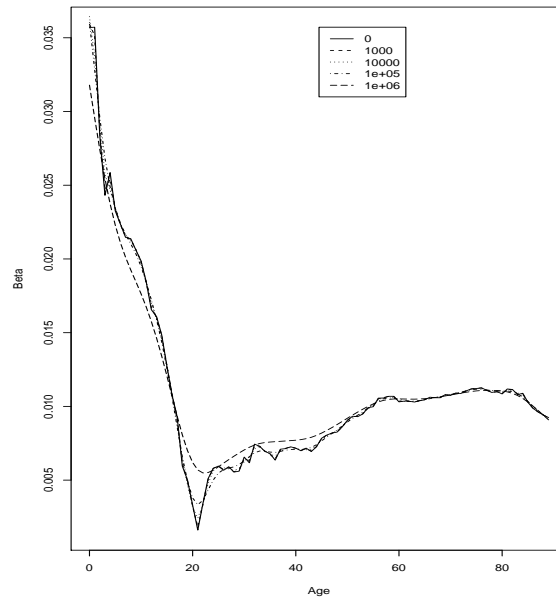


FIG. 3.6 – Estimation des  $\beta_x$  par la méthode des moindres carrés pénalisés pour quelques valeurs de  $\pi_\beta$ .

Appliquons cette méthode aux données françaises en ne considérant qu'un lissage des  $\hat{\beta}_x$ , donc en choisissant  $\pi_\alpha = \pi_\kappa = 0$ . La Figure 3.6 montre le lissage progressif de la courbe des  $\hat{\beta}_x$  pour des valeurs croissantes de  $\pi_\beta$ . Remarquez que cette opération n'est pas sans conséquences sur les estimateurs  $\hat{\alpha}_x$  et  $\hat{\kappa}_t$ , comme le montre la Figure 3.7.

L'opération de cross-validation est très coûteuse puisque, pour chaque valeur de  $\pi_\beta$ , il convient d'effectuer  $(x_{\max} - x_{\min} + 1)(t_{\max} - t_{\min} + 1)$  ajustements, lesquels ne peuvent être réalisés que par l'algorithme itératif (3.10) puisqu'un couple  $(x, t)$  de données est à chaque fois manquant. Nous trouvons les résultats suivants:

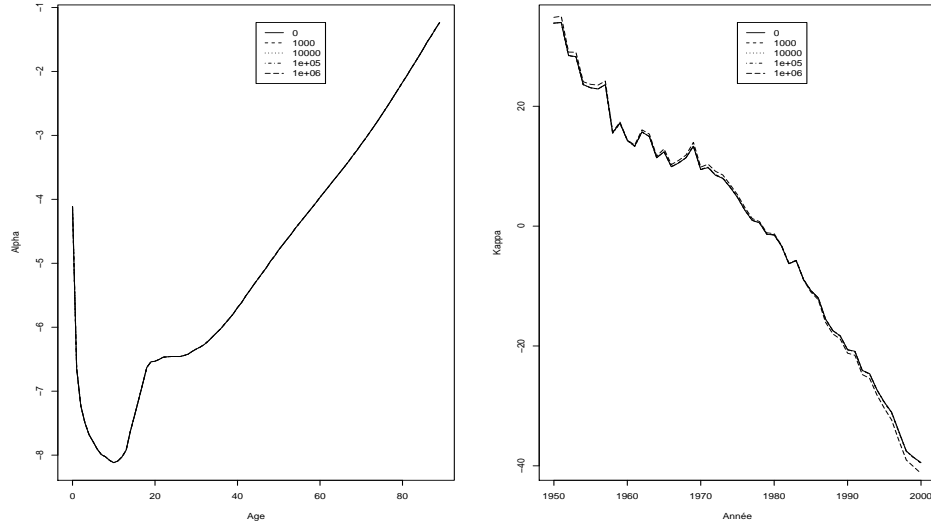


FIG. 3.7 – Estimation des  $\alpha_x$  et  $\kappa_t$  par la méthode des moindres carrés pénalisés pour quelques valeurs de  $\pi_\beta$ .

$\pi_\beta$	$\epsilon_{\bullet,0,\pi_\beta,0}^2$
0	51.242
1	51.241
$10^1$	51.236
$10^2$	51.188
$10^3$	50.876
$10^4$	50.348
$2 \cdot 10^4$	50.246
$4 \cdot 10^4$	50.172
<b><math>8 \cdot 10^4</math></b>	<b>50.137</b>
$10^5$	60.620
$10^6$	94.020

et le paramètre optimal est donc approximé par  $\hat{\pi}_\beta \approx 8 \cdot 10^4$ .

### 3.2.6 Mesure de la qualité de l'ajustement

#### Taux d'inertie

Un premier indicateur de la qualité de l'ajustement du modèle (3.1) réside dans le taux d'inertie (3.7). Ce pourcentage, qui permet de déterminer la pertinence de l'approximation (3.6), peut également servir comme aide à la décision pour l'incorporation ou non d'un second terme bilinéaire tel que décrit dans la section 3.2.9. Cette mesure n'est cependant disponible que dans le cas d'une estimation des paramètres par la méthode de la décomposition en valeurs singulières.

Dans le cas des données françaises, nous trouvons  $\tau_1 = 86.64\%$ ; l'approximation (3.6) est donc satisfaisante.

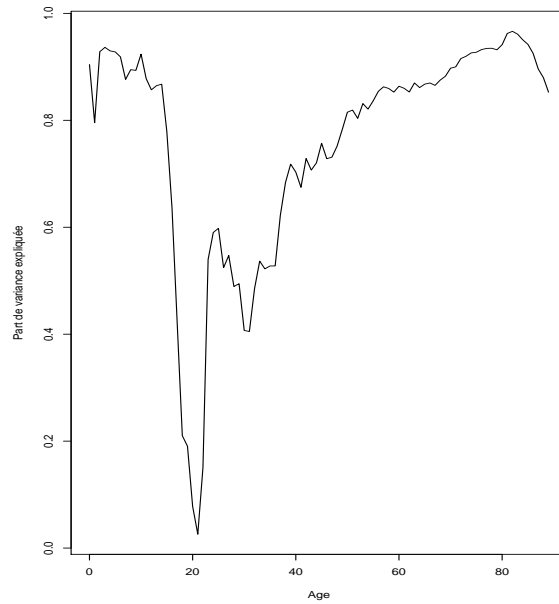
#### Variance temporelle

Il est également possible de déterminer la part de la variance temporelle à l'âge  $x$  expliquée par le modèle. À cette fin, nous utiliserons l'expression

$$1 - \frac{\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} (\widehat{\Delta}_x(t))^2 - \left(\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} \widehat{\Delta}_x(t)\right)^2}{\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} (\widehat{\mu}_x(t))^2 - \left(\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} \widehat{\mu}_x(t)\right)^2} \quad (3.19)$$

où  $\widehat{\Delta}_x(t) = \widehat{\mu}_x(t) - \exp(\widehat{\alpha}_x + \widehat{\beta}_x \widehat{\kappa}_t)$ .

L'application de cette formule aux données françaises est donnée par la Figure 3.8. On y constate des parts de variance expliquées par le modèle supérieures à 80%, hormis entre 20 et 45 ans, âges auxquels la qualité de l'ajustement se détériore fortement.

FIG. 3.8 – *Part de variance temporelle expliquée par le modèle (3.1).*

### 3.2.7 Analyse des résidus

Il est souvent nécessaire d'analyser les résidus d'un modèle tel que (3.1) pour vérifier son adéquation. Définissons à ce titre les résidus  $r_{xt}$ :

$$r_{xt} = \ln \hat{\mu}_x(t) - (\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t). \quad (3.20)$$

Idéalement, ces résidus ne devraient plus comporter de structure.

La Figure 3.9 illustre, pour nos données, les résidus du modèle (3.1) pour quelques âges choisis. On y remarque notamment une structure très prononcée à 0 et 20 ans. Ceci semble indiquer que (3.1) n'est pas adéquat, et nécessite un raffinement tel que celui proposé à la Section 3.2.9.

De tels graphiques poussent également à s'interroger sur la pertinence d'incorporer les jeunes âges dans le modèle (3.1). Ces derniers, qui jouissent d'une mortalité plus difficile à modéliser (comme les mortalités infantile et juvénile et la bosse accident), peuvent polluer le modèle. De plus, s'agissant de créer des tables de mortalité pour la tarification de rentes viagères, il convient de s'intéresser en priorité aux individus les plus enclins à souscrire ce genre de produits, donc âgés de plus de 50 ou 60 ans. À l'aide de données belges, DELWARDE & DENUIT (2002) apportent plus de précisions sur l'influence des âges considérés dans le modèle de Lee et Carter.

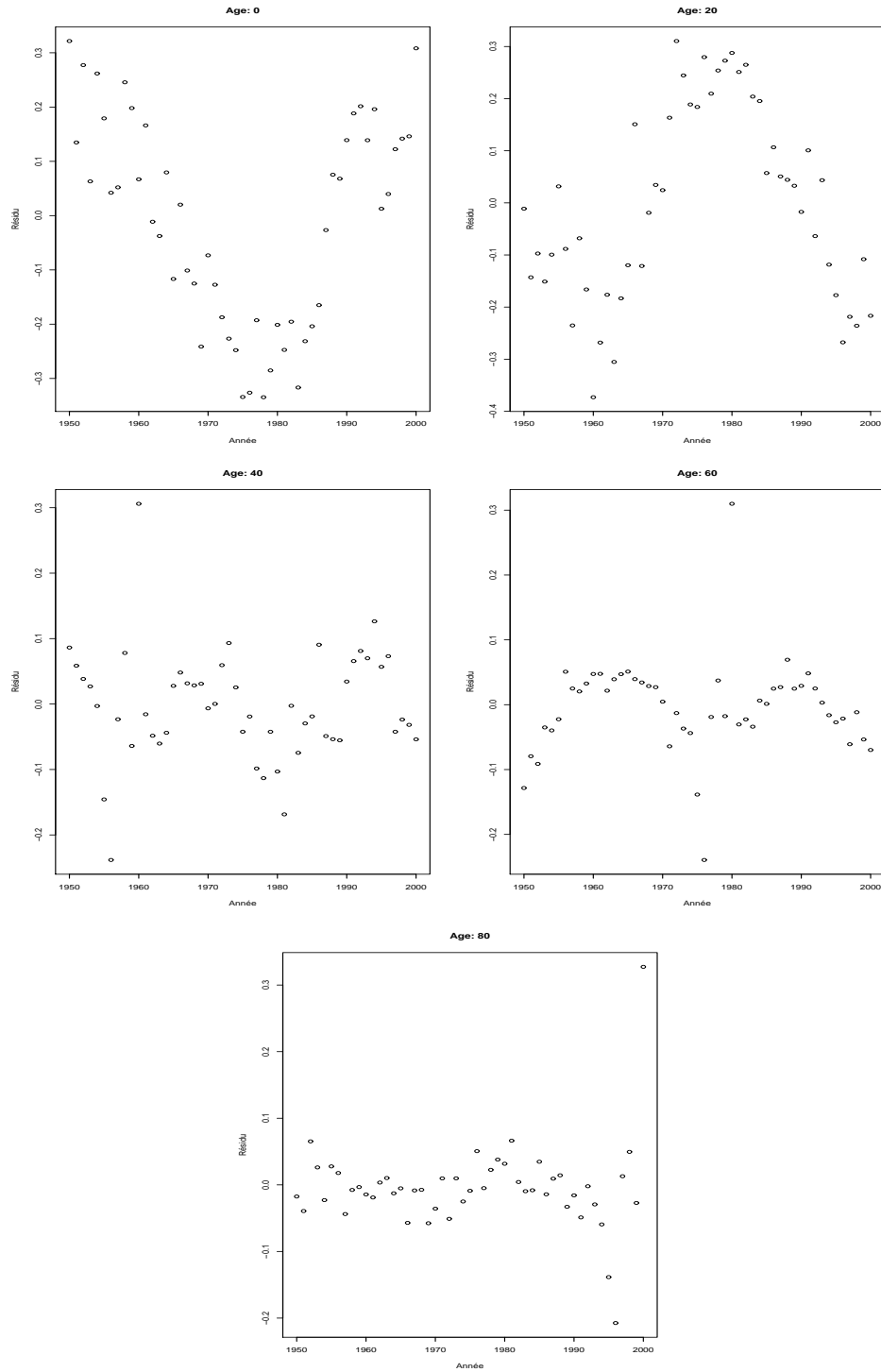


FIG. 3.9 – Résidus (3.20) pour quelques âges.

### 3.2.8 Extrapolation de la tendance temporelle

Le modèle classique de Lee et Carter présenté ci-dessus possède la propriété de synthétiser dans la série  $\{\kappa_t\}_t$  toute l'information relative à l'évolution de la mortalité dans le temps. L'objectif de cette section est donc de modéliser cette série temporelle afin d'être en mesure d'effectuer des prévisions sur les indices de mortalité des années futures. Cette manière de procéder possède les avantages et les inconvénients de l'objectivité : aucune attention n'est accordée à l'avis d'expert sur l'évolution présumée de la mortalité, sur les progrès de la médecine, l'apparition de nouvelles maladies ou encore l'évolution du style de vie. Le but est donc de se borner à extrapoler dans le futur les tendances observées dans le passé. Cette démarche a été critiquée, car elle est incapable de prévoir des variations subites de la mortalité (par exemple la découverte d'un nouveau traitement médical). Néanmoins, la mortalité étant influencée par l'interaction complexe de facteurs sociaux et biologiques difficiles à modéliser, la prédiction par extrapolation est en fait la *technique par défaut*.

L'une des spécifications les plus pertinentes pour l'extrapolation d'une série temporelle réside dans la modélisation ARIMA (pour AutoRegressive Integrated Moving Average). Adoptons les notations suivantes :

- $B$  est l'opérateur de retard, c'est-à-dire  $B\kappa_t = \kappa_{t-1}$
- $\nabla_d$  est l'opérateur de différence d'ordre  $d$ , c'est-à-dire  $\nabla_d = 1 - B^d$ .

$X_t$  est alors un processus ARIMA(p,d,q) si

$$\begin{aligned} Y_k &= \nabla_d X_k \\ Y_k - \sum_{k=1}^p a_k Y_{t-k} &= \theta + \zeta_t + \sum_{j=1}^q \theta_j \zeta_{t-j} \end{aligned}$$

avec  $\zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$  des erreurs indépendantes et identiquement distribuées.

Un tel ajustement peut être réalisé à l'aide de la désormais célèbre méthodologie de Box et Jenkins. Celle-ci nous rendra donc capable d'effectuer des projections de la mortalité dans le futur. Voyons à présent  $\{\hat{\kappa}_t, t = t_{\min}, \dots, t_{\max}\}$  comme une réalisation de dimension finie de la série chronologique  $\mathcal{K} = \{\kappa_t, t \in \mathbb{N}\}$ .

La modélisation paramétrique d'une série temporelle se construit essentiellement à partir de la structure de dépendance existant entre les éléments de la série. Celle-ci sera étudiée au moyen de la fonction d'autocovariance et de la fonction d'autocorrélation de la série. Pour une introduction aux modèles ARIMA, on pourra utilement se référer à BROCKWELL & DAVIS (1996); l'ouvrage de BROCKWELL & DAVIS (1991) offre un traitement plus théorique de la question.

Nous suivons ici pas à pas les étapes de la méthodologie de Box & Jenkins.

1. **Stationnarisation de la série.** La méthode proposée par Box et Jenkins s'applique à des processus stochastiques faiblement stationnaires d'ordre 2. Ceci signifie l'existence des deux premiers moments, la constance de

la moyenne et de la variance au cours du temps et des fonctions d'autocovariance et d'autocorrélation qui ne dépendent que du délai et pas de l'instant. La première étape de la modélisation consiste donc à s'assurer que  $\mathcal{K}$  satisfait bien ces conditions; dans le cas contraire, il s'agira de transformer  $\mathcal{K}$  adéquatement afin d'atteindre la stationnarité faible d'ordre 2.

**1.(a) Stationnarisation en variance.** La première condition nécessaire à l'ajustement d'un modèle ARIMA est la stabilité en variance de la série considérée. Il faut donc que la variabilité soit homogène sur toute la période d'observation.

**1.(b) Stationnarisation en moyenne.** La méthodologie de Box et Jenkins exige également des séries stationnaires en moyenne, c'est-à-dire dépourvues de tendance temporelle.

La méthode des différences itérées consiste à appliquer l'opérateur  $\nabla_1$  à la série initiale  $\mathcal{K}$  autant de fois que nécessaire jusqu'à l'annulation de sa tendance.

Les séries annuelles présentent régulièrement une non-stationnarité (i.e. une croissance ou une décroissance), mais sont généralement dépourvues de saisonnalité. Dans ce cas, il convient d'appliquer autant de différentiations que nécessaire pour obtenir la stationnarité en moyenne. En ce sens, il est utile de recourir à des tests formels comme ceux de Dickey-Fuller ou de Phillips-Perron, qui permettent d'éprouver l'hypothèse de non-stationnarité selon plusieurs aspects (notamment l'existence d'une tendance).

**2. Identification du modèle.** L'identification du modèle nécessite l'examen attentif des graphiques des fonctions d'autocorrélation et d'autocorrélation partielle de la série (éventuellement modifiée pour satisfaire à la stationnarité faible d'ordre 2). Si la première donne le coefficient de corrélation entre deux éléments de la série en fonction du délai qui les sépare, la seconde présente l'avantage d'éliminer l'influence des variables intermédiaires. Ces fonctions traduisent l'influence entre deux observations en fonction du délai; elles jouent donc un rôle majeur dans l'analyse de la dépendance temporelle de la série. De tels graphiques nous permettent d'apprécier la stationnarité de la série (une décroissance lente des autocorrélations partielles indique une non-stationnarité). Il est également possible de recourir à des tests éprouvant la nullité des premières autocorrélations.

En cas de doute, il peut être fait appel à des critères comme le *AIC* (pour Akaike Information Criterion), ou le *BIC* (pour Bayesian Information Criterion) qui pénalise davantage que le premier la surparamétrisation. Le modèle présentant le plus faible critère sera ainsi préféré.

**3. Estimation des paramètres.** L'estimation des paramètres par la méthode du maximum de vraisemblance (ou des moindres carrés) conditionnelle peut être réalisée par bon nombre de logiciels traitant de séries chrono-



giques. Une fois celle-ci réalisée, il convient de vérifier que les paramètres sont significativement non nuls. Ce test généralement fourni par les logiciels sus-mentionnés.

- 4. Analyse des résidus.** Pour que le modèle sélectionné soit jugé correct, il reste à vérifier que les résidus  $y$  relatifs sont la réalisation d'un bruit blanc. Il existe à cet effet un test global éprouvant la nullité des corrélations des résidus.

Finalement, nos modèles peuvent encore être améliorés par une modélisation des résidus de type (G)ARCH. Cette dernière s'impose lorsque les séries initiales adoptent un comportement hétéroscédastique: il existe des périodes de petites variations et d'autres durant lesquelles les changements sont plus importants. Les résidus peuvent alors présenter une dépendance quadratique indécélable dans les tests de corrélation. Ainsi, il est utile de jeter un oeil sur les QQ-plots normaux des résidus du modèle.

- 5. Prévision.** Finalement, nous pouvons projeter les valeurs de  $\mathcal{K}$  dans le futur. Il convient ici d'insister sur le fait que, en général, le faible nombre d'années d'observation ne nous autorise normalement pas à effectuer des projections à un horizon lointain.

En pratique, même si des modèles plus sophistiqués semblent parfois préférables, il est souvent fait usage d'une marche aléatoire avec tendance, soit le modèle ARIMA(0,1,0). Ce fut notamment le cas dans l'application de la méthode de Lee et Carter sur les populations américaine, canadienne, chilienne et japonaise. Formellement, ce modèle s'écrit

$$\widehat{\kappa}_t = \widehat{\kappa}_{t-1} + \theta + \zeta_t \quad (3.21)$$

avec  $\theta$  le paramètre de tendance et  $\zeta_t \sim \mathcal{N}(0, \sigma_\zeta^2)$  des erreurs indépendantes et identiquement distribuées.

L'estimation des paramètres par la méthode du maximum de vraisemblance fournit

$$\begin{aligned} \widehat{\theta} &= \frac{\widehat{\kappa}_{t_{\max}} - \widehat{\kappa}_{t_{\min}}}{t_{\max} - t_{\min}} \\ \widehat{\sigma}_\zeta^2 &= \frac{1}{t_{\max} - t_{\min}} \sum_{t=t_{\min}}^{t_{\max}-1} (\widehat{\kappa}_{t+1} - \widehat{\kappa}_t - \widehat{\theta})^2 \end{aligned}$$

et la variance de l'estimateur  $\widehat{\theta}$  vaut

$$\text{Var}[\widehat{\theta}] = \frac{\sigma_\zeta^2}{t_{\max} - t_{\min}}.$$

Des prédictions de  $\widehat{\kappa}_{t_{\max}+\tau}$ ,  $\tau \in \mathbb{N}$  peuvent alors être estimées par

$$\widehat{\kappa}_{t_{\max}+\tau} = \widehat{\kappa}_{t_{\max}} + \tau \widehat{\theta} + \sum_{u=1}^{\tau} \zeta_{t_{\max}+u}.$$

Grâce à l'hypothèse d'indépendance et d'identique distribution des  $\zeta_t$ , cette expression peut se réécrire

$$\widehat{\kappa}_{t_{\max}+\tau} = \widehat{\kappa}_{t_{\max}} + \tau\widehat{\theta} + \sqrt{\tau}\zeta_{t_{\max}}$$

et il vient donc

$$\begin{aligned} \mathbb{E}[\widehat{\kappa}_{t_{\max}+\tau} | \widehat{\kappa}_{t_{\min}}, \dots, \widehat{\kappa}_{t_{\max}}] &= \widehat{\kappa}_{t_{\max}} + \tau\widehat{\theta} \\ \text{Var}[\widehat{\kappa}_{t_{\max}+\tau} | \widehat{\kappa}_{t_{\min}}, \dots, \widehat{\kappa}_{t_{\max}}] &= \tau\sigma_{\zeta}^2. \end{aligned}$$

Par ailleurs, il existe une relation étroite entre, d'une part, le modèle de Lee et Carter (3.1) associé à une tendance temporelle telle que (3.21) et d'autre part une modélisation directe des logarithmes des taux de mortalité par une marche aléatoire avec tendance sur série multivariée:

$$\begin{aligned} \ln \widehat{\boldsymbol{\mu}}_t &= \ln \widehat{\boldsymbol{\mu}}_{t-1} + \boldsymbol{\psi} + \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_t &\sim \mathcal{N}(0, \boldsymbol{\Sigma}) \end{aligned}$$

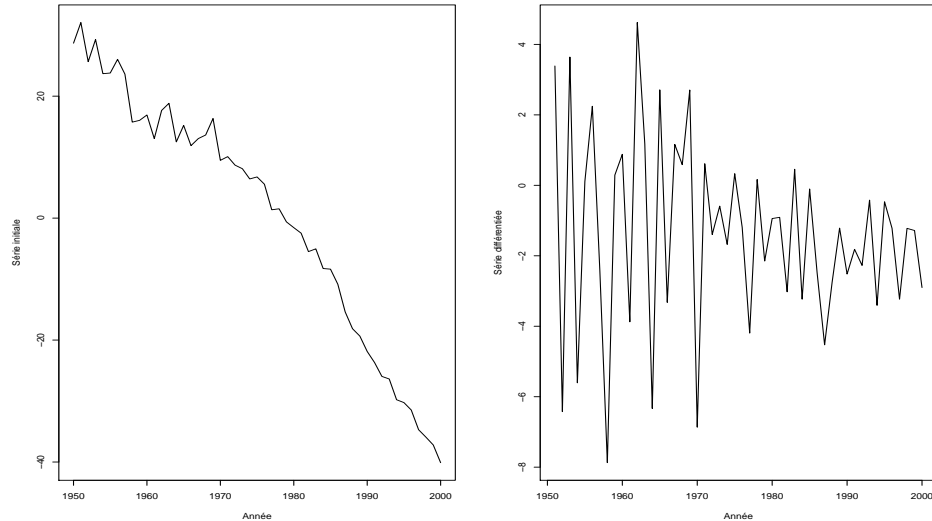
avec  $\widehat{\boldsymbol{\mu}}_t = (\widehat{\mu}_{x_{\min},t}, \dots, \widehat{\mu}_{x_{\max},t})'$ ,  $\boldsymbol{\psi}$  un vecteur de dimension  $(x_{\max} - x_{\min} + 1)$  et  $\boldsymbol{\Sigma}$  une matrice de variance-covariance de dimensions  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ . Pour plus de détails à ce sujet, voyez GIROSI & KING (2004).

La Figure 3.10 montre, pour nos données, la série initiale des  $\widehat{\kappa}_t$  ainsi que celle des différences premières  $\widehat{\kappa}_t - \widehat{\kappa}_{t-1}$ . Comme mentionné précédemment, le comportement de la mortalité entre les années 1950 et 1970 est chaotique. On y retrouve une variance nettement plus élevée que pour les années suivantes. Pour cette raison, nous choisissons de ne modéliser la série qu'à partir de 1970.

La Figure 3.11 illustre les fonctions d'aurocorrélation et d'autocorrélation partielle de la série différenciée. On y remarque une autocorrélation au délai 0 forcément égale à 1. Pour le reste, les autocorrélations et autocorrélations partielles sont non significatives au niveau 5%. Nous pouvons considérer être en présence d'un bruit blanc, et donc ajuster aux  $\widehat{\kappa}_t$  un modèle ARIMA(0,1,0). De plus, des tests globaux qui éprouvent la nullité des autocorrélations fournissent des p-valeurs toutes supérieures à 5% qui ne nous permettent pas de rejeter cette hypothèse.

