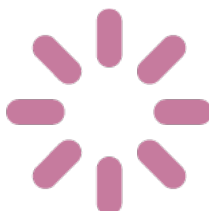


EURIA  
Euro-Institut d'Actuariat



## Rapport de Bureau d'Étude

---

**Objectif** : Élaboration d'un MOOC sur la construction d'une table de mortalité d'expérience

---

LAVENANT David

LEJEUNE Corentin

NEDELEC Pierre

Département des Sciences et Techniques  
Université de Bretagne Occidentale - EURIA  
6 Avenue Le Gorgeu - France  
Mai 2017

# Table des matières

<b>I</b>	<b>Gestion de projet</b>	<b>5</b>
<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Déroulement du projet</b>	<b>9</b>
2.1	Agenda de l'élaboration du MOOC . . . . .	9
2.2	Répartition du travail et gestion du projet . . . . .	10
2.2.1	Site Trello . . . . .	10
2.2.2	VSDC . . . . .	11
2.2.3	Création de QCM . . . . .	13
2.3	Problèmes rencontrés . . . . .	13
<b>3</b>	<b>Conclusion</b>	<b>14</b>
<b>II</b>	<b>