Le calcul des critères *AIC* et *BIC* pour quelques modèles ARIMA, repris ci-dessous, nous informe cependant que la marche aléatoire, sans être le pire, ne semble pas pour autant être le modèle idéal. Nous lui maintenons cependant notre confiance, ne serait-ce que pour sa simplicité.

Modèle	<i>AIC</i>	<i>BIC</i>
ARIMA(0,1,0)	119.80	121.23
ARIMA(1,1,0)	117.47	120.44
ARIMA(0,1,1)	117.44	120.31
ARIMA(2,1,0)	119.29	123.59
ARIMA(0,1,2)	119.33	123.64
ARIMA(1,1,1)	119.34	123.64

FIG. 3.10 – *Série initiale et série différenciée.*

Il est donc maintenant possible d'extrapoler les  $\widehat{\kappa}_t$  au-delà de  $t_{\max}$ , comme illustré à la Figure 3.12. On y calcule également des intervalles de prédiction à 95%, lesquels doivent leur étroitesse à la quasi linéarité des  $\widehat{\kappa}_t$  entre 1970 et 2000. Cette opération permet donc de reconstituer la mortalité pour les années supérieures à  $t_{\max}$ . Ainsi, la Figure 3.13 fournit les logarithmes des taux instantanés de mortalité pour les années 1960, 1990, 2020 et 2050. On remarque ainsi l'influence prépondérante des brisures dans la courbe des  $\widehat{\beta}_x$ . Comme remarqué à la Section 3.2.5, celles-ci engendrent des taux instantanés futurs très erratiques, notamment lors des premières années de la vie.

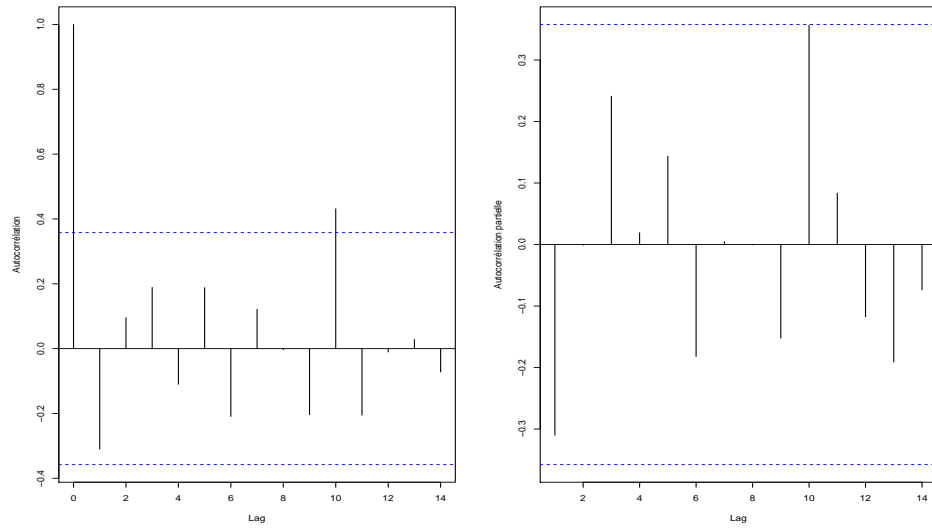


FIG. 3.11 – Autocorrélations et autocorrélations partielles.

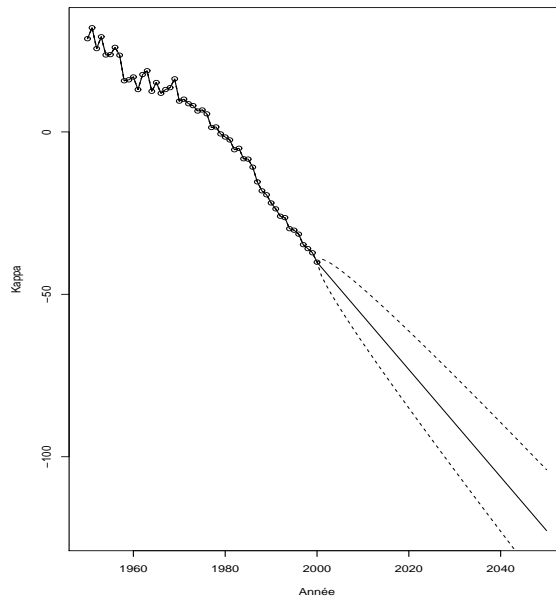


FIG. 3.12 – Prédiction des  $\hat{\kappa}_t$  à l'aide du modèle de marche aléatoire avec tendance.

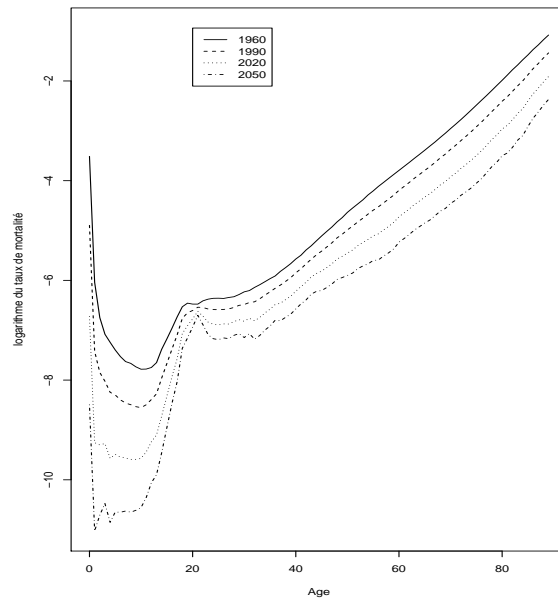


FIG. 3.13 – Prédiction des logarithmes des taux instantanés de mortalité pour quelques années à l'aide du modèle de marche aléatoire avec tendance.

### 3.2.9 Extension de Renshaw-Haberman

Comme l'ont noté RENSHAW & HABERMAN (2002), le modèle de Lee et Carter n'est pas toujours à même de décrire fidèlement les taux de mortalité de la population étudiée. Ceci se remarque sur les résidus, qui peuvent présenter une certaine structure. Il importe alors de raffiner le modèle (3.1).

Plutôt que de se limiter à une approximation au première ordre par la méthode de décomposition aux valeurs singulières, il peut être intéressant d'inclure un terme bilinéaire supplémentaire. Le modèle (3.1) devient alors

$$\ln \hat{\mu}_x(t) = \exp(\alpha_x + \beta_{x,1}\kappa_{t,1} + \beta_{x,2}\kappa_{x,2}) + \epsilon_{xt} \quad (3.22)$$

où les erreurs  $\epsilon_{xt}$  sont supposées centrées, indépendantes et de même variance  $\sigma^2$ . Les contraintes qui garantissent l'identifiabilité du modèle sont alors

$$\sum_{t=t_{\min}}^{t_{\max}} \kappa_{t,1} = \sum_{t=t_{\min}}^{t_{\max}} \kappa_{t,2} = 0 \quad (3.23)$$

$$\sum_{x=x_{\min}}^{x_{\max}} \beta_{x,1} = \sum_{x=x_{\min}}^{x_{\max}} \beta_{x,2} = 1. \quad (3.24)$$

Les  $\alpha_x$  sont toujours estimés par la moyenne des  $\ln \hat{\mu}_x(t)$  au cours du temps. Les  $\kappa_{t,j}$  et  $\beta_{x,j}$ ,  $j = 1, 2$ , sont obtenus à partir de la décomposition aux valeurs singulières de la matrice  $\mathbf{Z}$  habituelle. Il faut alors prendre le second terme de la décomposition, alors que le premier reste inchangé:

$$\begin{aligned} \kappa_{t,1} &= \kappa_t \\ \beta_{x,1} &= \beta_x. \end{aligned}$$

Seuls les  $\hat{\kappa}_{t,1}$  peuvent alors être réajustés afin de reproduire le nombre annuel total de décès.

Bien entendu, la tendance temporelle étant à présent résumée dans deux indices, une série chronologique bivariée  $\{(\kappa_{t,1}, \kappa_{t,2})\}$  doit être utilisée pour extrapoler l'évolution future de la mortalité. Cette remarque n'est pas sans importance, dans la mesure où la modélisation d'une série chronologique multivariée est beaucoup plus compliquée que dans le cas univarié.

La Figure 3.14 représente l'estimation des paramètres  $\beta_{x,1}$ ,  $\beta_{x,2}$ ,  $\kappa_{t,1}$  et  $\kappa_{t,2}$  pour le modèle (3.22) appliqué aux données françaises. On y constate que la seconde composante principale est nettement plus compliquée à interpréter que la première. Elle indique un comportement très atypique de la mortalité dans les années 70 par rapport aux premières et dernières années de la période d'observation. En outre, ce phénomène semble plus sensible chez les jeunes que chez les personnes âgées.

Bien entendu, il faut tenir compte dans cette analyse de la faiblesse de la seconde composante principale au regard de la première. En effet, si le taux d'inertie  $\tau_1$  valait 86.64%,  $\tau_2$  ne vaut plus que 4.61%, soit près de 20 fois moins.

L'approximation (3.6) est donc faiblement améliorée par l'ajout des paramètres  $\beta_{x,2}$  et  $\kappa_{t,2}$ .

L'apport en terme de part de variance temporelle expliquée est visible à la Figure 3.15. Il se fait surtout ressentir entre 15 et 40 ans, âges qui présentaient une déficience certaine dans le modèle initial à une composante principale. Cette amélioration est également visible à la Figure 3.16 qui compare les nouveaux résidus aux anciens. La structure  $y$  est moindre, même si elle n'a pas totalement disparu.

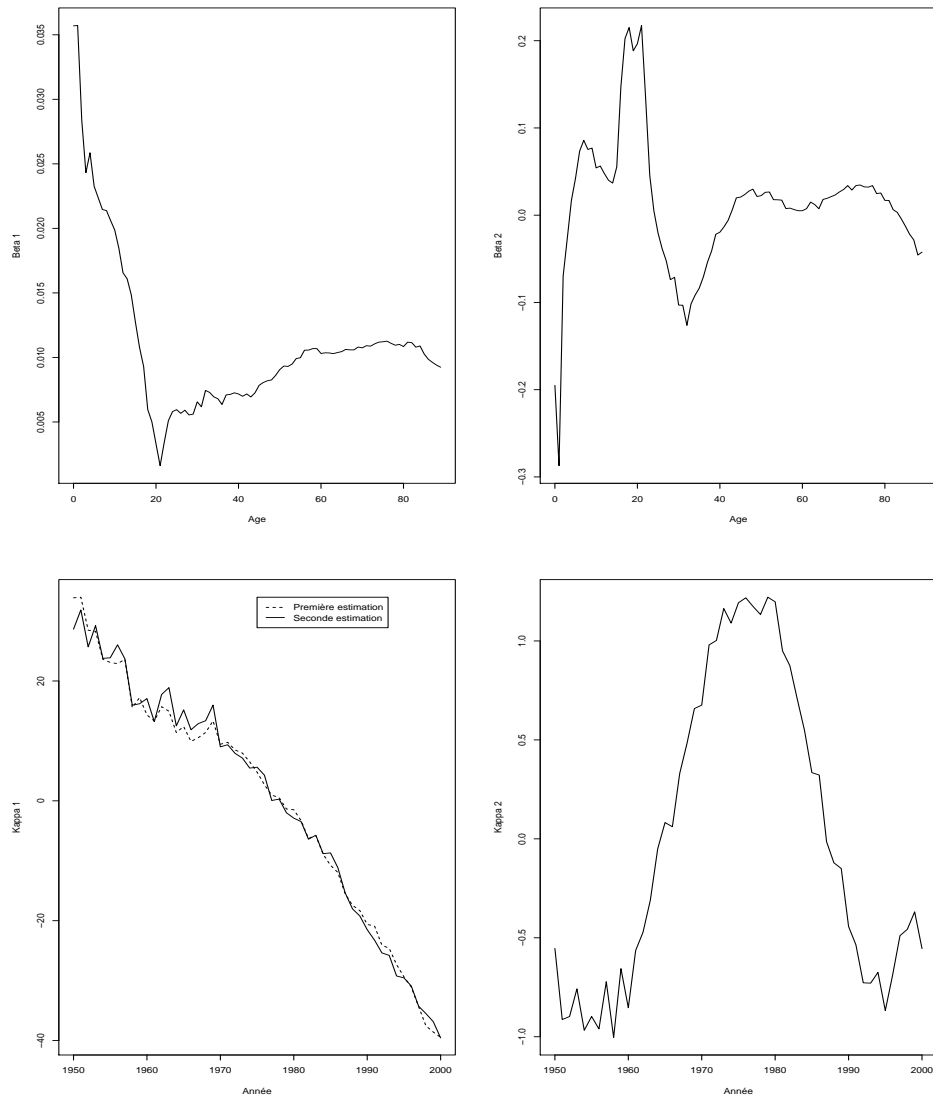
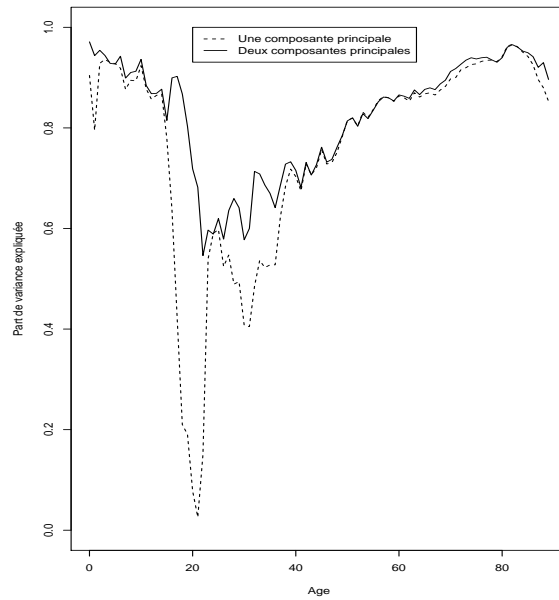


FIG. 3.14 – Estimation des  $\beta_{x,1}$  et  $\beta_{x,2}$  (en haut), et  $\kappa_{t,1}$  et  $\kappa_{t,2}$  (en bas) pour le modèle (3.22).



FIG. 3.15 – *Part de variance temporelle expliquée par le modèle (3.22).*

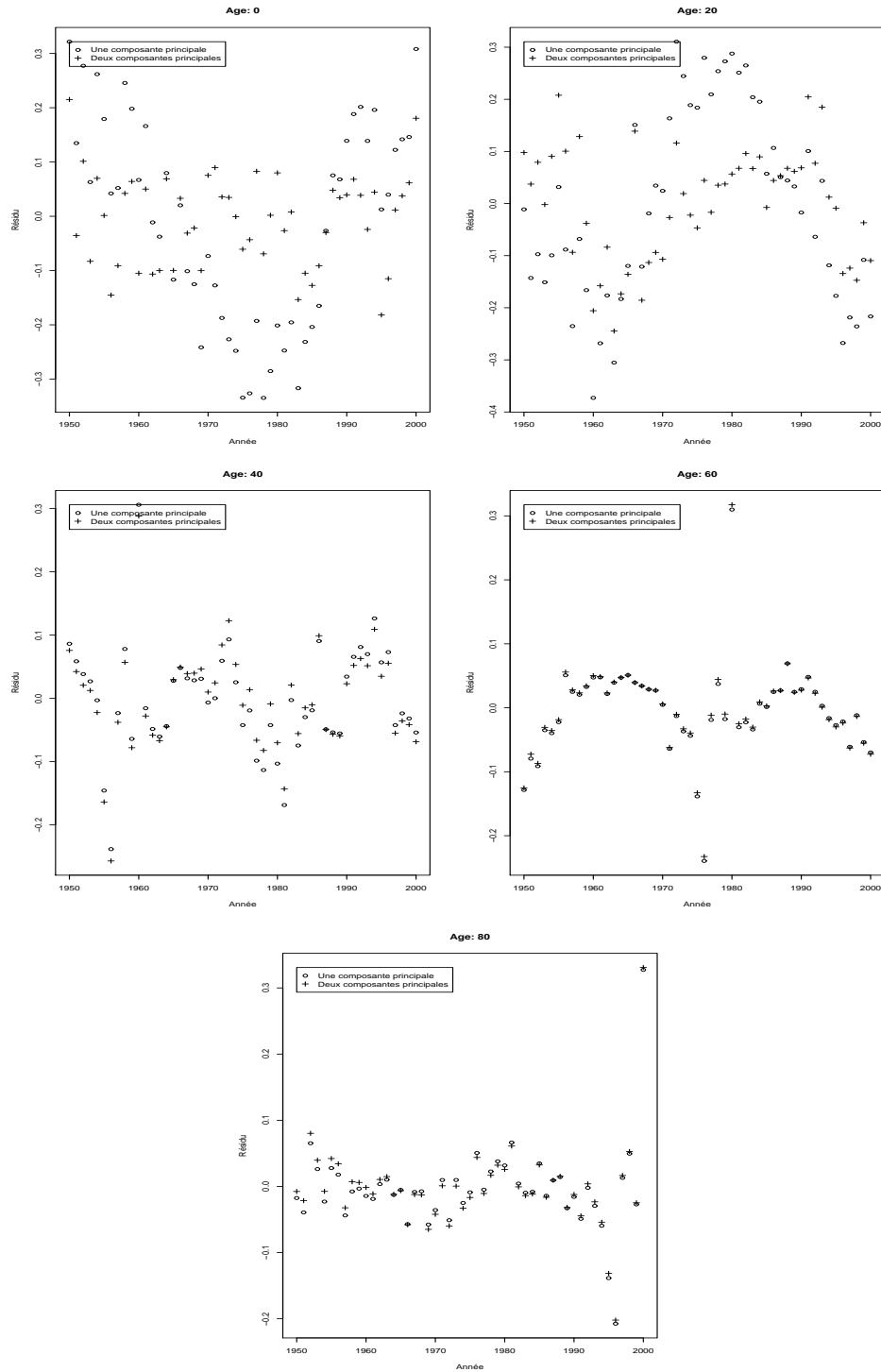


FIG. 3.16 – Résidus du modèle à deux composantes principales pour quelques âges.

### 3.2.10 Modèle avec tendance temporelle linéaire

Dans bien des pays industrialisés, l'application du modèle de Lee et Carter (3.1) fournit une courbe des  $\widehat{\kappa}_t$  fortement linéaire, spécialement à partir des années 70. En ce sens, il est utile de considérer des modèles de prédiction plus simples qu'auparavant. Deux alternatives s'offrent à nous. La première consiste à imposer le choix d'un modèle ARIMA(0,1,0) pour  $\mathcal{K}$ , i.e. retenir

$$\kappa_t = \kappa_{t-1} + \theta + \zeta_t$$

avec  $\zeta_t$  un bruit blanc de variance  $\sigma_\zeta^2$ . Les prédictions seront donc linéaires et les seuls paramètres à estimer sont  $\theta$  et  $\sigma_\zeta^2$ .

Une seconde possibilité, plus élégante, serait de modifier le modèle (3.1) pour qu'il tienne compte explicitement d'une tendance temporelle linéaire:

$$\ln \widehat{\mu}_x(t) = \alpha_x + \beta_x(a + bt) + \epsilon_{xt} \quad (3.25)$$

où les erreurs  $\epsilon_{xt}$  sont à nouveau supposées centrées, indépendantes et de même variance  $\sigma^2$ .

Notez qu'une telle spécification ne nécessite plus de contraintes telles (3.2)-(3.3), celles-ci étant implicitement remplacées par l'imposition des deux paramètres  $a$  et  $b$  de la tendance linéaire. L'alternative consiste à laisser  $a$  et  $b$  libres de manière à respecter (3.2)-(3.3).

Dans la situation simplifiée où la tendance linéaire est égale au temps calendaire  $t$ , ces coefficients prennent les valeurs  $a = 0$  et  $b = 1$  et le modèle devient

$$\ln \widehat{\mu}_x(t) = \alpha_x + \beta_x t + \epsilon_{xt} \quad (3.26)$$

Ce dernier est en fait équivalent à (3.25) puisqu'on peut retrouver (3.25) en effectuant la transformation

$$\begin{aligned} \alpha_x &\leftarrow \alpha_x + a\beta_x \\ \beta_x &\leftarrow b\beta_x. \end{aligned}$$

L'estimation des paramètres  $\alpha_x$  et  $\beta_x$  par le critère des moindres carrés est maintenant aisée à l'aide de n'importe quel logiciel traitant des modèles linéaires généralisés. Néanmoins, il est possible de recourir au même style d'algorithme itératif que précédemment. Les itérations (3.10) deviennent ainsi

$$\begin{aligned} \widehat{\alpha}_x^{(\nu+1)} &= \widehat{\alpha}_x^{(\nu)} + \frac{\sum_{t=t_{\min}}^{t_{\max}} (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu)} - \widehat{\beta}_x^{(\nu)} t)}{t_{\max} - t_{\min} + 1} \\ \widehat{\beta}_x^{(\nu+1)} &= \widehat{\beta}_x^{(\nu)} + \frac{\sum_{t=t_{\min}}^{t_{\max}} t (\ln \widehat{\mu}_x(t) - \widehat{\alpha}_x^{(\nu+1)} - \widehat{\beta}_x^{(\nu)} t)}{\sum_{t=t_{\min}}^{t_{\max}} t^2}. \end{aligned} \quad (3.27)$$

Il est également possible de modifier les estimateurs obtenus pour les rendre compatibles avec les contraintes (3.2)-(3.3). Cette opération s'effectue par les

transformations habituelles:

$$\begin{aligned}\widehat{\alpha}_x &\leftarrow \widehat{\alpha}_x + \widehat{\beta}_x \bar{t} \\ \widehat{\kappa}_t &\leftarrow (t - \bar{t}) \widehat{\beta}_\bullet \\ \widehat{\beta}_x &\leftarrow \frac{\widehat{\beta}_x}{\widehat{\beta}_\bullet}\end{aligned}$$

où  $\bar{t} = \frac{t_{\min} + t_{\max}}{2}$  est la moyenne des années retenues dans le modèle et  $\widehat{\beta}_\bullet$  la somme des  $\widehat{\beta}_x$  initiaux. Les estimateurs de la tendance temporelle  $\widehat{\kappa}_t$  ne sont donc plus égaux au temps calendaire  $t$  comme le spécifiait (3.26), mais ils restent cependant linéaires. Par ailleurs, sous les contraintes (3.2)-(3.3), l'algorithme itératif associé à la transformation précédente doit rétablir les estimateurs

$$\widehat{\alpha}_x = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t=t_{\min}}^{t_{\max}} \ln \widehat{\mu}_x(t).$$

Nous disposons donc d'un moyen supplémentaire pour vérifier la convergence des itérations (3.27).

L'application du modèle (3.25) associé aux contraintes (3.2)-(3.3) engendre les courbes présentes aux Figures 3.17 et 3.18. Nous avons ici utilisé toute la période d'observation 1950-2000. Dans un cas comme celui-ci, il serait sans doute préférable d'ajuster le modèle aux années supérieures à 1970, période au cours de laquelle la décroissance des  $\widehat{\kappa}_t$  du modèle (3.1) semblait linéaire (comme le montre la Figure 3.10). Ainsi, on pourrait par exemple choisir la période d'observation en déterminant une nouvelle année  $t_{\min}^*$  de la manière suivante:

$$t_{\min}^* = \arg \max_u R_{u, t_{\max}}^2 \quad (3.28)$$

où  $u < t_{\max} - 1$  et  $R_{u, t_{\max}}^2$  est la valeur de la statistique  $R^2$  d'une régression linéaire sur les paramètres  $\{\widehat{\kappa}_u, \dots, \widehat{\kappa}_{t_{\max}}\}$  du modèle classique de Lee et Carter (3.1).

La Figure 3.19 montre les valeurs de  $R_{u, t_{\max}}^2$  pour nos données françaises. On y remarque que le coefficient devient excellent entre 1970 et 1990. Pour garder un nombre suffisant d'observations, il semble donc raisonnable de choisir  $t_{\min}^* = 1970$ .

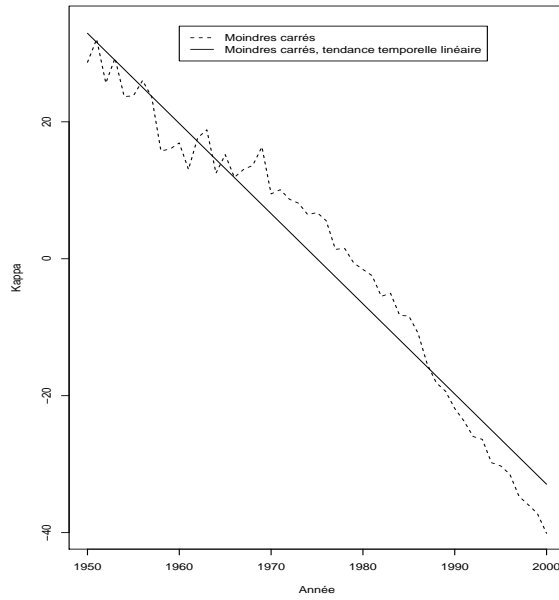


FIG. 3.17 – Estimation des  $\kappa_t$  pour le modèle à tendance temporelle linéaire.

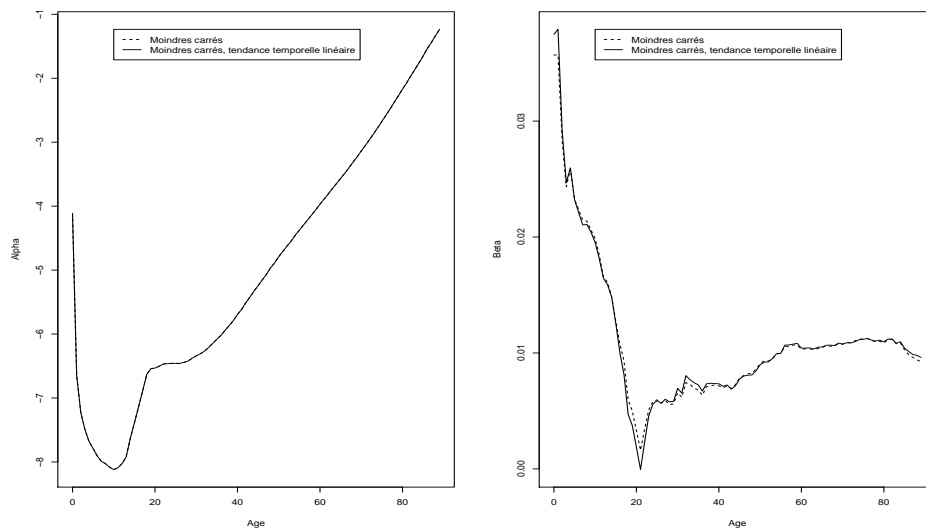


FIG. 3.18 – Estimation des  $\alpha_x$  et  $\beta_x$  pour le modèle à tendance temporelle linéaire.

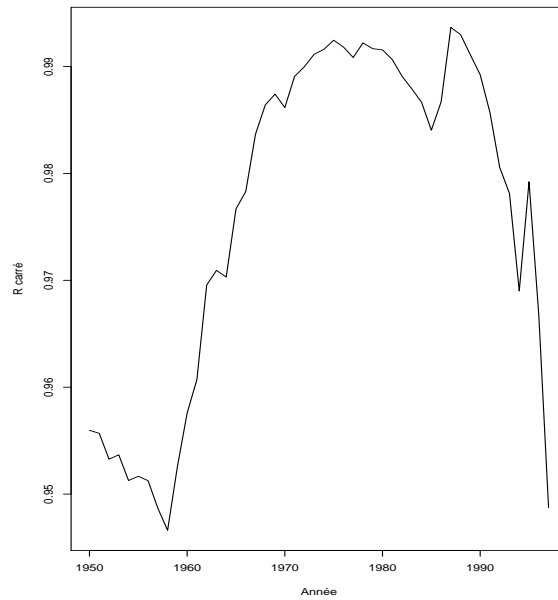


FIG. 3.19 –  $R_{u,t_{\max}}^2$  pour les différentes valeurs de  $u$ .

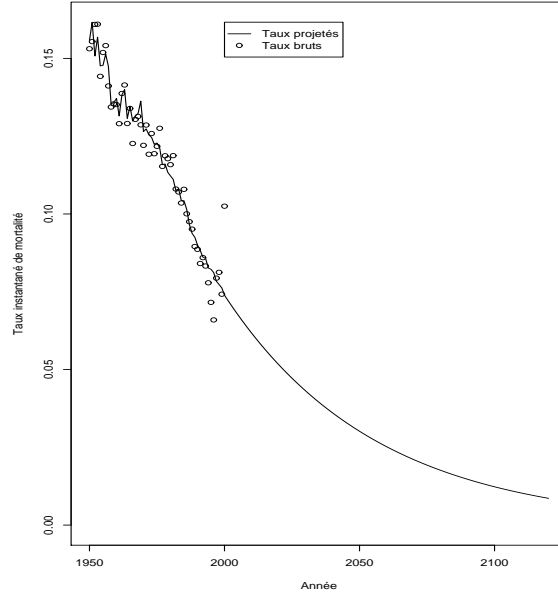


FIG. 3.20 – *Décroissance du taux instantané de mortalité à 80 ans sur base du modèle (3.1)-(3.21).*

### 3.2.11 Modèle avec table limite

Si le modèle log-bilinéaire possède de nombreux atouts, il souffre également de certaines carences. L'une d'entre elles réside dans le comportement asymptotique des taux instantanés de mortalité projetés dans le futur. Dans toutes les applications de ce modèle recensées à ce jour, nous trouvons une tendance temporelle décroissante, associée à des  $\hat{\beta}_x$  positifs. Ce contexte naturel est celui d'une mortalité qui décline avec le temps. Seuls quelques âges très avancés échappent à cette règle, avec des  $\hat{\beta}_x$  presque nuls, voire carrément négatifs. Donc, à quelques exceptions près, extrapoler les  $\hat{\kappa}_t$  à un horizon infini revient à supposer des taux instantanés de mortalité asymptotiquement nuls. En effet, si  $\hat{\beta}_x$  est positif et  $\hat{\alpha}_x$  est fini, il vient

$$\lim_{\hat{\kappa}_t \rightarrow -\infty} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t) = 0.$$

La Figure 3.20, qui montre la décroissance avec le temps du taux instantané de mortalité à 80 ans sur base du modèle (3.1)-(3.21), confirme cette analyse. Un tel comportement semble cependant improbable.

Supposons que nous disposions d'une estimation du comportement asymptotique de la mortalité, et notons  $\mu_x^{lim}$  et  $q_x^{lim}$  les taux instantanés et quotients y correspondant. Il est très aisé de tenir compte de ce phénomène dans le modèle (3.1), qui devient alors

$$\ln(\hat{\mu}_x(t) - \mu_x^{lim}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt} \tag{3.29}$$

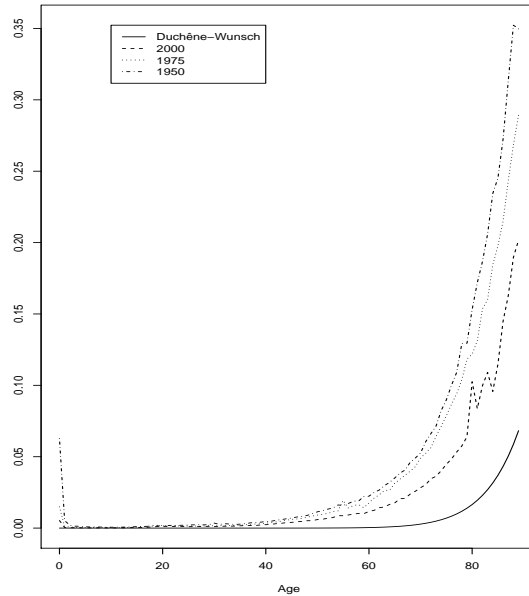


FIG. 3.21 – *Taux instantanés de mortalité de la table limite de Duchêne et Wunsch comparés aux estimations brutes en 1950, 1975 et 2000.*

avec comme toujours des erreurs  $\epsilon_{xt}$  supposées centrées, indépendantes et de même variance  $\sigma^2$ .

L'estimation des nouveaux paramètres est alors effectuée comme précédemment, soit pas la décomposition en valeurs singulières, soit pas l'algorithme itératif. Seuls les estimations brutes de taux de mortalité  $\hat{\mu}_x(t)$  doivent être remplacées dans les formules par les différences  $\hat{\mu}_x(t) - \mu_x^{lim}$ .

DUCHÊNE & WUNSCH (1988) ont établi une table unisexe, reprenant des quotients de mortalité minimaux qui représentent, selon eux, les limites naturelles humaines. Les taux instantanés de mortalité y sont nuls jusqu'à près de 40 ans, comme le montre la Figure 3.21 qui les compare aux estimations brutes françaises pour quelques années. L'incorporation de cette table dans le modèle (3.29) engendre pour nos données les graphiques repris aux Figures 3.22 et 3.23. Comme prévu, l'influence est nette aux âges élevés mais très faible avant. La courbe des  $\hat{\kappa}_t$  est pour sa part légèrement modifiée.

L'influence du nouveau modèle sur la projection des taux instantanés de mortalité peut être observée à la Figure 3.24. La convergence vers un niveau de mortalité nul y est remplacée par une convergence vers le taux instantané fourni par la table limite.



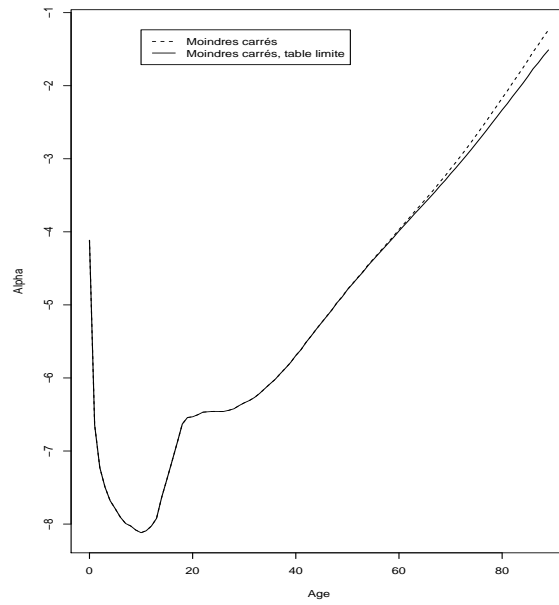


FIG. 3.22 – Estimation des  $\alpha_x$  pour le modèle avec table limite.

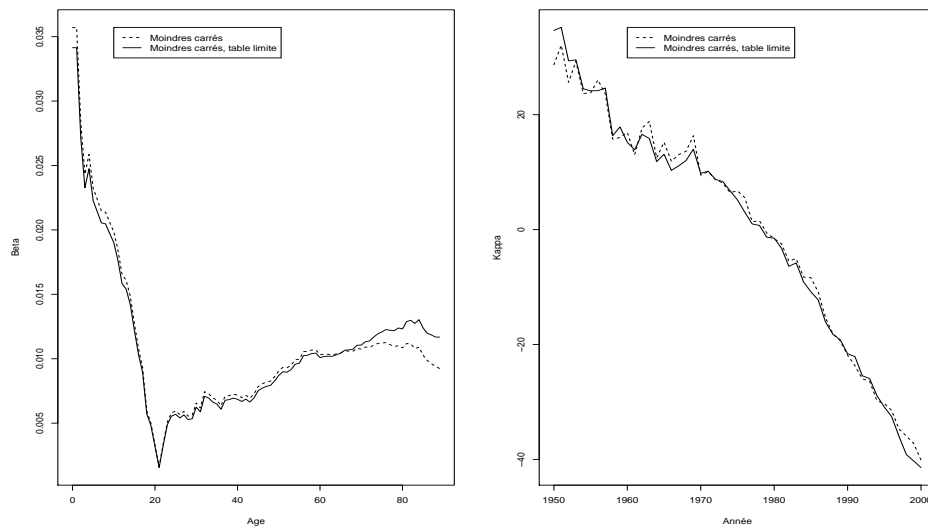


FIG. 3.23 – Estimation des  $\beta_x$  et  $\kappa_t$  pour le modèle avec table limite.

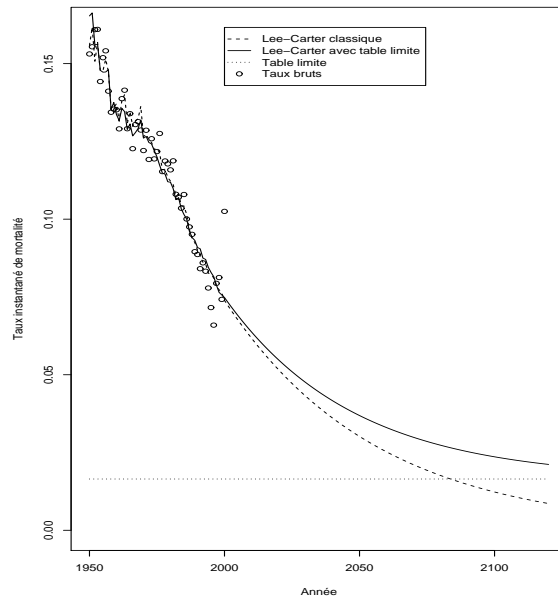


FIG. 3.24 – Décroissance du taux instantané de mortalité à 80 ans sur base du modèle (3.29)-(3.21).

### 3.3 Modèle de Poisson

Comme l'a très justement fait remarquer ALHO (2000), le critère des moindres carrés (qui mène à la décomposition aux valeurs singulières), voire sa version pondérée suggérée par WILMOTH (1993), est raisonnable mais certainement pas optimal. Le principal défaut du modèle de Lee et Carter réside dans l'hypothèse d'homoscédasticité des erreurs requise pour l'estimation par moindres carrés ordinaires. En effet, le logarithme du taux instantané de mortalité observé est généralement bien plus variable aux grands âges que chez les jeunes, étant donné le faible nombre d'observations. C'est pourquoi BROUHNS, DENNIT & VERMUNT (2002a) ont suggéré de modéliser le nombre  $D_{xt}$  de décès constatés à l'âge  $x$  durant l'année  $t$  à l'aide de la loi de Poisson, i.e. de spécifier

$$D_{xt} \sim \text{Poisson}(E_{xt}\mu_x(t)). \quad (3.30)$$

où  $E_{xt}$  est le nombre de personnes/années d'âge  $x$  durant l'année  $t$  (i.e. l'exposition au risque). La modélisation poissonnienne (3.30) peut se justifier sur base démographique; voyez par exemple BRILLINGER (1986).

Montrons dans quelle mesure l'hypothèse (2.11) corrobore la modélisation de Poisson (3.30). Intéressons-nous à un couple  $(x,t)$  particulier. Nous observons  $D_{xt}$  décès recensés parmi  $L_{xt}$  individus d'âge  $x$  au début de la période d'observation annuelle. On suppose que les durées de vie de ces individus peuvent être considérées comme des variables aléatoires indépendantes et identiquement distribuées.

À chacun des  $L_{xt}$  individus, on associe une variable  $\delta_i$  indiquant s'il est décédé ou non, i.e.

$$\delta_i = \begin{cases} 1 & \text{si l'individu } i \text{ est décédé à l'âge } x \text{ considéré} \\ 0 & \text{sinon,} \end{cases}$$

$i = 1, 2, \dots, L_{xt}$ . Définissons également  $\tau_i$  comme le temps vécu par l'individu  $i$ . Ainsi  $\tau_i = 1$  si l'individu  $i$  atteint l'âge  $x+1$  et  $\tau_i$  vaut la fraction d'année vécue par l'individu  $i$  s'il décède à l'âge  $x$ . Nous supposons disposer pour chacun des  $L_{xt}$  individus des observations  $(\delta_i, \tau_i)$ .

Grâce à l'hypothèse (2.11), la contribution du  $i^{\text{ème}}$  individu à la vraisemblance s'écrit comme suit:

- si l'individu  $i$  survit, la vraisemblance le concernant est

$$p_x(t) = \exp(-\mu_x(t));$$

- si l'individu  $i$  décède, la vraisemblance le concernant vaut

$$\tau_i p_x(t) \mu_{x+\tau_i}(t + \tau_i) = \exp(-\tau_i \mu_x(t)) \mu_x(t).$$

Dès lors, la contribution du  $i^{\text{ème}}$  individu se réécrit simplement sous la forme

$$\exp(-\tau_i \mu_x(t)) \{\mu_x(t)\}^{\delta_i}.$$

Grâce à l'hypothèse d'indépendance, la vraisemblance associée aux  $L_{xt}$  individus se met sous la forme

$$\mathcal{V}(\mu_x(t)) = \prod_{i=1}^{L_{xt}} \exp(-\tau_i \mu_x(t)) \{\mu_x(t)\}^{\delta_i} = \exp(-\mu_x(t) \tau_{\bullet}) \{\mu_x(t)\}^{\delta_{\bullet}} \quad (3.31)$$

où

$$\tau_{\bullet} = \sum_{i=1}^{L_{xt}} \tau_i \text{ et } \delta_{\bullet} = \sum_{i=1}^{L_{xt}} \delta_i,$$

$\delta_{\bullet}$  est ainsi le nombre de décès observé à l'âge  $x$  durant l'année  $t$ ,  $D_{xt}$ , et  $\tau_{\bullet}$  est le temps total vécu par les  $L_{xt}$  individus au cours de l'année  $t$ .

La vraisemblance (3.31) ne diffère qu'à un facteur constant près de la vraisemblance obtenue sous l'hypothèse (3.30). Dès lors, si on a recours au principe du maximum de vraisemblance pour estimer les paramètres, il est indifférent de travailler sur base de la vraisemblance (3.31) ou de celle déduite de l'hypothèse (3.30). La loi de Poisson, dite des événements rares, devrait d'ailleurs particulièrement bien convenir pour modéliser le nombre des décès survenus aux différents âges. Ainsi, supposer que  $D_{xt}$  est de loi de Poisson de moyenne  $\mu_x(t) \tau_{\bullet}$  semble raisonnable.

Notez que la spécification (3.30) n'est qu'une approximation du modèle naturel

$$D_{xt} \sim \text{Binomiale}(L_{xt}, q_x(t))$$

puisque (3.30) permet avec une probabilité positive d'observer plus de  $L_{xt}$  décès, ce qui est clairement impossible. Néanmoins, cette probabilité est insignifiante en pratique et ce modèle fournit souvent une excellente approximation.

D'autre part, BRILLINGER (1986) a montré que sous des hypothèses raisonnables formulées à propos des processus gouvernant les naissances et les décès, la loi de Poisson apparaissait naturellement pour modéliser le nombre de décès aux différents âges.

### 3.3.1 Modèle

RENSHAW & HABERMAN (1996) suggèrent d'utiliser la régression poissonnienne. Plus précisément, le nombre de décès à l'âge  $x$  au cours de l'année  $t$  est modélisé au moyen de la loi de Poisson de moyenne  $\mu_x(t)$  multiplié par l'exposition au risque correspondante. Ces auteurs proposent de retenir pour  $\mu_x(t)$  la forme

$$\ln \mu_x(t) = \sum_{j=1}^p \psi_j(x, t) \beta_j$$

où les fonctions  $\psi_j(\cdot, \cdot)$  de l'âge et du temps calendaire jouent le rôle de variables explicatives, et où les  $\beta_j$  sont autant de paramètres à estimer sur base des données. Cette méthode a été appliquée par SITHOLE, HABERMAN & VERRALL (2000). On lui reprochera cependant son côté ad hoc.

Comme dans BROUHNS, DENUIT & VERMUNT (2002a), nous proposons de compléter le modèle (3.30) par la spécification

$$\mu_x(t) = \exp(\alpha_x + \beta_x \kappa_t) \quad (3.32)$$

où les paramètres respectent toujours les contraintes (3.2) et (3.3). La forme log-bilinéaire des taux instantanés de mortalité proposée par LEE & CARTER (1992) n'est donc pas remise en question, mais la structure de l'erreur est modifiée: plutôt que de superposer au taux instantané de mortalité observé un bruit additif sur l'échelle logarithmique comme le prévoit (3.1), nous considérons ici la loi de Poisson pour le nombre observé de décès.

Dans cette reformulation du problème, l'interprétation des  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  reste essentiellement la même que dans le modèle classique de Lee et Carter. Le modèle (3.30)-(3.32) décompose donc l'évolution des taux instantanés de mortalité en différentes composantes, dont la signification est la suivante:

$\alpha_x$ : décrit le comportement moyen des  $\mu_x(t)$  au cours du temps.

$\beta_x$ : décrit (toujours sur l'échelle logarithmique) l'écart des  $\mu_x(t)$  par rapport au comportement moyen  $\alpha_x$ . Les âges  $x$  pour lesquels  $\beta_x$  est grand seront donc très sensibles à l'évolution temporelle. Si les  $\kappa_t$  sont (approximativement) linéaires, les taux de mortalité  $\mu_x(t)$  présentent une décroissance exponentielle à un taux constant.

$\kappa_t$ : décrit l'évolution de la mortalité au cours du temps.

La spécification (3.30) présente plusieurs avantages sur son homologue (3.1) de Lee et Carter, au nombre desquels

1. le modèle (3.30) reconnaît le caractère entier du nombre de décès  $D_{xt}$ , ce que ne fait pas (3.1);
2. la loi de Poisson utilisée dans (3.30) permet de lever l'hypothèse d'homoscédasticité de (3.1) et reconnaît la variabilité plus grande des taux bruts  $\hat{\mu}_x(t)$  aux âges élevés, due aux effectifs réduits;
3. la spécification (3.30) permet le recours au principe du maximum de vraisemblance pour estimer les paramètres. Comme l'ajustement de (3.30) ne repose plus sur une décomposition aux valeurs singulières de la matrice des  $\ln \hat{\mu}_x(t)$ , (3.30) ne nécessite plus des statistiques de mortalité sous forme rectangulaire;
4. la modélisation des observations  $D_{xt}$  plutôt que des taux de mortalité observés permet d'éviter la phase de réestimation des  $\kappa_t$  suggérée par LEE & CARTER (1992) afin de reproduire le nombre annuel total observé de décès. En effet, les équations de vraisemblance garantissent que le modèle de Poisson log-bilinéaire reproduit le nombre total de décès observés à chaque âge au cours de la période d'observation, i.e.

$$\frac{\partial}{\partial \alpha_x} \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) = 0 \Leftrightarrow \sum_{t=t_{\min}}^{t_{\max}} D_{xt} = \sum_{t=t_{\min}}^{t_{\max}} E_{xt} \exp(\alpha_x + \beta_x \kappa_t),$$

où  $\mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa})$  est la fonction de vraisemblance associée au modèle de Poisson. Enfin, le modèle (3.30) est particulièrement bien adapté aux applications actuarielles (comme par exemple les projections des flux financiers relatifs à un portefeuille de rentes).

### 3.3.2 Estimation des paramètres par maximum de vraisemblance

Un grand avantage du modèle (3.30)-(3.32) est qu'il est aisé d'en estimer les paramètres par la méthode de maximum de vraisemblance. En effet, en notant  $\lambda_{xt}$  le nombre attendu de décès à l'âge  $x$  durant l'année  $t$ , i.e.

$$\lambda_{xt} = \mathbb{E}[D_{xt}] = E_{xt} \exp(\alpha_x + \beta_x \kappa_t),$$

la fonction de vraisemblance s'écrit

$$\mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) = \prod_{t=t_{\min}}^{t_{\max}} \prod_{x=x_{\min}}^{x_{\max}} \frac{\lambda_{xt}^{D_{xt}} \exp(-\lambda_{xt})}{(D_{xt})!}.$$

La fonction de log-vraisemblance vaut donc, à un facteur additif près,

$$\ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) = \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt}(\alpha_x + \beta_x \kappa_t) - E_{xt} \exp(\alpha_x + \beta_x \kappa_t) \right).$$

Les estimations du maximum de vraisemblance des paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  sont les valeurs de ceux-ci qui maximisent  $\mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa})$ , i.e. qui rendent les statistiques de mortalité dont on dispose les plus vraisemblables dans le modèle (3.30)-(3.32).

La présence du terme bilinéaire  $\beta_x \kappa_t$  dans la spécification (3.32) du taux de mortalité rend l'estimation des paramètres irréalisable par les logiciels commerciaux traitant de la régression de Poisson (qui ne couvrent que des formes linéaires du score). Néanmoins, nous disposons d'un algorithme itératif proposé par GOODMAN (1979), et qui utilise la méthode de Newton-Raphson. Le principe en est simple : à chaque itération, chaque ensemble de paramètres est mis à jour en fixant les autres à leur valeur courante. Plus précisément, si on note  $\boldsymbol{\theta}$  les paramètres du modèle, à l'étape  $\nu + 1$ ,  $\theta_j$  est réestimé par

$$\widehat{\theta}_j^{(\nu+1)} = \widehat{\theta}_j^{(\nu)} - \frac{\partial \mathcal{V}^{(\nu)} / \partial \theta_j}{\partial^2 \mathcal{V}^{(\nu)} / \partial \theta_j^2}$$

où  $\mathcal{V}^{(\nu)} = \mathcal{V}(\widehat{\boldsymbol{\theta}}^{(\nu)})$ .

Dans le cas qui nous occupe, il existe trois ensembles de paramètres, à savoir les  $\alpha_x$ , les  $\beta_x$  et les  $\kappa_t$ . Le schéma itératif est le suivant: partant de valeurs

initiales  $\widehat{\alpha}_x^{(0)}$ ,  $\widehat{\beta}_x^{(0)}$  et  $\widehat{\kappa}_t^{(0)}$ , les estimations sont revues à l'aide des formules

$$\begin{aligned}
 \widehat{\alpha}_x^{(\nu+1)} &= \widehat{\alpha}_x^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\widehat{\alpha}_x^{(\nu)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \right)}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\widehat{\alpha}_x^{(\nu)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \right)} \\
 \widehat{\kappa}_t^{(\nu+1)} &= \widehat{\kappa}_t^{(\nu)} - \frac{\sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} - E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \right) \widehat{\beta}_x^{(\nu)}}{-\sum_{x=x_{\min}}^{x_{\max}} \left( E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \right) \left( \widehat{\beta}_x^{(\nu)} \right)^2} \quad (3.33) \\
 \widehat{\beta}_x^{(\nu+1)} &= \widehat{\beta}_x^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu+1)}) \right) \widehat{\kappa}_t^{(\nu+1)}}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu+1)}) \right) \left( \widehat{\kappa}_t^{(\nu+1)} \right)^2}.
 \end{aligned}$$

Cet algorithme nécessite une initialisation. Il faut évidemment éviter d'attribuer une valeur nulle à  $\widehat{\beta}_x^{(0)}$  sous peine de constater une division par 0 dans le calcul de  $\widehat{\kappa}_t^{(1)}$ . Cette exception mise à part, le choix des paramètres initiaux est libre. Il ne reste alors plus qu'à fixer un critère d'arrêt. Nous convenons de stopper les itérations lorsque l'augmentation relative de la fonction de log-vraisemblance n'excède plus une borne inférieure fixée.

Ces itérations peuvent être facilement réécrites sous forme matricielle:

$$\begin{aligned}
 (\mathbf{C}_\alpha^{(\nu)}) \widehat{\boldsymbol{\alpha}}^{(\nu+1)} &= (\mathbf{C}_\alpha^{(\nu)}) \widehat{\boldsymbol{\alpha}}^{(\nu)} + \mathbf{r}_\alpha^{(\nu)} \\
 (\mathbf{C}_\kappa^{(\nu)}) \widehat{\boldsymbol{\kappa}}^{(\nu+1)} &= (\mathbf{C}_\kappa^{(\nu)}) \widehat{\boldsymbol{\kappa}}^{(\nu)} + \mathbf{r}_\kappa^{(\nu)} \\
 (\mathbf{C}_\beta^{(\nu)}) \widehat{\boldsymbol{\beta}}^{(\nu+1)} &= (\mathbf{C}_\beta^{(\nu)}) \widehat{\boldsymbol{\beta}}^{(\nu)} + \mathbf{r}_\beta^{(\nu)}
 \end{aligned} \quad (3.34)$$

avec  $\mathbf{C}_\alpha^{(\nu)}$ ,  $\mathbf{C}_\kappa^{(\nu)}$  et  $\mathbf{C}_\beta^{(\nu)}$  des matrices diagonales respectivement de dimensions  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ ,  $(t_{\max} - t_{\min} + 1) \times (t_{\max} - t_{\min} + 1)$  et  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ , et dont les éléments valent

$$\begin{aligned}
 (\mathbf{C}_\alpha^{(\nu)})_{xx} &= \sum_{t=t_{\min}}^{t_{\max}} E_{xt} \exp(\widehat{\alpha}_x^{(\nu)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \\
 (\mathbf{C}_\kappa^{(\nu)})_{tt} &= \sum_{x=x_{\min}}^{x_{\max}} \left( \widehat{\beta}_x^{(\nu)} \right)^2 E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu)}) \\
 (\mathbf{C}_\beta^{(\nu)})_{xx} &= \sum_{t=t_{\min}}^{t_{\max}} \left( \widehat{\kappa}_t^{(\nu+1)} \right)^2 E_{xt} \exp(\widehat{\alpha}_x^{(\nu+1)} + \widehat{\beta}_x^{(\nu)} \widehat{\kappa}_t^{(\nu+1)})
 \end{aligned}$$

et  $\mathbf{r}_\alpha^{(\nu)}$ ,  $\mathbf{r}_\kappa^{(\nu)}$  et  $\mathbf{r}_\beta^{(\nu)}$  des vecteurs-colonnes respectivement de dimensions  $(x_{\max} -$

$x_{\min} + 1$ ),  $(t_{\max} - t_{\min} + 1)$  et  $(x_{\max} - x_{\min} + 1)$ , et dont les éléments valent

$$\begin{aligned} (\mathbf{r}_{\hat{\alpha}}^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} (D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu)})) \\ (\mathbf{r}_{\hat{\kappa}}^{(\nu)})_t &= \sum_{x=x_{\min}}^{x_{\max}} \hat{\beta}_x^{(\nu)} (D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu)})) \\ (\mathbf{r}_{\hat{\beta}}^{(\nu)})_x &= \sum_{t=t_{\min}}^{t_{\max}} \hat{\kappa}_t^{(\nu+1)} (D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu+1)})). \end{aligned}$$

Remarquons qu'aucune des contraintes (3.2)-(3.3) n'est respectée par les estimations  $\hat{\alpha}_x$ ,  $\hat{\beta}_x$  et  $\hat{\kappa}_t$  résultant de (3.33) ou (3.34). C'est pourquoi il est nécessaire de modifier les estimateurs ainsi obtenus. Il y a donc lieu d'effectuer la transformation

$$\begin{aligned} \hat{\alpha}_x &\leftarrow \hat{\alpha}_x + \hat{\beta}_x \bar{\kappa} \\ \hat{\kappa}_t &\leftarrow (\hat{\kappa}_t - \bar{\kappa}) \hat{\beta}_{\bullet} \\ \hat{\beta}_x &\leftarrow \frac{\hat{\beta}_x}{\hat{\beta}_{\bullet}} \end{aligned}$$

où  $\bar{\kappa}$  est la moyenne des  $\hat{\kappa}_t$  initiaux et  $\hat{\beta}_{\bullet}$  la somme des  $\hat{\beta}_x$  initiaux.

Les nouveaux estimateurs satisfont maintenant les contraintes (3.2)-(3.3) tout en maintenant inchangée la valeur de  $(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$ .

L'application du modèle (3.30)-(3.32) aux données françaises donne les graphiques repris aux Figures 3.25 et 3.26. On y retrouvent les mêmes caractéristiques qu'aux Figures 3.1 et 3.2 relatives au modèle classique de Lee et Carter.



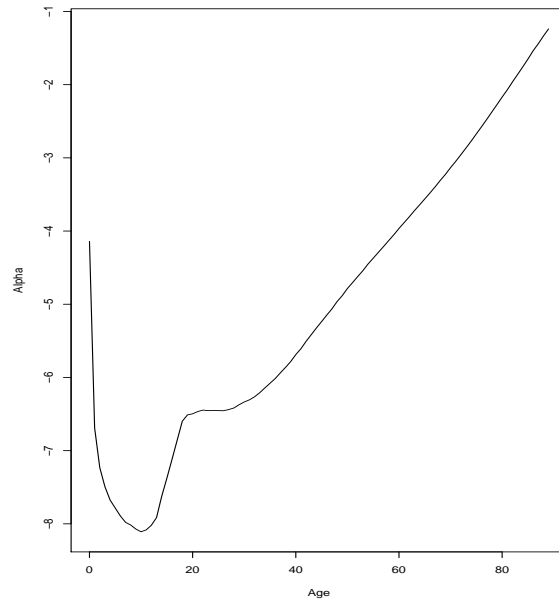


FIG. 3.25 – Estimation des  $\alpha_x$  du modèle (3.30)-(3.32).

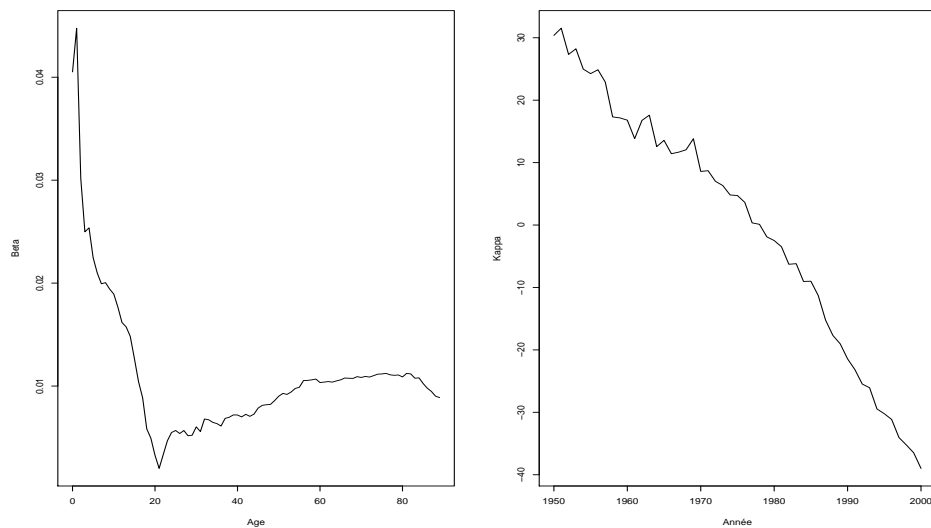


FIG. 3.26 – Estimation des  $\beta_x$  et  $\kappa_t$  du modèle (3.30)-(3.32) avec contraintes (3.2)-(3.3) par l'algorithme de Goodman.

### 3.3.3 Estimation des paramètres par maximum de vraisemblance pénalisée

Nous sommes déjà revenus sur l'importance d'obtenir des courbes de paramètres lisses. Dans le cadre du modèle de Lee et Carter, une des solutions à ce problème était d'utiliser le critère des moindres carrés pénalisés. Il est possible d'adopter la même démarche pour le modèle (3.30)-(3.32), en pénalisant cette fois la fonction de log-vraisemblance. EILERS & MARX (1996) définissent donc la fonction de log-vraisemblance pénalisée:

$$\begin{aligned} & \ln \mathcal{V}_P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) \\ = & \ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) - \frac{1}{2} \boldsymbol{\alpha}' \mathbf{P}_\alpha \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\beta}' \mathbf{P}_\beta \boldsymbol{\beta} - \frac{1}{2} \boldsymbol{\kappa}' \mathbf{P}_\kappa \boldsymbol{\kappa} \\ = & \ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) - \frac{1}{2} \pi_\alpha \boldsymbol{\alpha}' \mathbf{D}'_\alpha \mathbf{D}_\alpha \boldsymbol{\alpha} - \frac{1}{2} \pi_\beta \boldsymbol{\beta}' \mathbf{D}'_\beta \mathbf{D}_\beta \boldsymbol{\beta} - \frac{1}{2} \pi_\kappa \boldsymbol{\kappa}' \mathbf{D}'_\kappa \mathbf{D}_\kappa \boldsymbol{\kappa} \end{aligned} \quad (3.35)$$

où  $\mathbf{P}_\alpha = \pi_\alpha \mathbf{D}'_\alpha \mathbf{D}_\alpha$ ,  $\mathbf{P}_\beta = \pi_\beta \mathbf{D}'_\beta \mathbf{D}_\beta$  et  $\mathbf{P}_\kappa = \pi_\kappa \mathbf{D}'_\kappa \mathbf{D}_\kappa$  sont des matrices de pénalité et  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  des paramètres de lissage.

Les itérations (3.34) deviennent alors, en utilisant les mêmes notations que précédemment,

$$\begin{aligned} (\mathbf{C}_\alpha^{(\nu)} + \text{diag} \mathbf{P}_\alpha) \hat{\boldsymbol{\alpha}}^{(\nu+1)} &= (\mathbf{C}_\alpha^{(\nu)} + \text{diag} \mathbf{P}_\alpha - \mathbf{P}_\alpha) \hat{\boldsymbol{\alpha}}^{(\nu)} + \mathbf{r}_\alpha^{(\nu)} \\ (\mathbf{C}_\kappa^{(\nu)} + \text{diag} \mathbf{P}_\kappa) \hat{\boldsymbol{\kappa}}^{(\nu+1)} &= (\mathbf{C}_\kappa^{(\nu)} + \text{diag} \mathbf{P}_\kappa - \mathbf{P}_\kappa) \hat{\boldsymbol{\kappa}}^{(\nu)} + \mathbf{r}_\kappa^{(\nu)} \\ (\mathbf{C}_\beta^{(\nu)} + \text{diag} \mathbf{P}_\beta) \hat{\boldsymbol{\beta}}^{(\nu+1)} &= (\mathbf{C}_\beta^{(\nu)} + \text{diag} \mathbf{P}_\beta - \mathbf{P}_\beta) \hat{\boldsymbol{\beta}}^{(\nu)} + \mathbf{r}_\beta^{(\nu)} \end{aligned} \quad (3.36)$$

où  $\text{diag} \mathbf{P}_\alpha$ ,  $\text{diag} \mathbf{P}_\kappa$  et  $\text{diag} \mathbf{P}_\beta$  sont les matrices composées exclusivement de la diagonale principale des matrices  $\mathbf{P}_\alpha$ ,  $\mathbf{P}_\kappa$  et  $\mathbf{P}_\beta$ .

Le procédé de sélection des paramètres de lissage  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  est à nouveau basé sur la cross-validation. Définissons cette fois l'erreur de prédiction  $\epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}$  de la manière suivante:

$$\begin{aligned} \epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa} &= \sqrt{2} \cdot \text{sign}(D_{xt} - \hat{D}_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)) \\ &\quad \sqrt{D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)} - (D_{xt} - \hat{D}_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t))} \end{aligned} \quad (3.37)$$

avec

$$\hat{D}_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) = E_{xt} \exp(\hat{\alpha}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) + \hat{\beta}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t) \hat{\kappa}_{t, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t))$$

et  $\hat{\alpha}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$ ,  $\hat{\beta}_{x, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$  et  $\hat{\kappa}_{t, \pi_\alpha, \pi_\beta, \pi_\kappa}^{-}(x, t)$  les estimateurs par maximum de vraisemblance pénalisée par les paramètres  $\pi_\alpha$ ,  $\pi_\beta$  et  $\pi_\kappa$  du modèle duquel on a supprimé le couple de données  $(x, t)$ . La définition de ces erreurs de prédiction est en fait basée sur la formule des résidus de déviance, lesquels sont définis à la section 3.3.5. La notion d'erreur de prédiction doit cependant être préférée puisqu'il ne s'agit pas à proprement parler de résidus, les données relatives au couple  $(x, t)$  considéré ayant été volontairement omises de l'ajustement. Les

paramètres de lissage optimaux  $\hat{\pi}_\alpha$ ,  $\hat{\pi}_\beta$  et  $\hat{\pi}_\kappa$  sont alors solutions du problème d'optimisation

$$(\hat{\pi}_\alpha, \hat{\pi}_\beta, \hat{\pi}_\kappa) = \arg \min_{\pi_\alpha, \pi_\beta, \pi_\kappa} \epsilon_{\bullet\bullet, \pi_\alpha, \pi_\beta, \pi_\kappa}^2 \quad (3.38)$$

avec

$$\epsilon_{\bullet\bullet, \pi_\alpha, \pi_\beta, \pi_\kappa}^2 = \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \epsilon_{xt, \pi_\alpha, \pi_\beta, \pi_\kappa}^2.$$

La méthode consiste donc à minimiser une pseudo déviance basée sur des erreurs de prédiction.

À nouveau, observons l'influence de  $\pi_\beta$  sur l'estimation des paramètres, tout en maintenant  $\pi_\alpha = \pi_\kappa = 0$ . Les courbes obtenues se trouvent aux Figures 3.27 et 3.28. On remarque cette fois que le lissage des  $\hat{\beta}_x$  n'a que peu d'influence sur les valeurs des  $\hat{\alpha}_x$  et  $\hat{\kappa}_t$ . Calculons donc  $\epsilon_{\bullet\bullet, 0, \pi_\beta, 0}^2$  pour un panel de valeurs de  $\pi_\beta$ :

$\pi_\beta$	$\epsilon_{\bullet\bullet, 0, \pi_\beta, 0}^2$
0	97356.10
1	97356.10
$10^1$	97356.09
$10^2$	97356.05
$10^3$	97355.65
$10^4$	97351.71
$10^5$	97321.19
$10^6$	97102.99
<b><math>10^7</math></b>	<b>97092.17</b>
$2 \cdot 10^7$	97208.05
$4 \cdot 10^7$	97450.30
$8 \cdot 10^7$	98092.59
$10^8$	98356.71

Le paramètre optimal est donc approximé par  $\hat{\pi}_\beta \approx 10^7$ .

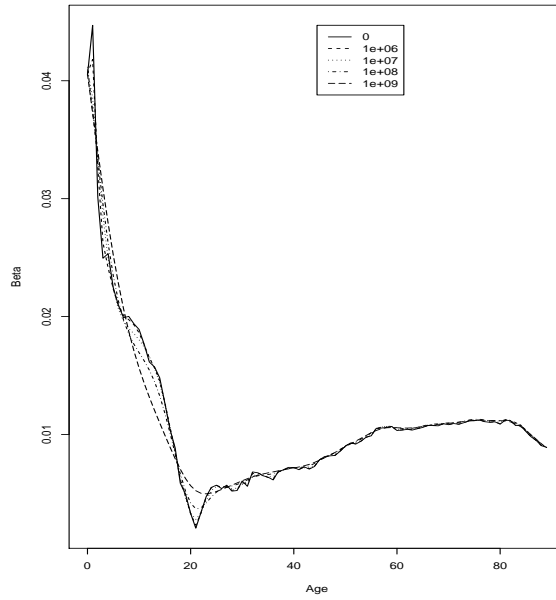


FIG. 3.27 – Estimation des  $\beta_x$  par la méthode des moindres carrés pénalisés pour quelques valeurs de  $\pi_\beta$ .

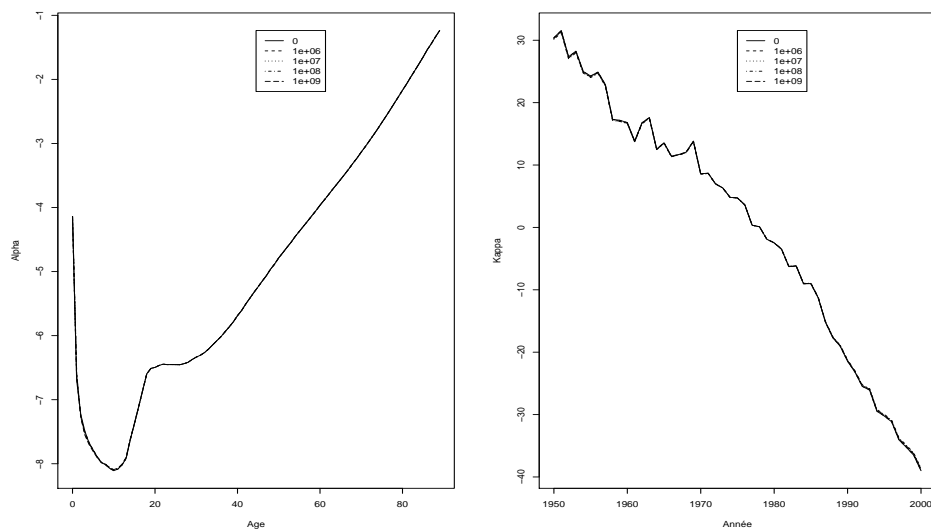


FIG. 3.28 – Estimation des  $\alpha_x$  et  $\kappa_t$  par la méthode des moindres carrés pénalisés pour quelques valeurs de  $\pi_\beta$ .

### 3.3.4 Mesure de la qualité de l'ajustement

#### Déviance

Dans les modèles de régression, il est de bon usage d'apprécier la qualité d'un ajustement par la mesure de la déviance. Celle-ci est définie par le double de la différence entre la log-vraisemblance du modèle saturé (i.e. celui comptant autant de paramètres que de données) et celle du modèle choisi. Dans la situation de notre régression poissonnienne, chaque donnée est représentée par un couple  $(x, t)$ . Formellement, la déviance vaut donc

$$\begin{aligned} \mathcal{D}(D, \hat{D}) &= 2(\ln \mathcal{V}^{sat} - \ln \mathcal{V}(\hat{\alpha}, \hat{\beta}, \hat{\kappa})) \\ &= 2 \left( \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} \ln D_{xt} - D_{xt} - \sum_{i=1}^{D_{xt}} \ln i \right) \right. \\ &\quad \left. - \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} \ln \hat{D}_{xt} - \hat{D}_{xt} - \sum_{i=1}^{D_{xt}} \ln i \right) \right) \\ &= 2 \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt}} - (D_{xt} - \hat{D}_{xt}) \right) \end{aligned}$$

avec  $\hat{D}_{xt} = E_{xt} \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$ . Cette expression peut se simplifier en

$$\mathcal{D}(D, \hat{D}) = 2 \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt}} \right)$$

puisque la maximisation de la log-vraisemblance entraîne

$$\sum_{t=t_{\min}}^{t_{\max}} \hat{D}_{xt} = \sum_{t=t_{\min}}^{t_{\max}} D_{xt}.$$

#### Pseudo- $R^2$

La généralisation de la statistique usuelle  $R^2$  du modèle linéaire se base sur la décomposition de la déviance en

$$\mathcal{D}(D, D^*) = \mathcal{D}(D, \hat{D}) + \mathcal{D}(\hat{D}, D^*)$$

où

1.  $\mathcal{D}(D, D^*)$  est la déviance d'un modèle de base.
2.  $\mathcal{D}(D, \hat{D})$  est la déviance associée au modèle considéré;
3.  $\mathcal{D}(\hat{D}, D^*)$  est la déviance non expliquée par le modèle.

Cette décomposition mène à la définition

$$\text{pseudo-}R^2 = 1 - \frac{\mathcal{D}(D, \hat{D})}{\mathcal{D}(D, D^*)}.$$

La statistique pseudo- $R^2$  mesure donc la diminution de déviance qui résulte du choix d'un modèle de meilleure qualité que le modèle de base, par exemple en incorporant à ce dernier des variables explicatives. Par définition, le pseudo- $R^2$  est toujours compris entre 0 et 1, et est d'autant plus proche de 1 que le modèle choisi est adéquat.

Deux possibilités s'offrent à nous dans le choix du modèle de base. La première consiste à remplacer la spécification (3.32) par

$$\mu_x(t) = \exp(\alpha_x) \quad (3.39)$$

qui ne comporte aucune tendance temporelle. Les estimateurs des  $\alpha_x$  sont alors fournis par

$$\exp \hat{\alpha}_x = \frac{\sum_{t=t_{\min}}^{t_{\max}} D_{xt}}{\sum_{t=t_{\min}}^{t_{\max}} E_{xt}}$$

et  $\hat{D}_{xt} = E_{xt} \exp \hat{\alpha}_x$ . Le pseudo- $R^2$  obtenu peut alors être comparé au taux d'inertie (3.7) issu de décomposition en valeurs singulières pour le modèle (3.1) de Lee et Carter.

Le modèle de base peut également ne comprendre qu'un seul paramètre; il devient alors

$$\mu_x(t) = \exp(\alpha) \quad (3.40)$$

et l'estimateur de  $\alpha$  satisfait la relation

$$\exp \hat{\alpha} = \frac{\sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} D_{xt}}{\sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} E_{xt}}.$$

Les nombres de décès prédits par le modèle sont alors  $\hat{D}_{xt} = E_{xt} \exp \hat{\alpha}$ .

Remarquons que la qualité du pseudo- $R^2$  diminuera avec la pertinence du modèle de base. Ainsi, le pseudo- $R^2$  associé au modèle de base (3.40) sera meilleur que celui associé au modèle (3.39).

Avec nos données, le pseudo- $R^2$  associé à (3.39) vaut 91.33% tandis que celui associé à (3.40) vaut 99.76%. Ces valeurs laissent présager que le modèle (3.30)-(3.32) est adéquat pour nos données.

### Variance temporelle

Afin de rendre l'ajustement du modèle de Poisson (3.30)-(3.32) comparable avec celui du modèle de Lee et Carter (3.1), on peut à nouveau avoir recours à la part de variance temporelle expliquée. L'expression de celle-ci est la même pour les deux modèles sus-nommés. Elle vaut, pour chaque âge  $x$ ,

$$1 - \frac{\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} (\hat{\Delta}_x(t))^2 - \left( \frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} \hat{\Delta}_x(t) \right)^2}{\frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} (\hat{\mu}_x(t))^2 - \left( \frac{1}{t_{\max}-t_{\min}+1} \sum_{t=t_{\min}}^{t_{\max}} \hat{\mu}_x(t) \right)^2} \quad (3.41)$$

où  $\hat{\mu}_x(t)$  est l'estimation brute de  $\mu_x(t)$  et  $\hat{\Delta}_x(t) = \hat{\mu}_x(t) - \exp(\hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_t)$ .

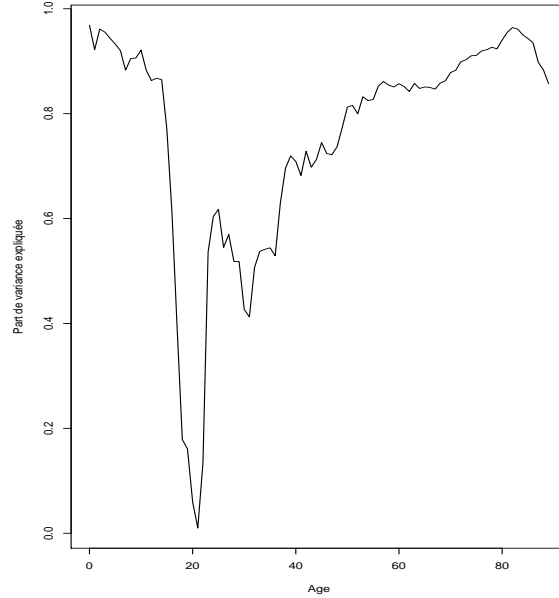


FIG. 3.29 – Part de variance temporelle expliquée par le modèle (3.30)-(3.32).

La courbe correspondant aux données françaises est reprise à la Figure 3.29. Elle est remarquablement proche de celle associée au modèle classique de Lee et Carter (reprise à la Figure 3.8).

### 3.3.5 Analyse des résidus

Dans le cadre d'un modèle de régression, il importe de vérifier que les résidus ne présentent aucune structure. Si tel n'est pas le cas, notre modèle n'est pas à même de rendre compte d'une partie du mécanisme prévalant à l'évolution de la mortalité. Typiquement, c'est l'indice temporel unidimensionnel qui pourrait ne pas contenir toute l'information à propos de l'évolution de la mortalité, ou le fait que les  $\beta_x$  pourraient dépendre également du temps calendaire.

Il est donc d'usage d'inspecter les résidus de déviance  $r_{xt}^{deviance}$ . Ceux-ci sont définis par la racine carrée signée de la contribution de chaque couple  $(x,t)$  à la déviance du modèle, i.e.

$$r_{xt}^{deviance} = \sqrt{2} \cdot \text{sign}(D_{xt} - \hat{D}_{xt}) \sqrt{D_{xt} \ln \frac{D_{xt}}{\hat{D}_{xt}} - (D_{xt} - \hat{D}_{xt})}.$$

Remarquons que, par définition, nous obtenons la relation

$$\mathcal{D} = \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} (r_{xt}^{deviance})^2.$$

L'analyse de l'évolution temporelle des résidus de déviance pour quelques âges choisis peut dans certains cas mettre en évidence une structure résiduelle. Le cas échéant, le modèle choisi peut s'avérer inopportun, voire nécessiter une extension telle celle de Renshaw-Haberman développée dans la section 3.3.7. La Figure 3.30 reprend les résidus de déviance pour les données françaises. On y retrouve les mêmes signes de structure que pour les résidus du modèle classique de Lee et Carter (à la Figure 3.9).

Il est également possible d'effectuer la même analyse sur les résidus de Pearson, lesquels sont définis par

$$r_{xt}^{Pearson} = \frac{D_{xt} - \hat{D}_{xt}}{\sqrt{\hat{D}_{xt}}}.$$

Les résultats sont en général très similaires à ceux obtenus par les résidus de déviance.



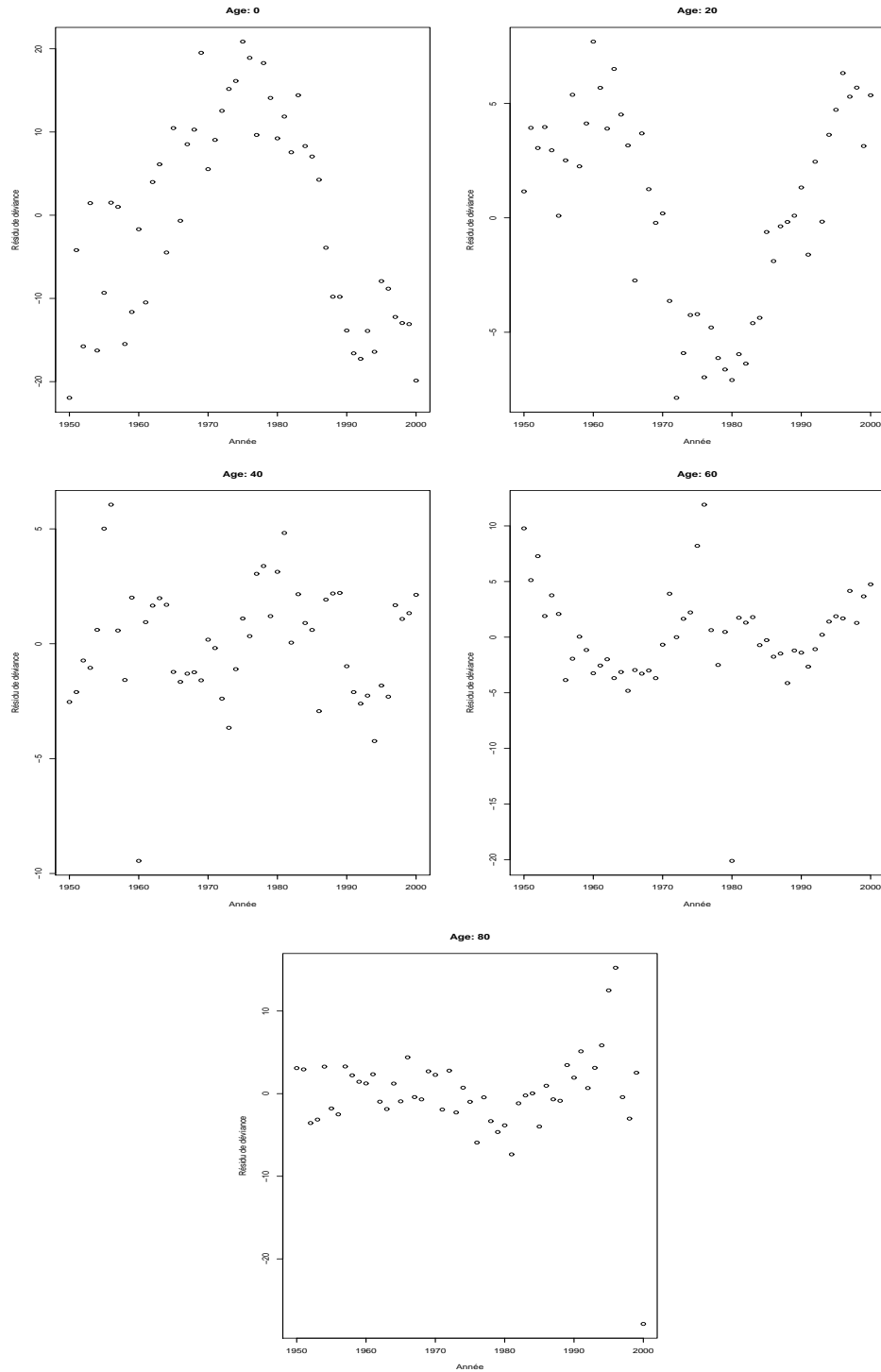


FIG. 3.30 – Résidus de déviance pour quelques âges.

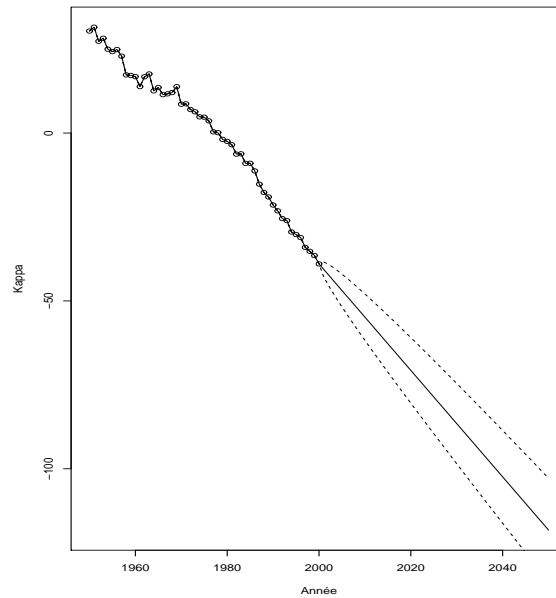


FIG. 3.31 – Prédiction des  $\hat{\kappa}_t$  à l'aide du modèle de marche aléatoire avec tendance.

### 3.3.6 Extrapolation de la tendance temporelle

Cette phase est en tout point identique à celle de la méthode de Lee et Carter.

La première étape de la modélisation consiste à transformer  $\mathcal{K}$  adéquatement afin d'atteindre la stationnarité faible d'ordre 2. Ensuite, un modèle doit être identifié à l'aide des graphiques de fonctions d'autocorrélation et d'autocorrélation partielle, ainsi que des critères *AIC* et *BIC*. Les paramètres du modèle ARIMA sélectionné sont alors estimés et confrontés à un test de nullité. Finalement, après examen attentif des résidus, il peut être procédé à l'extrapolation de la série.

Comme précédemment, nous choisissons comme modèle la marche aléatoire avec tendance. La Figure 3.31 montre les prédictions de  $\hat{\kappa}_t$  ainsi que des intervalles de confiance à 95%. Les taux instantanés de mortalité futurs sont pour leur part repris à la Figure 3.32.

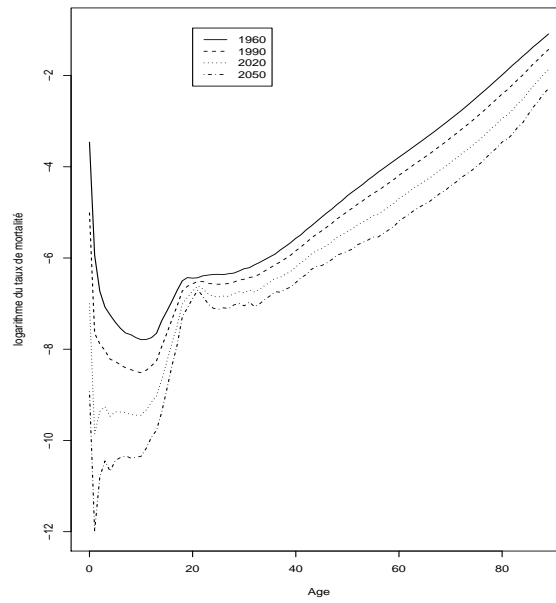


FIG. 3.32 – Prédiction des logarithmes des taux instantanés de mortalité pour quelques années à l'aide du modèle de marche aléatoire avec tendance.

### 3.3.7 Extension de Renshaw-Haberman

Dans le cas où les graphes des résidus laissent apparaître une structure, Renshaw et Haberman proposent d'ajouter une seconde tendance temporelle. La spécification (3.32) devient alors

$$\mu_x(t) = \exp(\alpha_x + \beta_{x,1}\kappa_{t,1} + \beta_{x,2}\kappa_{t,2}) \quad (3.42)$$

et il faut recourir aux contraintes (3.23) et (3.24) pour garantir l'identifiabilité du modèle.

La fonction de log-vraisemblance devient donc, à un facteur additif près,

$$\begin{aligned} & \ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) \\ &= \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt}(\alpha_x + \beta_{x,1}\kappa_{t,1} + \beta_{x,2}\kappa_{t,2}) - E_{xt} \exp(\alpha_x + \beta_{x,1}\kappa_{t,1} + \beta_{x,2}\kappa_{t,2}) \right). \end{aligned}$$

L'estimation des paramètres par maximum de vraisemblance peut être réalisée en utilisant un algorithme du même type que (3.33), dont les itérations sont

$$\begin{aligned} \hat{\alpha}_x^{(\nu+1)} &= \hat{\alpha}_x^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right)}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right)} \\ \hat{\kappa}_{t,1}^{(\nu+1)} &= \hat{\kappa}_{t,1}^{(\nu)} - \frac{\sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \hat{\beta}_{x,1}^{(\nu)}}{-\sum_{x=x_{\min}}^{x_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \left( \hat{\beta}_{x,1}^{(\nu)} \right)^2} \\ \hat{\beta}_{x,1}^{(\nu+1)} &= \hat{\beta}_{x,1}^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \hat{\kappa}_{t,1}^{(\nu+1)}}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \left( \hat{\kappa}_{t,1}^{(\nu+1)} \right)^2} \quad (3.43) \\ \hat{\kappa}_{t,2}^{(\nu+1)} &= \hat{\kappa}_{t,2}^{(\nu)} - \frac{\sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu+1)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \hat{\beta}_{x,2}^{(\nu)}}{-\sum_{x=x_{\min}}^{x_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu+1)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu)}) \right) \left( \hat{\beta}_{x,2}^{(\nu)} \right)^2} \\ \hat{\beta}_{x,2}^{(\nu+1)} &= \hat{\beta}_{x,2}^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu+1)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu+1)}) \right) \hat{\kappa}_{t,2}^{(\nu+1)}}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_{x,1}^{(\nu+1)} \hat{\kappa}_{t,1}^{(\nu+1)} + \hat{\beta}_{x,2}^{(\nu)} \hat{\kappa}_{t,2}^{(\nu+1)}) \right) \left( \hat{\kappa}_{t,2}^{(\nu+1)} \right)^2}. \end{aligned}$$

À nouveau, le choix des paramètres initiaux  $\hat{\alpha}_x^{(0)}$ ,  $\hat{\kappa}_{t,1}^{(0)}$  et  $\hat{\kappa}_{t,2}^{(0)}$  est libre, tandis que  $\hat{\beta}_{x,1}^{(0)}$  et  $\hat{\beta}_{x,2}^{(0)}$  ne peuvent être nuls.

Afin de satisfaire les contraintes (3.23)-(3.24), il faut à nouveau faire subir une transformation aux estimateurs issus de (3.43):

$$\begin{aligned}
\widehat{\alpha}_x &\leftarrow \widehat{\alpha}_x + \widehat{\beta}_{x,1}\overline{\kappa}_1 + \widehat{\beta}_{x,2}\overline{\kappa}_2 \\
\widehat{\kappa}_{t,1} &\leftarrow (\widehat{\kappa}_{t,1} - \overline{\kappa}_1)\widehat{\beta}_{\bullet,1} \\
\widehat{\beta}_{x,1} &\leftarrow \frac{\widehat{\beta}_{x,1}}{\widehat{\beta}_{\bullet,1}} \\
\widehat{\kappa}_{t,2} &\leftarrow (\widehat{\kappa}_{t,2} - \overline{\kappa}_2)\widehat{\beta}_{\bullet,2} \\
\widehat{\beta}_{x,2} &\leftarrow \frac{\widehat{\beta}_{x,2}}{\widehat{\beta}_{\bullet,2}}
\end{aligned}$$

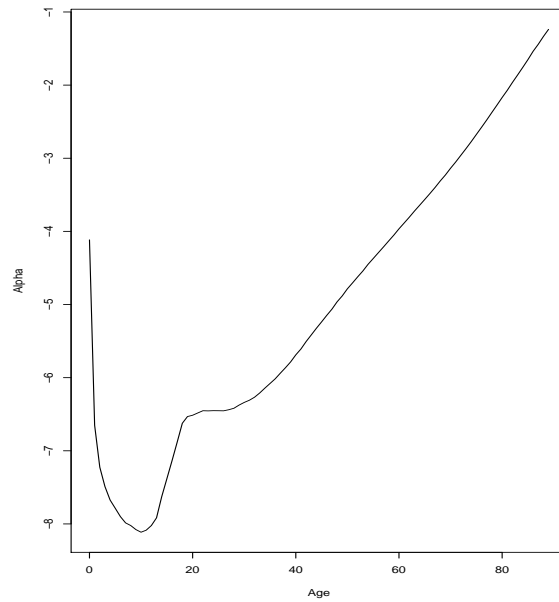
avec comme d'habitude  $\overline{\kappa}_1$  et  $\overline{\kappa}_2$  les moyennes des  $\widehat{\kappa}_{t,1}$  et  $\widehat{\kappa}_{t,2}$  initiaux, et  $\widehat{\beta}_{\bullet,1}$  et  $\widehat{\beta}_{\bullet,2}$  les sommes de  $\widehat{\beta}_{x,1}$  et  $\widehat{\beta}_{x,2}$  initiaux.

Bien évidemment, l'augmentation du nombre de paramètres accroîtra la qualité de l'ajustement. Néanmoins, une telle approche souffre d'un défaut majeur. Les deux indices temporels constituent autant de séries chronologiques vraisemblablement corrélées. Leur extrapolation dans le futur est ainsi rendue beaucoup plus complexe que dans le cas d'une série unique.

Remarquons finalement que, contrairement au cas du modèle de Lee et Carter, les séries  $\widehat{\beta}_x$  du modèle classique et  $\widehat{\beta}_{x,1}$  du modèle étendu peuvent être singulièrement différentes. Il en va de même pour les séries  $\widehat{\kappa}_t$  et  $\widehat{\kappa}_{t,1}$ . En effet, dans le cas du modèle de Poisson, l'estimation des paramètres du modèle étendu (3.30)-(3.42) s'effectue indépendamment de celle des paramètres du modèle classique (3.30)-(3.32). La situation est tout à fait différente pour le modèle de Lee et Carter où, au moyen d'une unique décomposition en valeurs singulières, l'entière des paramètres  $\alpha_x$ ,  $\beta_x = \beta_{x,1}$ ,  $\kappa_t = \kappa_{t,1}$ ,  $\beta_{x,2}$  et  $\kappa_{t,2}$  est estimée.

Les Figures 3.33 et 3.34 représentent les nouveaux estimateurs et sont à comparer avec les Figures 3.25 et 3.26 du modèle initial. On y constate des courbes de  $\widehat{\kappa}_{t,1}$  et  $\widehat{\kappa}_{t,2}$  beaucoup plus chaotiques que dans le modèle initial. De plus, le comportement de la mortalité aux jeunes âges semble une nouvelle fois atypique, ce qui nous conforte dans l'idée que le modèle ne devrait sans doute être appliqué qu'aux plus de 40 ou 50 ans.

Le gain en terme de part de variance expliquée est visible à la Figure 3.35. Comme dans le modèle de Lee et Carter, l'introduction d'un second terme bilinéaire dans le modèle initial y influence majoritairement les âges compris entre 15 et 40 ans. Enfin, nous voyons à la Figure 3.36 que la structure des résidus du modèle à deux termes bilinéaires est sensiblement plus faible que celle des résidus du modèle initial.

FIG. 3.33 – Estimation des  $\alpha_x$  pour le modèle (3.30)-(3.42).

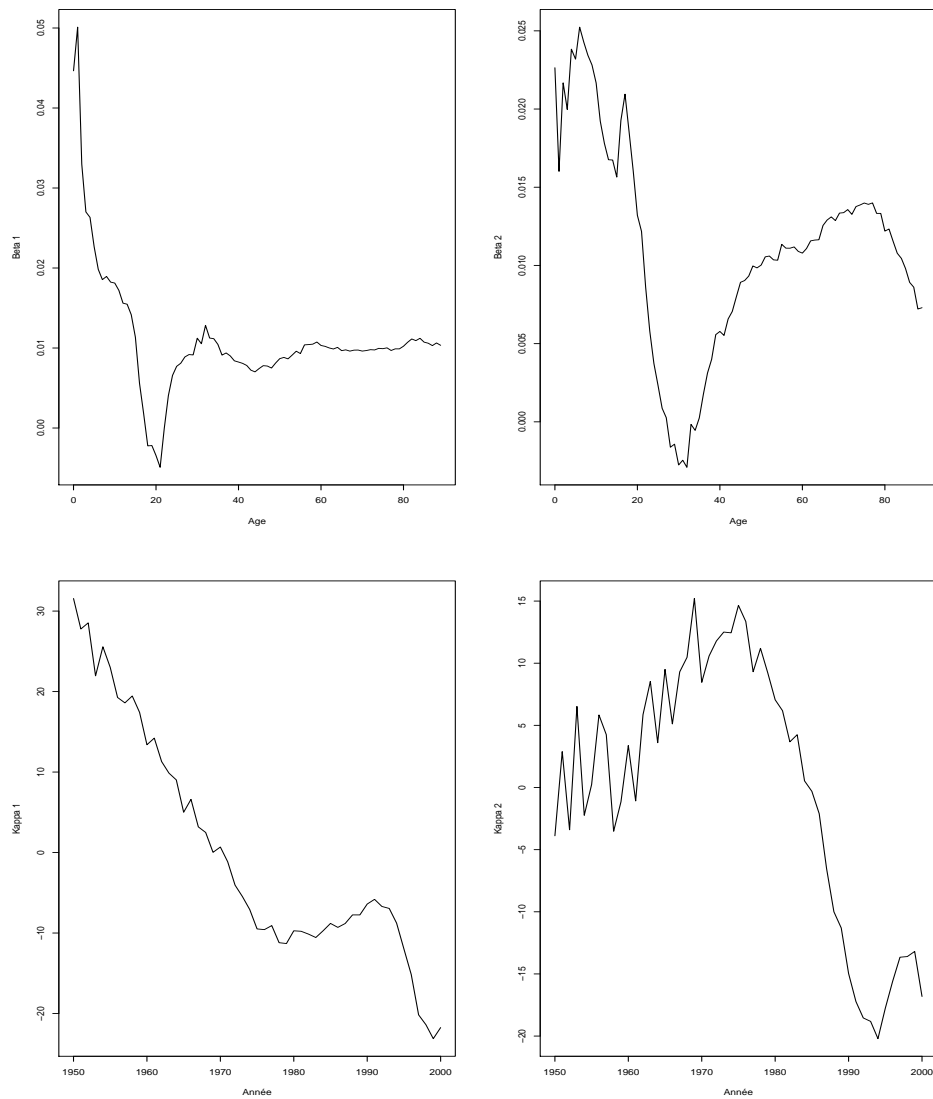


FIG. 3.34 – Estimation des  $\beta_{x,1}$ ,  $\beta_{x,2}$ ,  $\kappa_{t,1}$  et  $\kappa_{t,2}$  pour le modèle (3.30)-(3.42).

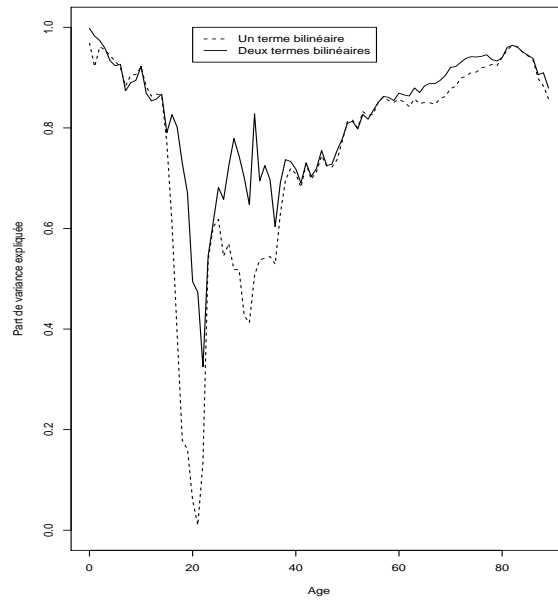


FIG. 3.35 – Part de variance temporelle expliquée par le modèle (3.30)-(3.42).



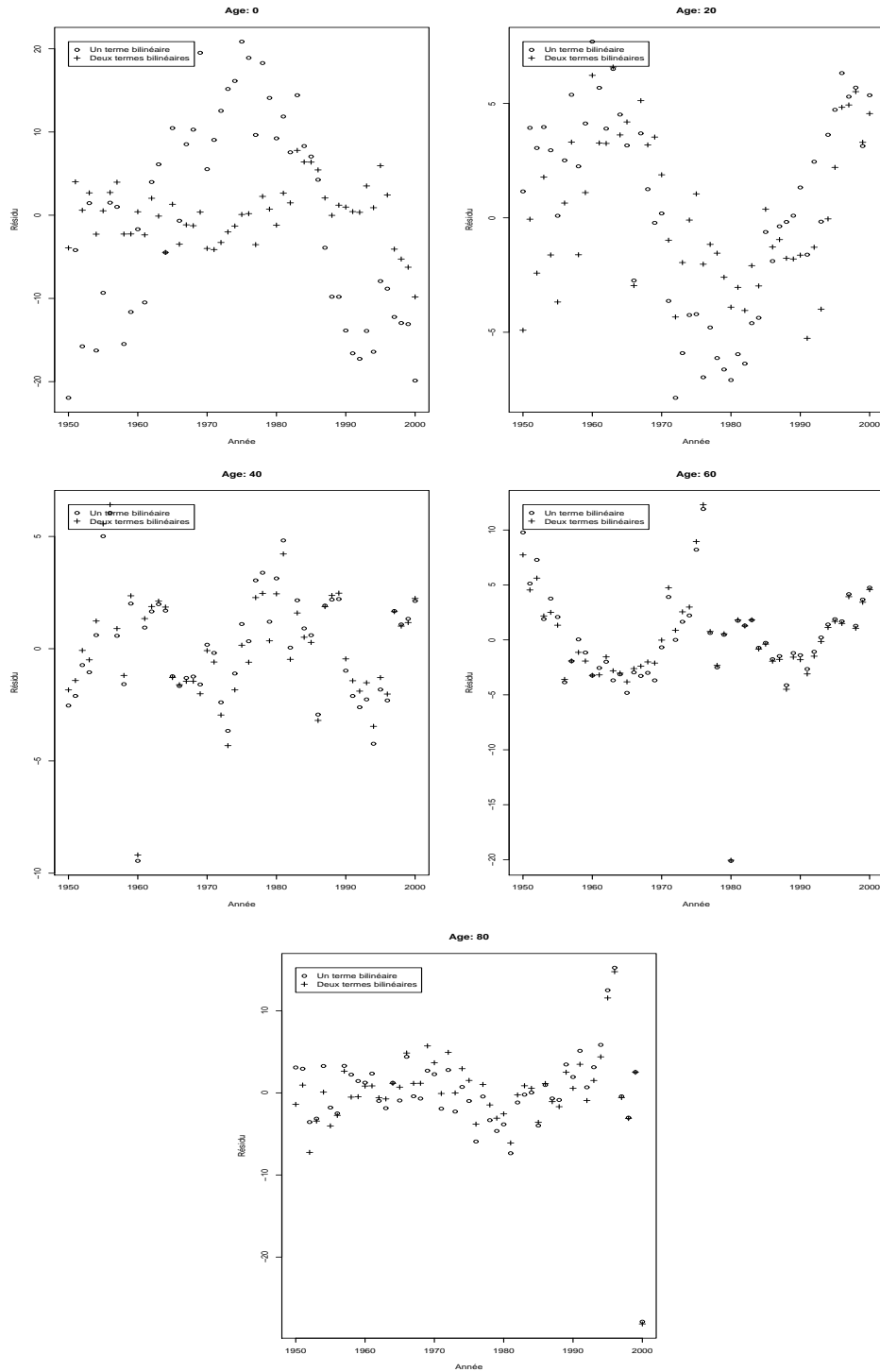


FIG. 3.36 – Résidus du modèle à deux termes bilinéaires pour quelques âges.

### 3.3.8 Modèle avec tendance temporelle linéaire

Comme le modèle classique de Lee et Carter, le modèle de Poisson log-bilinéaire peut être adapté de manière à tenir compte d'une tendance temporelle linéaire. (3.30) est alors complété par

$$\mu_x(t) = \exp(\alpha_x + \beta_x(a + bt)). \quad (3.44)$$

Notez qu'une telle spécification ne nécessite plus de contraintes telles (3.2)-(3.3), celles-ci étant implicitement remplacées par l'imposition des deux paramètres  $a$  et  $b$  de la tendance linéaire.

Dans la situation simplifiée où la tendance linéaire est égale au temps calendaire  $t$ , ces coefficients prennent les valeurs  $a = 0$  et  $b = 1$  et le modèle devient

$$\mu_x(t) = \exp(\alpha_x + \beta_x t) \quad (3.45)$$

qui revient à (3.44) par la transformation

$$\begin{aligned} \alpha_x &\leftarrow \alpha_x + a\beta_x \\ \beta_x &\leftarrow b\beta_x. \end{aligned}$$

L'estimation des paramètres  $\alpha_x$  et  $\beta_x$  par maximum de vraisemblance peut maintenant être réalisée à l'aide d'un logiciel traitant de modèles linéaires généralisés. Néanmoins, il est possible de recourir au même style d'algorithme itératif que précédemment. La fonction de log-vraisemblance associée à ce nouveau modèle devient, à un facteur additif près,

$$\ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt}(\alpha_x + \beta_x t) - E_{xt} \exp(\alpha_x + \beta_x t) \right)$$

et les itérations y associées sont donc

$$\begin{aligned} \hat{\alpha}_x^{(\nu+1)} &= \hat{\alpha}_x^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_x^{(\nu)} t) \right)}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_x^{(\nu)} t) \right)} \\ \hat{\beta}_x^{(\nu+1)} &= \hat{\beta}_x^{(\nu)} - \frac{\sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} - E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_x^{(\nu)} t) \right) t}{-\sum_{t=t_{\min}}^{t_{\max}} \left( E_{xt} \exp(\hat{\alpha}_x^{(\nu+1)} + \hat{\beta}_x^{(\nu)} t) \right) t^2}. \end{aligned} \quad (3.46)$$

Il est également possible de modifier les estimateurs obtenus pour les rendre compatibles avec les contraintes (3.2)-(3.3). Cette opération s'effectue par les transformations habituelles:

$$\begin{aligned} \hat{\alpha}_x &\leftarrow \hat{\alpha}_x + \hat{\beta}_x \bar{t} \\ \hat{\kappa}_t &\leftarrow (t - \bar{t}) \hat{\beta}_\bullet \\ \hat{\beta}_x &\leftarrow \frac{\hat{\beta}_x}{\hat{\beta}_\bullet} \end{aligned}$$

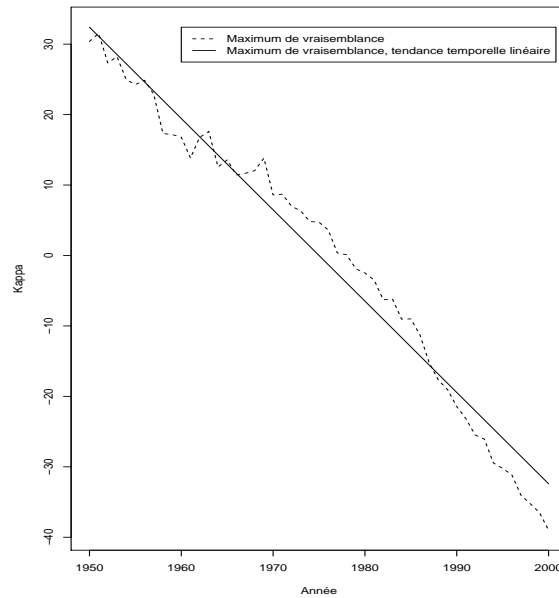


FIG. 3.37 – Estimation des  $\kappa_t$  pour le modèle à tendance temporelle linéaire.

où  $\bar{t} = \frac{t_{\min} + t_{\max}}{2}$  est la moyenne des années retenues dans le modèle et  $\hat{\beta}_\bullet$  la somme des  $\hat{\beta}_x$  initiaux. Les estimateurs de la tendance temporelle  $\hat{\kappa}_t$  ne sont donc plus égaux au temps calendaire  $t$  comme le spécifiait (3.45), mais ils restent cependant linéaires.

L'application de ce nouveau modèle à nos données est illustrée aux Figures 3.37 et 3.38. Comme pour le modèle de Lee et Carter, il serait sans doute préférable de n'ajuster le modèle (3.30)-(3.44) qu'aux données au-delà de 1970. Une nouvelle borne  $t_{\min}^*$  pour la période d'observation peut par exemple être estimée par (3.28).

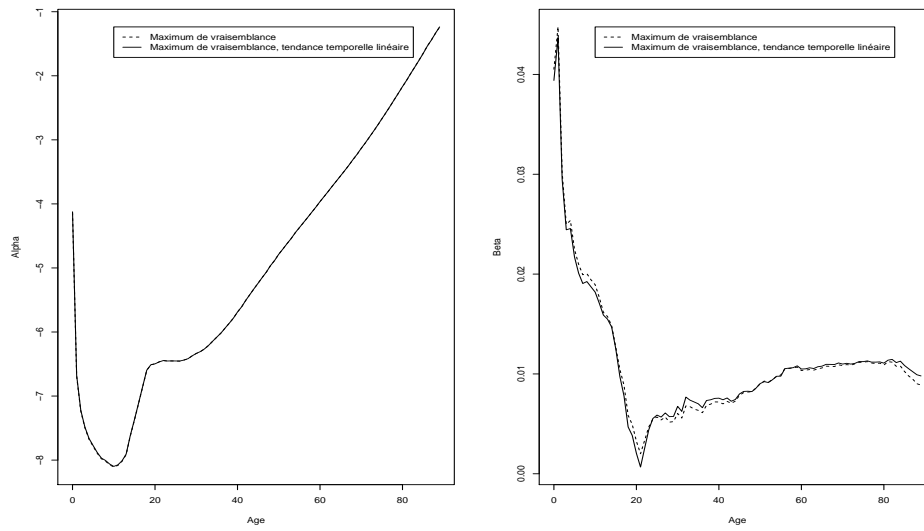


FIG. 3.38 – Estimation des  $\alpha_x$  et  $\beta_x$  pour le modèle à tendance temporelle linéaire.

### 3.3.9 Modèle avec table limite

Comme nous l'avons vu précédemment, le comportement asymptotique de la mortalité projetée par le modèle log-bilinéaire n'est pas toujours satisfaisant.

Supposons donc que nous ayons une estimation du comportement asymptotique de la mortalité, et notons  $\mu_x^{lim}$  et  $q_x^{lim}$  les taux instantanés et quotients  $y$  correspondant. Cette table représente donc un niveau de mortalité sous lequel il sera impossible de descendre dans le futur. Ainsi, l'incorporation de cet artifice dans (3.30)-(3.32) devrait nous permettre de contrôler le niveau asymptotique de la mortalité projetée. Le modèle devient donc

$$D_{xt} \sim \text{Poisson}\left(E_{xt}(\mu_x^{lim} + \mu_x(t))\right) \quad (3.47)$$

associé à (3.32) et aux contraintes (3.2)-(3.3). En notant

$$\lambda_{xt} = E_{xt}(\mu_x^{lim} + \exp(\alpha_x + \beta_x \kappa_t)),$$

la fonction de log-vraisemblance devient, à un facteur additif près,

$$\begin{aligned} & \ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) \\ = & \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} \left( D_{xt} \ln(\mu_x^{lim} + \exp(\alpha_x + \beta_x \kappa_t)) - E_{xt} \exp(\alpha_x + \beta_x \kappa_t) \right). \end{aligned}$$

Un nouvel algorithme itératif pour estimer les paramètres peut donc être imaginé. En effectuant le même raisonnement que pour les algorithmes précédents, et en notant  $\hat{\mu}_x^{(\nu_\alpha, \nu_\beta, \nu_\kappa)}(t) = \exp(\hat{\alpha}_x^{(\nu_\alpha)} + \hat{\beta}_x^{(\nu_\beta)} \hat{\kappa}_t^{(\nu_\kappa)})$ , nous trouvons les itérations suivantes:

$$\begin{aligned}
 \widehat{\alpha}_x^{(\nu+1,1)} &= \sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} \frac{\widehat{\mu}_x^{(\nu,\nu,\nu)}(t)}{\mu_x^{lim} + \widehat{\mu}_x^{(\nu,\nu,\nu)}(t)} - E_{xt} \widehat{\mu}_x^{(\nu,\nu,\nu)}(t) \right) \\
 \widehat{\alpha}_x^{(\nu+1,2)} &= \sum_{t=t_{\min}}^{t_{\max}} \left( D_{xt} \frac{\mu_x^{lim} \widehat{\mu}_x^{(\nu,\nu,\nu)}(t)}{(\mu_x^{lim} + \widehat{\mu}_x^{(\nu,\nu,\nu)}(t))^2} - E_{xt} \widehat{\mu}_x^{(\nu,\nu,\nu)}(t) \right) \\
 \widehat{\alpha}_x^{(\nu+1)} &= \widehat{\alpha}_x^{(\nu)} - \frac{\widehat{\alpha}_x^{(\nu+1,1)}}{\widehat{\alpha}_x^{(\nu+1,2)}} \\
 \widehat{\kappa}_t^{(\nu+1,1)} &= \sum_{x=x_{\min}}^{x_{\max}} \widehat{\beta}_x^{(\nu)} \left( D_{xt} \frac{\widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t)}{\mu_x^{lim} + \widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t)} - E_{xt} \widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t) \right) \\
 \widehat{\kappa}_t^{(\nu+1,2)} &= \sum_{x=x_{\min}}^{x_{\max}} (\widehat{\beta}_x^{(\nu)})^2 \left( D_{xt} \frac{\mu_x^{lim} \widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t)}{(\mu_x^{lim} + \widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t))^2} - E_{xt} \widehat{\mu}_x^{(\nu+1,\nu,\nu)}(t) \right) \\
 \widehat{\kappa}_t^{(\nu+1)} &= \widehat{\kappa}_t^{(\nu)} - \frac{\widehat{\kappa}_t^{(\nu+1,1)}}{\widehat{\kappa}_t^{(\nu+1,2)}} \\
 \widehat{\beta}_x^{(\nu+1,1)} &= \sum_{t=t_{\min}}^{t_{\max}} \widehat{\kappa}_t^{(\nu+1)} \left( D_{xt} \frac{\widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t)}{\mu_x^{lim} + \widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t)} - E_{xt} \widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t) \right) \\
 \widehat{\beta}_x^{(\nu+1,2)} &= \sum_{t=t_{\min}}^{t_{\max}} (\widehat{\kappa}_t^{(\nu+1)})^2 \left( D_{xt} \frac{\mu_x^{lim} \widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t)}{(\mu_x^{lim} + \widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t))^2} - E_{xt} \widehat{\mu}_x^{(\nu+1,\nu,\nu+1)}(t) \right) \\
 \widehat{\beta}_x^{(\nu+1)} &= \widehat{\beta}_x^{(\nu)} - \frac{\widehat{\beta}_x^{(\nu+1,1)}}{\widehat{\beta}_x^{(\nu+1,2)}}. \tag{3.48}
 \end{aligned}$$

Il faut également recourir à une modification des estimateurs ainsi obtenus afin de respecter les contraintes (3.2)-(3.3):

$$\begin{aligned}
 \widehat{\alpha}_x &\leftarrow \widehat{\alpha}_x + \widehat{\beta}_x \bar{\kappa} \\
 \widehat{\kappa}_t &\leftarrow (\widehat{\kappa}_t - \bar{\kappa}) \widehat{\beta}_\bullet \\
 \widehat{\beta}_x &\leftarrow \frac{\widehat{\beta}_x}{\widehat{\beta}_\bullet}
 \end{aligned}$$

où  $\bar{\kappa}$  est la moyenne des  $\widehat{\kappa}_t$  initiaux et  $\widehat{\beta}_\bullet$  la somme des  $\widehat{\beta}_x$  initiaux.

L'application du modèle (3.47)-(3.32) aux données françaises avec la table limite de DUCHÊNE & WUNSCH (1988) donne les graphiques des Figures 3.39 et 3.40. Comme prévu, les principales différences avec les estimateurs de (3.30)-(3.32) se situent aux âges où la table limite est significativement non nulle, soit à partir de 60 ans. L'influence sur les  $\widehat{\kappa}_t$  est relativement faible, même si ceux-ci semblent décroître légèrement plus vite.

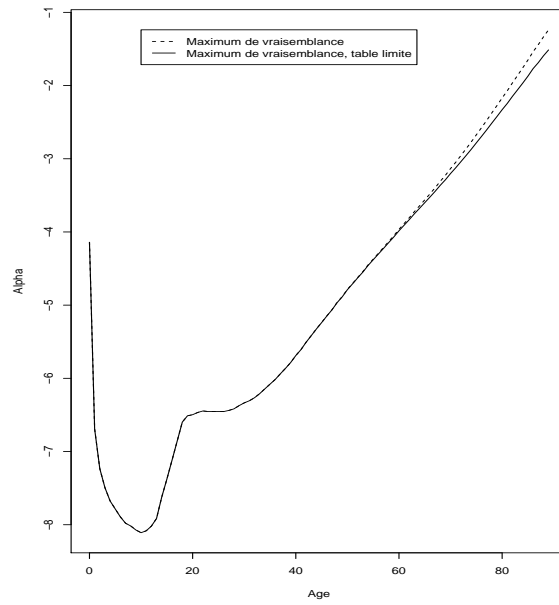


FIG. 3.39 – Estimation des  $\alpha_x$  pour le modèle avec table limite.

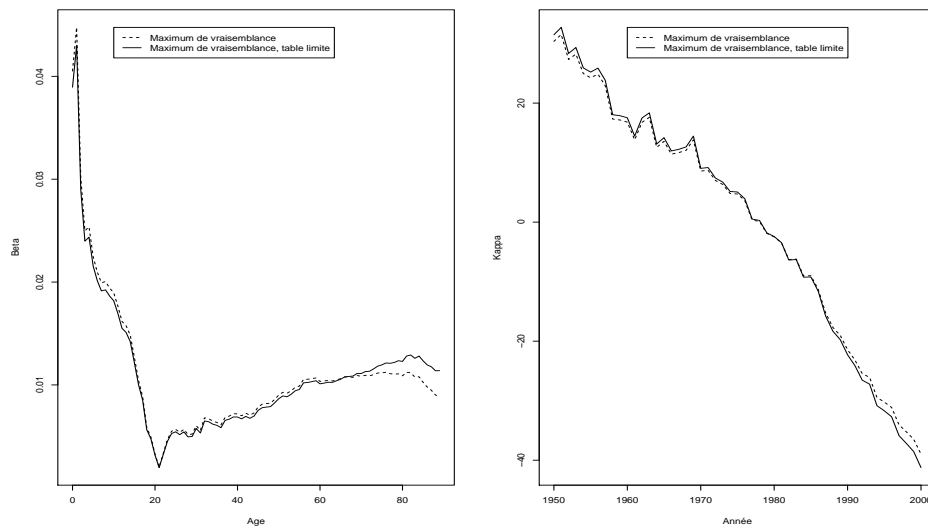


FIG. 3.40 – Estimation des  $\beta_x$  et  $\kappa_t$  pour le modèle avec table limite.

## Chapitre 4

# Intervalles de confiance pour les indicateurs de mortalité

En prédiction, il est important de fournir de l'information sur la variabilité des quantités projetées. Dans le cas présent, ceci permettra à l'actuaire d'évaluer le degré de confiance qu'il peut accorder à la projection de la prime relative à la rente viagère, et donc de choisir la marge de sécurité adéquate, voire d'opter pour un programme de réassurance approprié. L'outil le plus indiqué est certainement l'intervalle de confiance. L'objectif de ce chapitre est donc de construire de tels intervalles pour les indicateurs de mortalité (comme l'espérance de vie) à partir du modèle de Poisson log-bilinéaire.

Remarquons d'emblée qu'il est impossible d'obtenir analytiquement les intervalles de confiance recherchés. Il faut effectivement combiner deux sources très différentes de variabilité : la première est constituée par les erreurs d'échantillonnage relatives aux paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  tandis que la seconde réside dans les erreurs de prédiction des  $\kappa_t$ . En outre, les indicateurs qui nous intéressent sont des fonctions non-linéaires relativement complexes des paramètres tant du modèle de Poisson que du modèle de prédiction. Par conséquent, nous aurons recours à la simulation.

### 4.1 Approche par simulation Monte Carlo

#### 4.1.1 Principe

La première approche développée ci-dessous considère que le modèle (3.30)-(3.32) est adéquat. Étant donné que les paramètres ont été estimés par la méthode du maximum de vraisemblance, nous savons qu'ils obéissent approximativement à la loi normale multivariée, de moyenne  $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa})$  et de matrice de variance-covariance l'inverse de la matrice d'information de Fisher  $\mathcal{I}$ . Les éléments de cette dernière valent l'opposé de l'espérance des dérivées secondes de la log-vraisemblance  $\ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa})$  par rapport aux paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$ . Les détails relatifs au calcul de cette matrice sont repris plus bas.



Ensuite, la seconde source de variabilité concerne le modèle de prédiction. Dans le cas d'une projection ARIMA, toute l'information nécessaire est contenue dans les estimateurs des paramètres et de la variance  $\sigma^2$  du bruit blanc.

Notons  $\alpha_x^m$ ,  $\beta_x^m$  et  $\kappa_t^m$  les valeurs des paramètres du  $m^{\text{ème}}$  échantillon,  $m = 1, \dots, M$ , jusqu'à  $t = t_{\max}$  et  $\kappa_t^m$ ,  $t \geq t_{\max} + 1$ , les projections de l'indice temporel dans le futur. La simulation Monte Carlo peut alors être résumée en quatre étapes :

1. Générer  $\alpha_x^m$ ,  $\beta_x^m$  et  $\kappa_t^m$  à partir de la loi gaussienne multivariée à laquelle obéit asymptotiquement l'estimateur du maximum de vraisemblance.
2. Estimer les paramètres du modèle ARIMA dont les ordres ont été sélectionnés sur base des observations en utilisant la série  $\mathcal{K}^m = \{\kappa_t^m, t = t_{\min}, \dots, t_{\max}\}$ .
3. Générer les projections  $\kappa_t^m$  en utilisant les paramètres ARIMA estimés à l'étape 2. Les erreurs  $\epsilon_t^m$  sont simulées à partir de la loi normale  $\mathcal{N}(0, \sigma^m)$ .
4. Calculer pour chaque échantillon la valeur de l'indicateur désiré.

La première étape modélise le risque sur les paramètres du modèle (3.30)-(3.32). La seconde étape permet de considérer le fait que le risque sur les paramètres ARIMA dépend du risque sur les paramètres du modèle de Poisson. Par la troisième étape, nous considérons que le risque sur les prédictions  $\kappa_t$  ne dépend pas uniquement de l'erreur standard du modèle ARIMA, mais également du risque sur les paramètres ARIMA. Les calculs des indicateurs de mortalité à la quatrième étape permettent donc de prendre en compte tous les types de risque.

#### 4.1.2 Calcul de la matrice d'information de Fisher

Afin de simplifier les notations, nous allons reformuler le modèle (3.30)-(3.32) d'une manière légèrement différente. Nous modifions notre paramétrisation et choisissons de remplacer la contrainte (3.2) par

$$\beta_{x_{\min}} = 1. \quad (4.1)$$

Les nouveaux estimateurs sont facilement obtenus par le changement de variable

$$\begin{aligned} \widehat{\kappa}_t &\leftarrow \widehat{\kappa}_t \widehat{\beta}_{x_{\min}} \\ \widehat{\beta}_x &\leftarrow \frac{\widehat{\beta}_x}{\widehat{\beta}_{x_{\min}}}. \end{aligned}$$

Réécrivons le nombre attendu de décès d'individus d'âge  $x$  au cours de l'année  $t$ :

$$\begin{aligned} \lambda_{xt} &= E_{xt} \exp(\alpha_x + \beta_x \kappa_t) \\ &= E_{xt} \exp \left[ \left( \sum_{y=x_{\min}}^{x_{\max}} a_{xy} \alpha_y \right) + \left( \sum_{y=x_{\min}}^{x_{\max}} b_{xy} \beta_y \right) \left( \sum_{r=t_{\min}}^{t_{\max}-1} k_{tr} \kappa_r \right) \right] \end{aligned}$$

où  $a_{xy}$ ,  $b_{xy}$  et  $k_{tr}$  sont les éléments des trois matrices de pondération  $\mathbf{A}$  et  $\mathbf{B}$  de dimensions  $(x_{\max} - x_{\min} + 1) \times (x_{\max} - x_{\min} + 1)$ , et  $\mathbf{K}$  de dimensions  $(t_{\max} - t_{\min} + 1) \times (t_{\max} - t_{\min})$ . Afin de satisfaire les contraintes (3.3) et (4.1), il est nécessaire de définir ces matrices de la manière suivante :

$$\mathbf{A} = \mathbf{B} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \text{ et } \mathbf{K} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ -1 & \dots & -1 \end{pmatrix}.$$

Les éléments de la matrice d'information de Fisher pour les paramètres libres  $\alpha_{x_{\min}}, \dots, \alpha_{x_{\max}}, \beta_{x_{\min}+1}, \dots, \beta_{x_{\max}}$  et  $\kappa_{t_{\min}}, \dots, \kappa_{t_{\max}-1}$  sont donc donnés par

$$\begin{aligned} -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \alpha_y \partial \alpha_{y'}} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt} a_{xy} a_{xy'} \\ -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \beta_y \partial \beta_{y'}} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt} (\kappa_t b_{xy}) (\kappa_t b_{xy'}) \\ -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \kappa_r \partial \kappa_{r'}} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt} (\beta_x k_{tr}) (\beta_x k_{tr'}) \\ -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \alpha_y \partial \beta_{y'}} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt} a_{xy} (\kappa_t b_{xy'}) \\ -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \alpha_y \partial \kappa_r} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt} a_{xy} (\beta_x k_{tr}) \\ -\mathbb{E} \left( \frac{\partial^2 \ln \mathcal{V}}{\partial \beta_y \partial \kappa_r} \right) &= \sum_{x=x_{\min}}^{x_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \lambda_{xt}(t) (\kappa_t b_{xy}) (\beta_x k_{tr}). \end{aligned}$$

En remplaçant  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  par leurs estimateurs, nous trouvons une estimation de la matrice d'information de Fisher  $\widehat{\mathcal{I}}$ . Enfin, après avoir généré une simulation des paramètres libres, les valeurs simulées de  $\kappa_{t_{\max}}$  découlent directement de (3.3). Il est alors aisé de revenir de (4.1) à la paramétrisation d'origine satisfaisant la contrainte (3.2) en effectuant la transformation

$$\begin{aligned} \widehat{\kappa}_t &\leftarrow \widehat{\kappa}_t \widehat{\beta}_\bullet \\ \widehat{\beta}_x &\leftarrow \frac{\widehat{\beta}_x}{\widehat{\beta}_\bullet} \end{aligned}$$

où  $\widehat{\beta}_\bullet$  est la somme des  $\widehat{\beta}_x$  satisfaisant (4.1).

La Figure 4.1 montre les logarithmes des taux de mortalité générés par  $M = 1000$  échantillons. Naturellement, ceux-ci sont beaucoup plus variables lorsque nous dépassons l'année  $t_{\max}$ , puisque les simulations des erreurs du modèle ARIMA viennent s'y greffer. À la Figure 4.2, nous trouvons les intervalles de

confiance à 95% et les comparons avec les estimations du modèle classique de Lee et Carter et de celui de Poisson log-bilinéaire. Remarquons que les espérances de vie en 1950 générées par le modèle de Lee et Carter ne se situent pas dans ces intervalles. Ceci prouve qu'il existe une différence significative entre l'ajustement des deux modèles, du moins pour les données françaises.

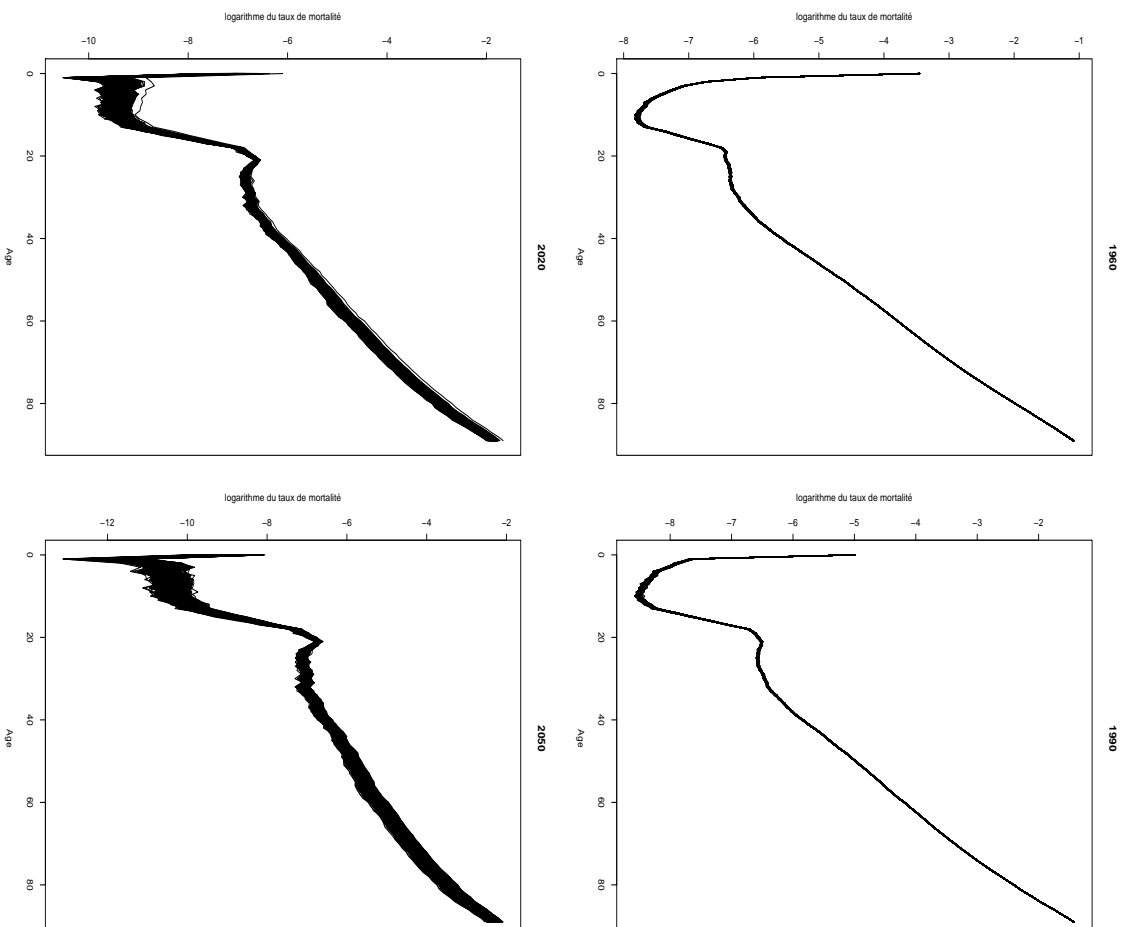


FIG. 4.1 – Prédiction des logarithmes des taux de mortalité pour quelques années par les simulations Monte Carlo.

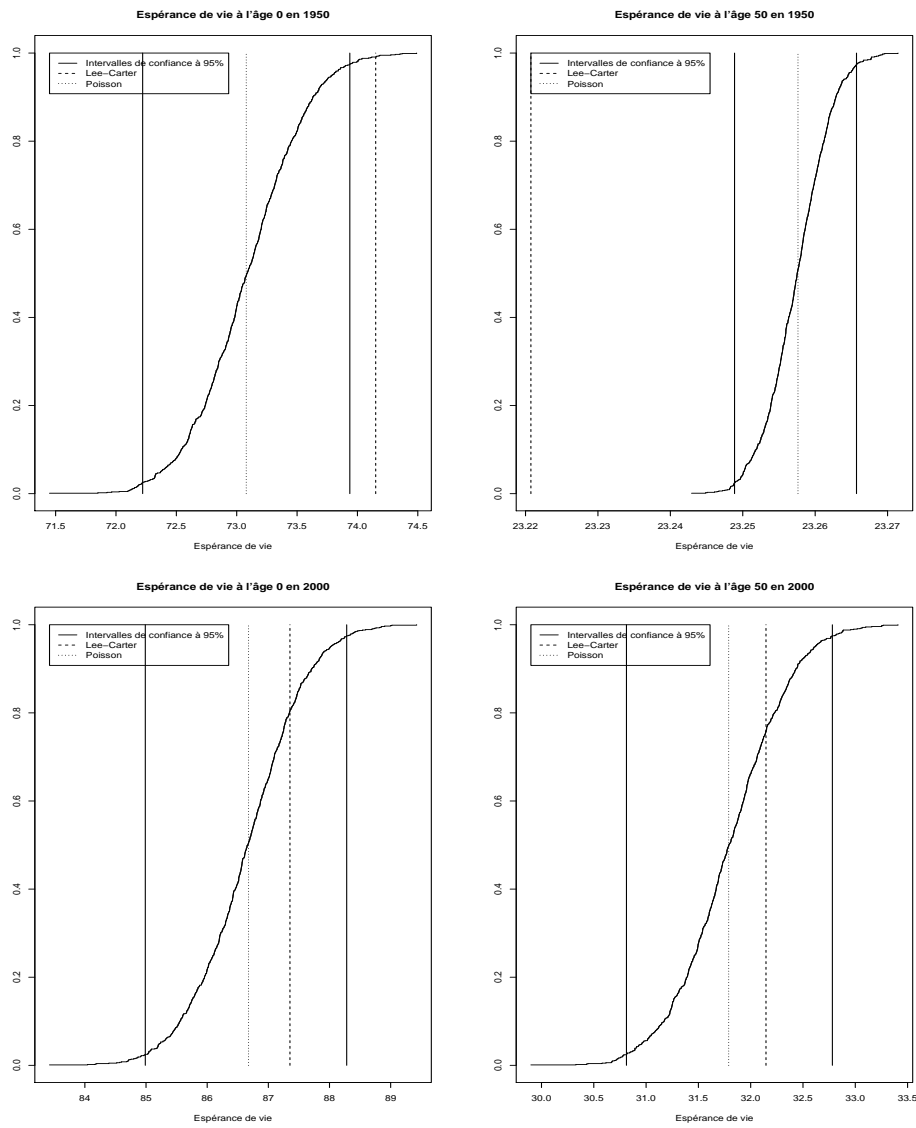


FIG. 4.2 – Fonctions de répartition empiriques des espérances de vie par les simulations Monte Carlo.

## 4.2 Méthodes de rééchantillonnage des décès

L'approche Monte Carlo détaillée plus haut suppose que le modèle (3.30)-(3.32) est correctement spécifié et que la taille de l'échantillon est suffisamment grande que pour pouvoir recourir à la loi asymptotique des estimateurs du maximum de vraisemblance des paramètres. Les méthodes de rééchantillonnage (ou bootstrap) proposées ici retiennent uniquement l'hypothèse de Poisson (3.30), sans pour autant utiliser la structure log-bilinéaire (3.32) des taux instantanés de mortalité.

### 4.2.1 Bootstrap paramétrique

Partant des données observées  $(E_{xt}, D_{xt})$ , nous créons des échantillons bootstrap  $(E_{xt}, D_{xt}^b)$ ,  $b = 1, \dots, B$ , où les  $D_{xt}^b$  sont obtenus par simulation à l'aide de la loi de Poisson de moyenne  $D_{xt}$ .

Pour chaque échantillon bootstrap, il convient de réestimer les paramètres  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$  puis de procéder comme pour la simulation Monte-Carlo.

Les étapes de la méthode du bootstrap paramétrique sont donc

1. Générer de nouveaux nombres de décès suivant la loi

$$D_{xt}^b \sim \text{Poisson}(D_{xt}).$$

2. Estimer les paramètres du modèle (3.30)-(3.32) à partir de chaque échantillon bootstrap  $(E_{xt}, D_{xt}^b)$
3. Estimer les paramètres du modèle ARIMA dont les ordres ont été sélectionnés sur base des observations en utilisant la série  $\mathcal{K}^b = \{\kappa_t^b, t = t_{\min}, \dots, t_{\max}\}$ .
4. Générer les projections  $\kappa_t^b$  en utilisant les paramètres ARIMA estimés à l'étape 3. Les erreurs  $\epsilon_t^m$  sont simulées à partir de la loi normale  $\mathcal{N}(0, \sigma^b)$ .
5. Calculer pour chaque échantillon la valeur de l'indicateur désiré.

La Figure 4.3 présente les logarithmes des taux de mortalité générés par  $B = 1000$  échantillons tandis qu'à la Figure 4.4, nous trouvons les intervalles de confiance à 95% pour les espérances de vie et les comparons avec les estimations du modèle classique de Lee et Carter et de celui de Poisson log-bilinéaire. Remarquons que ces intervalles sont très similaires à ceux obtenus dans le cadre des simulations Monte Carlo.

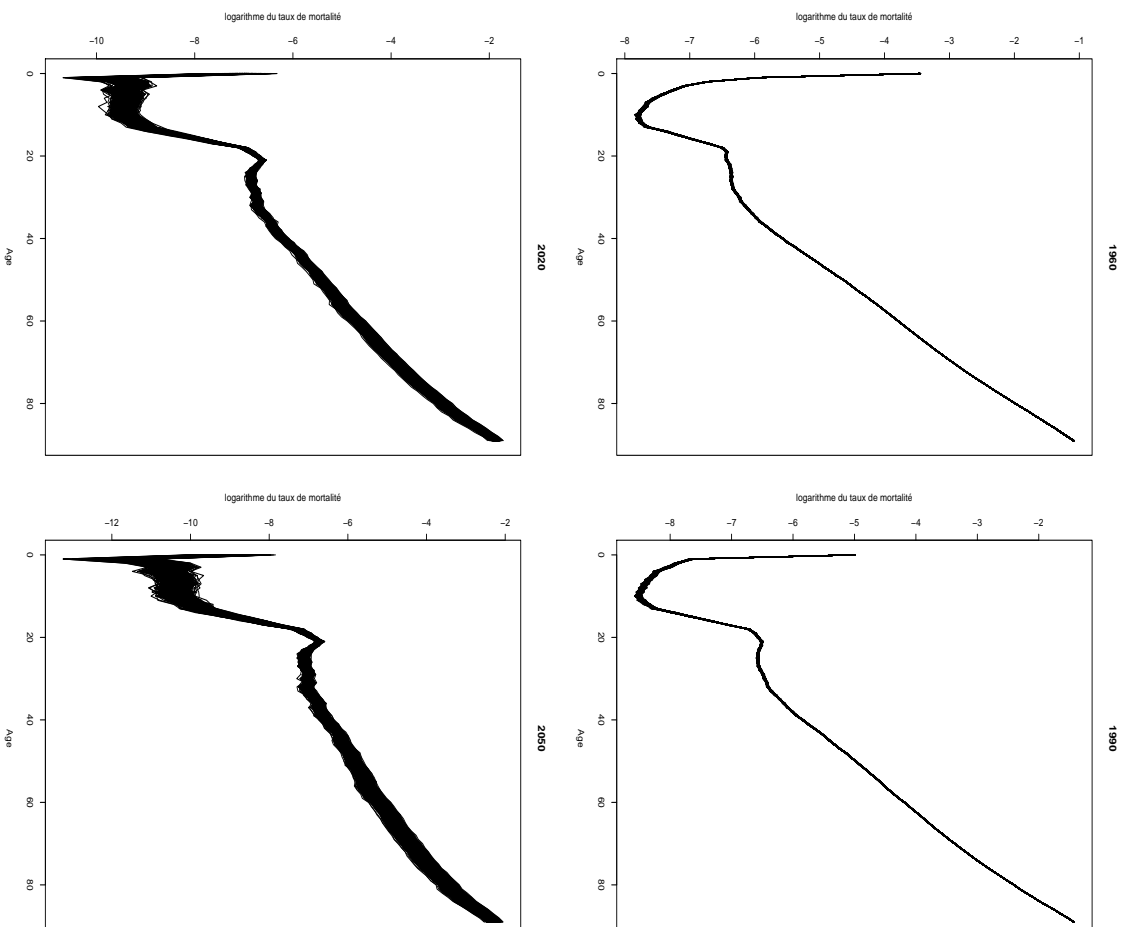


FIG. 4.3 – Prédiction des logarithmes des taux de mortalité pour quelques années par le bootstrap paramétrique.

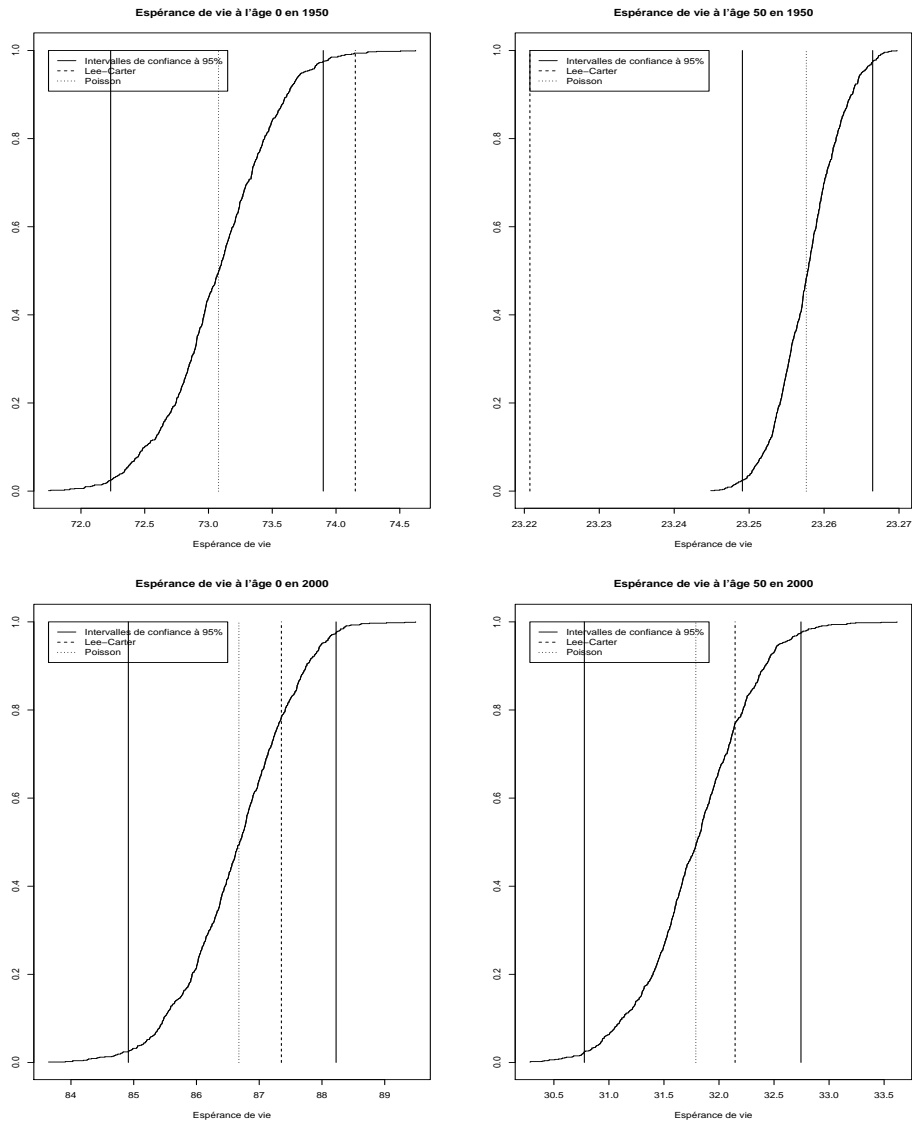


FIG. 4.4 – Fonctions de répartition empiriques des espérances de vie par le bootstrap paramétrique.



### 4.2.2 Bootstrap non paramétrique

Cette troisième approche, plus générale, reconnaît la réalité du problème par le truchement d'un schéma multinomial. Ce dernier redistribue le nombre de décès observé pour chaque pseudo-génération aux différents âges, conformément aux probabilités estimées sur base des statistiques de mortalité disponibles.

Plus précisément, nous suivons pour  $x$  et  $t$  donnés la pseudo-génération  $(E_{xt}, D_{xt})$ ,  $(E_{x+1,t+1}, D_{x+1,t+1})$ ,  $(E_{x+2,t+2}, D_{x+2,t+2})$ , ... Il s'agit en réalité d'une génération fictive, puisqu'elle mêle des individus nés lors des années  $t - x$  et  $t - x - 1$ . Des nouveaux nombres de décès  $D_{xt}^b$ ,  $D_{x+1,t+1}^b$ , ... sont générés pour chacune de ces générations fictives à l'aide de la loi multinomiale d'exposant

$$D_{\bullet} = \sum_{k \geq 0} D_{x+k,t+k}$$

et des paramètres

$$\begin{aligned} \hat{q}_x(t) &= \frac{D_{xt}}{D_{\bullet}} \\ \hat{q}_{x+1}(t+1) &= \frac{D_{x+1,t+1}}{D_{\bullet}} \\ &\vdots \end{aligned}$$

Nous pouvons alors procéder avec les nouveaux échantillons comme décrit aux étapes 2 à 5 du bootstrap paramétrique.

Remarquons que la relation

$$D_{\bullet} = \sum_{k \geq 0} D_{x+k,t+k}^b$$

est satisfaite par les échantillons bootstrap. Autrement dit, le nombre total de décès au sein d'une génération fictive est maintenu par le rééchantillonnage. Cette caractéristique est propre à la méthode du bootstrap non paramétrique. En effet, ni la simulation Monte Carlo, ni le bootstrap paramétrique, ni le bootstrap sur résidus vu à la section suivante ne permettent de maintenir cette propriété.

L'application du bootstrap non paramétrique aux données françaises avec  $B = 1000$  est reprise aux Figures 4.5 et 4.6. Une nouvelle fois, les intervalles de confiance sont très similaires à ceux obtenus à partir des deux premières méthodes.

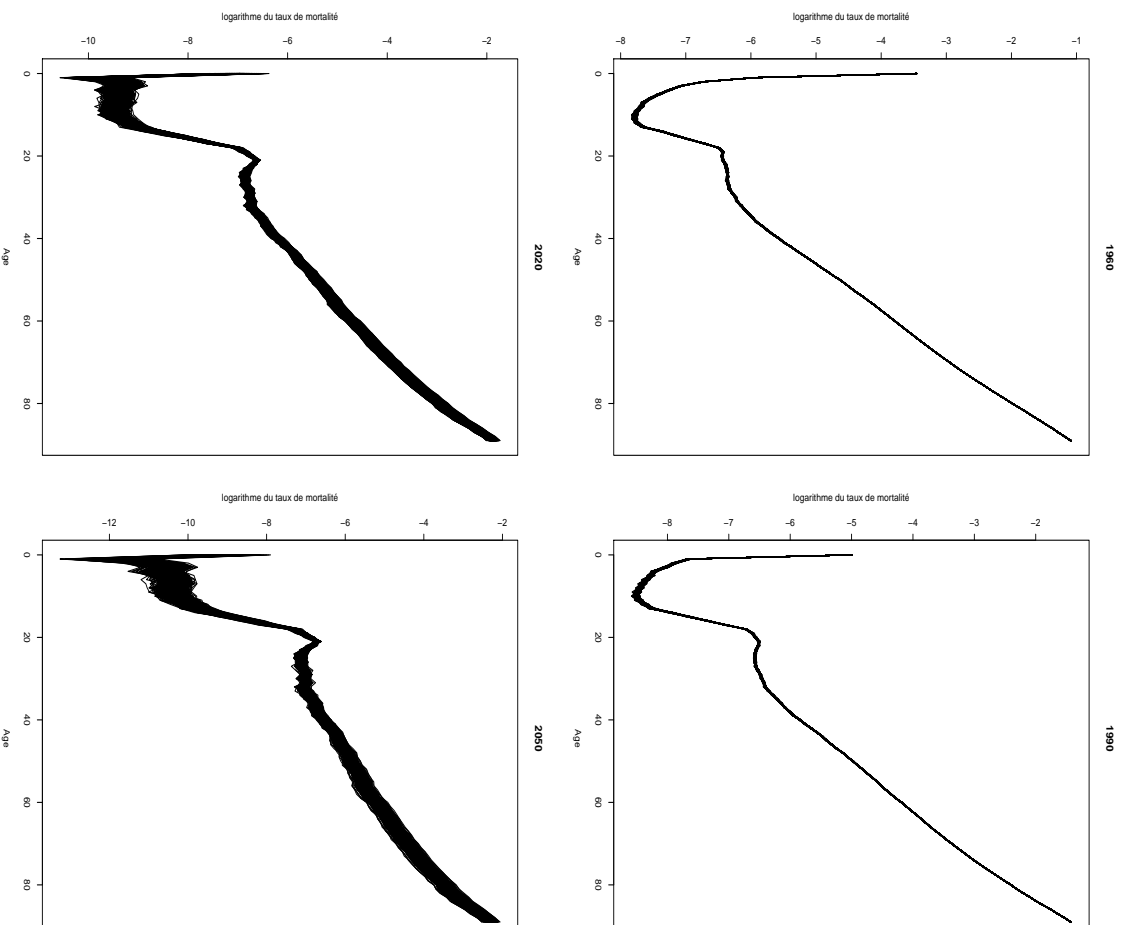


FIG. 4.5 – Prédiction des logarithmes des taux de mortalité pour quelques années par le bootstrap non paramétrique.

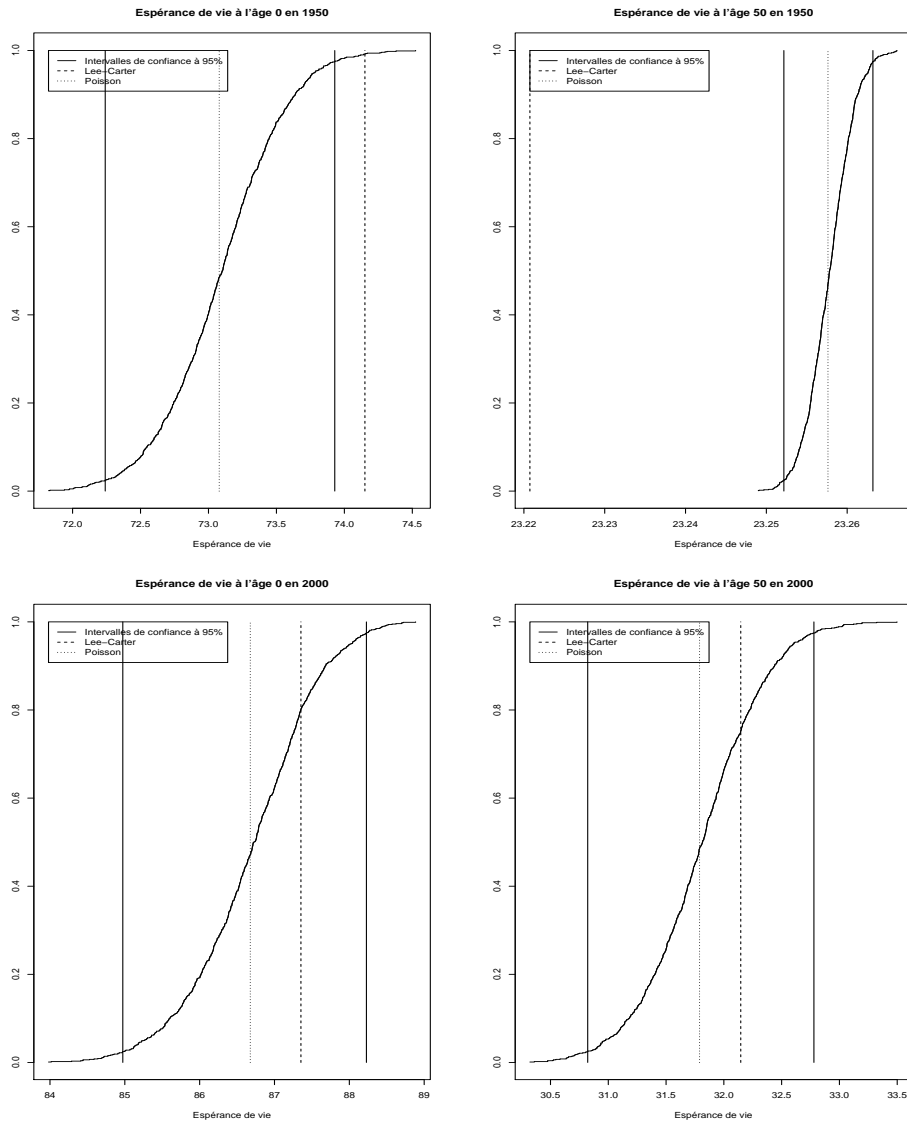


FIG. 4.6 – Fonctions de répartition empiriques des espérances de vie par le bootstrap non paramétrique.

### 4.3 Méthode de rééchantillonnage des résidus

Il est également possible d'utiliser une méthode de rééchantillonnage sur les résidus de Pearson

$$r_{xt}^{Pearson} = \frac{D_{xt} - \widehat{D}_{xt}}{\sqrt{\widehat{D}_{xt}}}$$

si ceux-ci ne souffrent pas de structure apparente et peuvent être considérés indépendants et identiquement distribués.

Définissons les résidus de Pearson centrés

$$\tilde{r}_{xt}^{Pearson} = r_{xt}^{Pearson} - \bar{r}^{Pearson}$$

avec  $\bar{r}^{Pearson}$  la moyenne arithmétique des résidus du modèle. Des échantillons bootstrap  $r_{xt}^{Pearson,b}$ ,  $b = 1, \dots, B$  peuvent alors être générés à partir de la fonction de distribution empirique des résidus centrés  $\tilde{r}_{xt}^{Pearson}$ . En notant  $n$  le nombre d'observations de l'échantillon initial, cette distribution est définie par

$$\mathbb{P}[r_{xt}^{Pearson,b} = \tilde{r}_{x't'}^{Pearson}] = \frac{1}{n}$$

pour tous les couples  $(x,t)$  et  $(x',t')$  de l'échantillon initial.

Des échantillons bootstrap pour les décès  $D_{xt}^b$  découlent alors de ces nouveaux résidus:

$$D_{xt}^b = \widehat{D}_{xt} + \sqrt{\widehat{D}_{xt}} r_{xt}^{Pearson,b}$$

et nous pouvons effectuer les mêmes étapes 2 à 5 que pour les méthodes de rééchantillonnage des décès.

Les Figures 4.7 et 4.8 présentent les résultats du bootstrap sur résidus appliqué à nos données avec  $B = 1000$ . Cette fois, les intervalles de confiance sont beaucoup plus étendus que ceux obtenus par les trois méthodes précédentes. Ceci provient de la structure observée dans les résidus du modèle (pour rappel, la Figure 3.30 illustre les résidus de déviance, très similaires aux résidus de Pearson).

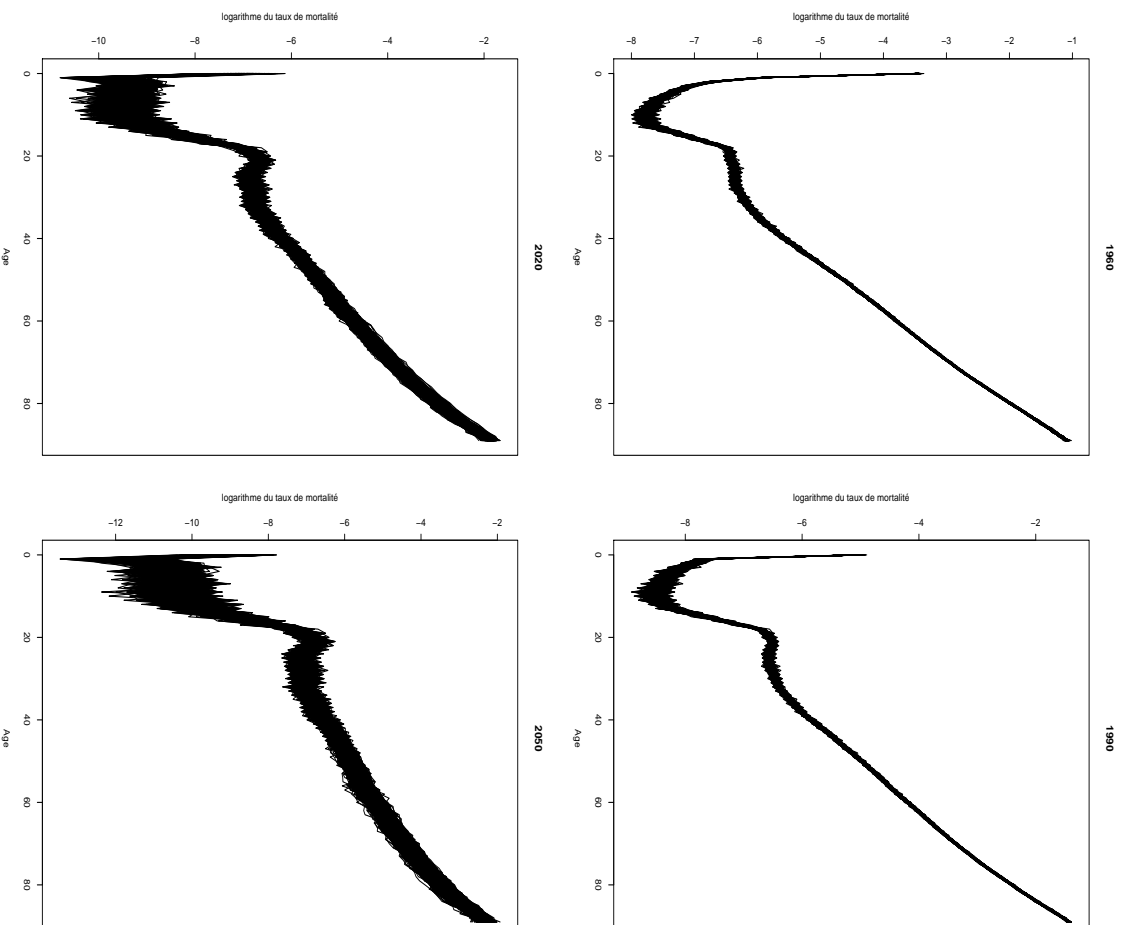


FIG. 4.7 – Prédiction des logarithmes des taux de mortalité pour quelques années par le bootstrap sur résidus.

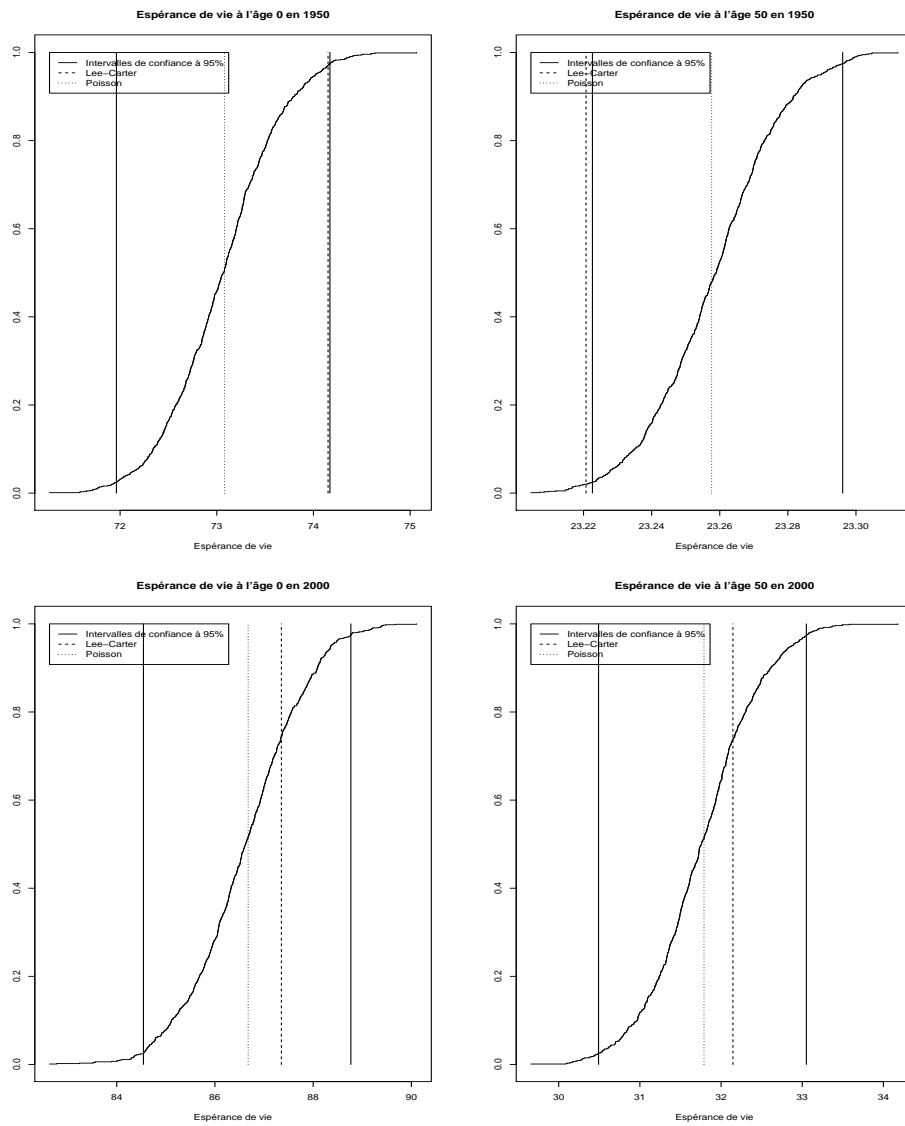


FIG. 4.8 – Fonctions de répartition empiriques des espérances de vie par le bootstrap sur résidus.

# Conclusion

## Modèle binomial

Nous avons exposé dans les sections précédentes comment, à partir des données usuellement mises à disposition par les instituts nationaux de statistique, construire les tableaux nécessaires aux projections de la mortalité. Ensuite, nous avons étudié la méthode de Lee et Carter dont l'objectif principal consiste à établir les tables de mortalité prospectives indispensables à une tarification correcte de produits d'assurance sur la vie, et notamment des rentes viagères. Ce modèle souffre de défauts certains, au nombre desquels il faut souligner l'hypothèse d'homoscédasticité utilisée dans (3.1). Il est donc utile de le raffiner, en proposant par exemple une modélisation des décès plutôt que des taux instantanés de mortalité. Ceci fait l'objet du modèle de Poisson log-bilinéaire, qui reconnaît la variabilité plus grande des taux instantanés de mortalité aux âges élevés.

Comme nous l'avons vu précédemment, le modèle de Poisson peut en fait être vu comme une approximation du modèle naturel

$$D_{xt} \sim \mathcal{B}inomiale(L_{xt}, q_x(t)).$$

En vertu de l'hypothèse de constance par morceaux des taux instantanés de mortalité, cette spécification peut être complétée par

$$q_x(t) = 1 - \exp(-\exp(\alpha_x + \beta_x \kappa_t))$$

et il convient également de lui ajouter des contraintes telles que (3.2)-(3.3) pour en assurer l'identifiabilité.

Sous l'hypothèse d'indépendance des couples  $(x, t)$  de données, la fonction de vraisemblance associée à ce modèle s'écrit

$$\mathcal{V}(\alpha, \beta, \kappa) = \prod_{t=t_{\min}}^{t_{\max}} \prod_{x=x_{\min}}^{x_{\max}} \frac{L_{xt}!}{D_{xt}!(L_{xt} - D_{xt})!} q_x(t)^{D_{xt}} (1 - q_x(t))^{L_{xt} - D_{xt}}$$

et la fonction de log-vraisemblance vaut donc, à un facteur additif près,

$$\begin{aligned} \ln \mathcal{V}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}) &= \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} D_{xt} \ln q_x(t) + (L_{xt} - D_{xt}) \ln(1 - q_x(t)) \\ &= \sum_{t=t_{\min}}^{t_{\max}} \sum_{x=x_{\min}}^{x_{\max}} D_{xt} \ln \left[ \left( 1 - \exp(-\exp(\alpha_x + \beta_x \kappa_t)) \right) \right] \\ &\quad - (L_{xt} - D_{xt}) \exp(\alpha_x + \beta_x \kappa_t) \Big]. \end{aligned}$$

Il est alors possible d'établir un nouvel algorithme itératif pour maximiser cette fonction de log-vraisemblance. De même, chacun des raffinements des Sections 3.3.3, 3.3.7, 3.3.8 et 3.3.9 peut être appliqué.

Ce petit exemple prouve qu'à partir du modèle initial (3.1) de Lee et Carter, il est possible de construire une multitude d'extensions. Celles-ci diffèrent de (3.1) par la variable dépendante, par la méthode d'estimation des paramètres, voire par le nombre ou la spécification de la tendance temporelle. Ainsi, afin de choisir le modèle le plus adéquat, il est primordial d'une part d'analyser en détails les données à utiliser, et d'autre part de garder en tête le but premier de la projection. S'agissant notamment de la tarification de produits d'assurance, il est indispensable de n'utiliser dans le modèle que les âges les plus appropriés aux produits visés. Il est ainsi malvenu d'introduire les adolescents (dont la mortalité est si délicate à modéliser) si l'objectif est de tarifer des rentes viagères.

## Estimation simultanée des paramètres

Un autre point délicat du modèle log-bilinéaire réside dans la non-simultanéité des estimations des paramètres. En général, il est d'abord procédé à l'estimation des  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$ ; le modèle de prédiction ARIMA est ensuite ajusté sur les estimateurs  $\hat{\kappa}_t$ . Il serait sans doute préférable d'effectuer ces ajustements simultanément: après avoir choisi l'ordre du modèle ARIMA, l'intégralité des paramètres serait estimée à l'aide d'un seul et même algorithme. Ainsi, dans le cas de la marche aléatoire avec tendance (3.21),  $\theta$  et  $\sigma_\zeta$  serait estimés en même temps que les  $\alpha_x$ ,  $\beta_x$  et  $\kappa_t$ .

Un tel algorithme est très difficile à imaginer; il est toutefois réalisable à l'aide de techniques de simulation telles que les MCMC (pour Markov Chain Monte Carlo). Cette approche permettrait également d'estimer directement des intervalles de confiance tels que ceux obtenus à l'aide des méthodes du chapitre 4.

Un algorithme MCMC pour le modèle de Poisson log-bilinéaire fait actuellement l'objet de recherches et devrait voir le jour très prochainement.



## Antisélection et fermeture des tables

Rappelons que les tables construites à l'aide des méthodes développées ne peuvent pas être utilisées telles quelles pour rendre compte de la mortalité subie par les assurés d'une compagnie. Il est en effet indispensable de recourir à une modélisation supplémentaire pour considérer le phénomène redouté de l'antisélection. Par celui-ci, les détenteurs d'un contrat d'assurance avec des garanties en cas de vie (notamment les rentiers) auront probablement des statistiques de mortalité plus favorables que le citoyen moyen, allongeant substantiellement leur espérance de vie. Ces caractéristiques doivent évidemment être considérées dans la tarification. Ce phénomène d'antisélection mériterait à coup sûr un examen attentif.

Enfin, l'actuaire en quête de tables prospectives pour tarifier des produits tels que les rentes viagères doit faire face à un dernier type de problème. En général, les tableaux de données ne sont disponibles que jusqu'à des âges relativement peu élevés (95 ou 100 ans). Il est donc nécessaire d'étendre ces tableaux jusqu'aux âges limites de la vie humaine, soit 120 ou 130 ans.

Cette extrapolation peut être réalisée de deux manières. La première possibilité consiste à extrapoler directement les estimateurs  $\hat{\alpha}_x$  et  $\hat{\beta}_x$ . Néanmoins, si la courbe des  $\hat{\alpha}_x$  peut sembler facile à étendre, il en va tout autrement pour celle des  $\hat{\beta}_x$ , et cette technique se révèle être très hasardeuse.

L'alternative consiste à étendre directement les tableaux de données disponibles, donc les taux bruts  $\hat{\mu}_x(t)$ , les quotients  $\hat{q}_x(t)$ , ou les nombres de décès  $D_{xt}$  et effectifs au premier janvier  $L_{xt}$ . Il faut donc modéliser autant de séries qu'il y a d'années d'observation, au contraire de la méthode précédente où seules deux séries devaient être traitées. Cette approche a déjà été utilisée, notamment par COALE & KISKER (1990) et JAUMAIN (2001). Un modèle de projection de la mortalité tel que ceux présentés au chapitre 3 peut alors être appliqué aux tableaux complétés. Néanmoins, si l'extrapolation a été réalisée sur les taux bruts  $\hat{\mu}_x(t)$  ou les quotients  $\hat{q}_x(t)$ , le modèle de Poisson log-bilinéaire ne sera plus accessible car il modélise directement les décès, au contraire du modèle de Lee et Carter qui utilise seulement les taux bruts  $\hat{\mu}_x(t)$ .

Modélisation de l'antisélection et fermeture des tables représentent donc deux étapes non négligeables dans la tarification de produits comme les rentes viagères.

## Références

- ACTUARIAL EDUCATION SERVICE (1995). *Actuarial Mathematics - Subject D2*. The Faculty and Institute of Actuaries, Oxford, Edinburgh.
- ALHO, J.M. (2000). Discussion of Lee (2000). *North American Actuarial Journal* **4**, 91-93.
- AUJOUX, D. & CARBONEL, G. (1996). Mortalité d'expérience et risque financier pour un portefeuille de rentes. *Bulletin Trimestriel de l'Institut des Actuaire Français*, 1-18.
- BELL, W.R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics* **13**, 279-303.
- BENJAMIN, B. & POLLARD, J.H. (1993). *The Analysis of Mortality and Other Actuarial Statistics (3d Edition)*. Heinemann, London.
- BENJAMIN, B. & SOLIMAN, A.S. (1993). *Mortality on the Move*. Institute of Actuaries, Oxford.
- BRILLINGER, D.R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* **42**, 693-734.
- BROCKWELL, P.J. & DAVIS, R.A. (1991). *Time Series: Theory and Methods*. Springer Verlag. New York.
- BROCKWELL, P.J. & DAVIS, R.A. (1996). *Introduction to Time Series and Forecasting*. Springer Verlag. New York.
- BROUHNS, N. & DENUIT, M. (2001a). Risque de longévité et rentes viagères I. Evolution de la mortalité en Belgique de 1880 à nos jours. *Belgian Actuarial Bulletin* **2**.
- BROUHNS, N. & DENUIT, M. (2001b). Risque de longévité et rentes viagères. II. Tables de mortalité prospectives pour la population belge. *Belgian Actuarial Bulletin* **2**.
- BROUHNS, N. & DENUIT, M. (2001c). Risque de longévité et rentes viagères III. Elaboration de tables de mortalité prospectives pour la population assurée belge, et évaluation du coût de l'antisélection. *Belgian Actuarial Bulletin* **2**.
- BROUHNS, N., DENUIT, M. & VAN KEILEGOM, I. (2004). Bootstrapping the Poisson log-bilinear model for mortality forecasting.
- BROUHNS, N., DENUIT, M. & VERMUNT, J. (2002a). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance:*

- Mathematics & Economics* **31**, 373-393.
- BROUHNS, N., DENUIT, M. & VERMUNT, J. (2002b). Measuring the longevity risk in mortality projections. *Bulletin of the Swiss Association of Actuaries*, 105-130.
- BROWN, R.L. (1997). Issues in the modelling of mortality at advanced ages. Institute of Insurance and Pension Research, University of Waterloo, Research Report 97-05.
- CARTER, L. (1996). Forecasting US mortality: A comparison of Box-Jenkins ARIMA and structural time series models. *The Sociological Quarterly* **37**, 127-144.
- CHIANG, C.L. (1984). *The Life Table and Its Applications*. R.E. Krieger Publishing Company, Malabar, Florida.
- CMIR (1990). *Continuous Mortality Investigation Report Number 10*. Institute and Faculty of Actuaries.
- CMIR (1999). *Continuous Mortality Investigation Report Number 17*. Institute and Faculty of Actuaries.
- COALE, A. & KISKER, E.E. (1990). Defects in data on old age mortality in the United States: New procedures for calculating approximately accurate mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum* **4**, 1-31.
- COPPOLA, M., DI LORENZO, E. & SIBILLO, M. (2000). Risk sources in a life annuity portfolio: decomposition and measurement tools. *Journal of Actuarial Practice* **8**, 43-61.
- DELWARDE, A. (2002). *Analyse Prospective de la Charge des Pensions pour l'Etat Belge*. Mémoire présenté en vue de l'obtention du grade d'Ingénieur Civil en Mathématiques Appliquées à l'Université catholique de Louvain, Louvain-la-Neuve.
- DELWARDE, A. & DENUIT, M. (2002). Importance de la période d'observation et des âges considérés dans la projection de la mortalité selon la méthode de Lee-Cardé. *Belgian Actuarial Bulletin* **3**.
- DENUIT, M. (1999). Bases techniques de l'assurance-vie individuelle en Belgique. *Revue Générale des Assurances et des Responsabilités* **72(4)**, 13078/1 - 13078/11.
- DENUIT, M. (2002). *Assurances de Personnes*. Cours professé à l'Institut des sciences actuarielles de l'Université catholique de Louvain.
- DENUIT, M. & CORNET, A. (1999). Premium calculation with dependent time-until-death random variables: the widow's pension. *Journal of Actuarial Practice* **7**, 147-180.
- DENUIT, M., DHAENE, J., LE BAILLY DE TILLEGHEM, C. & TEGHEM, S. (2001). Measuring the impact of a dependence among insured lifetimes. *Belgian Actuarial Bulletin* **1**, 18-39.
- EILERS, P.H.C. & MARX, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Sciences* **11**, 89-121.

- GIROSI F. & KING G. (2003). Demographic Forecasting. Draft book manuscript, version 11/3/2003.
- GOCKEL, S. (2000). *Elaboration d'une Méthodologie de Construction de Tables de Mortalité Prospectives Belges*. Mémoire présenté en vue de l'obtention du DES en sciences actuarielles à l'Université catholique de Louvain, Louvain-la-Neuve.
- GOMPERTZ, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical transactions* **27**, 513-585.
- GOODMAN, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association* **74**, 537-552.
- GUTTERMAN, S., & VANDERHOOF, I.T. (1999). Forecasting changes in mortality: a search for a law of causes and effects. *North American Actuarial Journal* **2**, 135-138.
- HELIGMAN, L. & POLARD, J.H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries* **107**, 49-80.
- HÜRLIMANN, W. (1990). On life table applications of ordering risks. *Insurance: Mathematics & Economics* **9**, 277-279.
- INSTITUT NATIONAL DE STATISTIQUE, BUREAU FÉDÉRAL DU PLAN & COMMUNAUTÉ SCIENTIFIQUE (1996). *Perspectives de Population 1995-2050*. Institut National de Statistique, Ministère des Affaires Economiques, Bruxelles.
- JAUMAIN, C. (2001). Tables de mortalité en Belgique: questions d'actualité. *Actu-L* **1**, 67-108.
- JAUMAIN, C. (2002). *La longévité humaine en Belgique. Evolution au cours du 20e siècle. Prospective. Assurance du risque de longévité*. Edité par l'auteur.
- KANNISTO, V. (1988). On the survival of centenarians and the span of life. *Population Studies* **42**, 389-406.
- KEYFITZ, N. (1985). *Applied Mathematical Demography*. Springer Verlag, New York.
- KINSELLA, K. & GIST, Y.J. (1998). Gender and aging: Mortality and Health. *International Brief* **98-02** US Bureau of the Census.
- KLING, B.M. (1993). *Life Insurance: a Non-Life Approach*. Thesis Publishers, Amsterdam.
- KLING, B.M. & WOLTHUIS, H. (1992). Ordering of risks in life insurance. *Insurance: Mathematics & Economics* **11**, 139-152.
- LAMBRECHT, M. (1997). Le vieillissement démographique: de l'analyse des évolutions de population au risque de conclusions hâtives. *Planning Paper* **81**, Bureau Fédéral du Plan, Bruxelles.
- LAMBRECHTS, K. (1996). *Belgische Sterftetafels: Van Vorige to Volgende Eeuw*. Eindwerk aangeboden to het behalen van de graad van gediplomeerde in de actuariële wetenschappen, KULeuven.

- LEE, R.D. (2000). The Lee-Carter method of forecasting mortality, with various extensions and applications. *North American Actuarial Journal* **4**, 80-93.
- LEE, R.D. & CARTER, L. (1992). Modelling and forecasting the time series of US mortality. *Journal of the American Statistical Association* **87**, 659-671.
- LEE, R.D., & MILLER, T. (2000). Assessing the performance of the Lee-Carter approach to modelling and forecasting mortality.  
See [www.demog.berkeley.edu/~rlee](http://www.demog.berkeley.edu/~rlee)
- LEE, R.D. & NAULT, F. (1993). Modeling and forecasting provincial mortality in Canada. Paper presented at the World Congress of International Union for the Scientific Study of Population, Montréal.
- LEE, R.D. & ROFMAN, R. (1994). Modelacion y proyeccion de la mortalidad in Chile. *Natas* **22(59)**, 182-313. (available in English from the authors, titled "Modeling and forecasting mortality in Chile")
- LERIDON, H. & TOULEMON, L. (1997). *Démographie*. Economica, Paris.
- LOUIS, A. (2001). *Etablissement de Tables de Mortalité Prospectives*. Mémoire présenté en vue de l'obtention du grade de Licencié en Sciences Mathématiques à l'Université catholique de Louvain, Louvain-la-Neuve.
- MAROCCHO, P. & PITACCO, E. (1998). Longevity risk and life annuity reinsurance. *Transactions of the 26th International Congress of Actuaries*, Birmingham, 453-479.
- MCDONALD, A.S. (1997) Editor. *The Second Actuarial Study of Mortality in Europe*. Groupe Consultatif des Associations d'Actuaires des Pays des Communautés Européennes, Oxford.
- MCDONALD, A.S., CAIRNS, A.J.C., GWILT, P.L. & MILLER, K.A. (1998). An international comparison of recent trends in mortality. *British Actuarial Journal* **4**, 3-141.
- MCNOWN, R. & ROGERS, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography* **26**, 645-660.
- MCNOWN, R. & ROGERS, A. (1992). Forecasting cause-specific mortality using time series method. *International Journal of Forecasting* **8**, 413-432.
- MENTHONNEX, J. & WANNER, PH. (1998). *Tables de Mortalité Longitudinales pour la Suisse (Génération 1880-1980)*. Office Fédéral de la Statistique, Berne.
- MESLÉ, F. MENTHONNEX, J. & WANNER, PH. (1998). *Tables de Mortalité Longitudinales pour la Suisse (Génération 1880-1980)*. Office Fédéral de la Statistique, Berne.
- MESLÉ, F. & VALLIN, J. (2002). Comment améliorer la précision des tables de mortalité aux grands âges? Le cas de la France. *Population* **4-5**, 603-631.
- OLIVIERI, A. (2001). Uncertainty in mortality projections: an actuarial perspective. *Insurance: Mathematics & Economics* **29**, 231-245.

- OLIVIERI, A. & PITACCO, E. (1999). Funding sickness benefits for the elderly. *Proceedings of the 30th ASTIN Colloquium*, Tokyo, 135-155.
- OLIVIERI, A. & PITACCO, E. (2000). Solvency requirements for life annuities. *Proceedings of the AFIR 2000 Colloquium*, Tromsø, Norway, 547-571.
- OLIVIERI, A. & PITACCO, E. (2001). Facing LTC risks. *Paper presented at the 5th International Congress on Insurance: Mathematics & Economics*, Penn State, July 2001.
- PITACCO, E. (2001). Assessing and facing the longevity risk: Financial requirements. *Proceedings of the 4th Italian-Spanish Conference on Financial Mathematics*, Alghero
- POH, S.B. & BROWN, R.L. (1995). Methods of projection of fertility and mortality rates. Institute of Insurance and Pension Research, University of Waterloo, Research Report 95-13.
- RENSHAW, A.E. & HABERMAN, S. (1996). The modelling of recent mortality trends in the United Kingdom male assured lives. *British Actuarial Journal* **2**, 449-477.
- RENSHAW, A.E. & HABERMAN, S. (2001). On the forecasting of mortality reduction factors. Actuarial Research Paper **135**, Department of Actuarial Science and Statistics, City University, London.
- RENSHAW, A.E. & HABERMAN, S. (2003a). Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projection. *Applied Statistics* **52**, 119-137.
- RENSHAW, A.E. & HABERMAN, S. (2003b). Lee-Carter mortality forecasting with age specific enhancement. *Insurance: Mathematics & Economics* **33**, 255-272.
- SITHOLE, T.Z., HABERMAN, S. & VERRALL, R.J. (2000). An investigation into parametric models for mortality projections, with applications to immediate annuitants and life office pensioners' data. *Insurance: Mathematics & Economics* **27**, 285-312.
- VALLIN, J. (1973). *La Mortalité par Génération en France depuis 1899*. INED, Paris.
- VANDESCHRIK, CH. (1995). *Analyse Démographique*. Academia Bruylant & L'Harmattan, Louvain-la-Neuve & Paris.
- VAUPEL, J.W. (1997). Trajectory of mortality at advanced ages. In *Between Zeus and the Salmon: The Biodemography of Longevity*, pp. 17-37, National Academy of Science.
- VINCENT, P. (1951). La mortalité des vieillards. *Population* **6** 181-204.
- WANG, S.S. & BROWN, R.L. (1998). A frailty model for projection of human mortality improvements. *Journal of Actuarial Practice* **6**, 221-241.
- WILMOTH, J.R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical Report, Department of Demography, University of California, Berkeley.

- WILMOTH, J.R. (1996). Mortality projections for Japan: A comparison of four methods. In *Health and Mortality among Elderly Populations*, Caselli, G., and Lopez, A. editors, pp. 266-287. Oxford University Press, New York.
- WILMOTH, J.R. (1997). In search of limits. In *Between Zeus and the Salmon: The Biodemography of Longevity*, pp. 38-64, National Academy of Science.
- WILMOTH, J.R., DEEGAN, L.J., LUNDSTRÖM, H. & HORIUCHI, S. (2000). Increase of maximum life-span in Sweden, 1861-1999. *Science* **289**, 2366-2369.
- YAARI, M.E. (1965). Uncertain lifetime, life insurance and the theory of the consumer. *Review of Economic Studies* **32**.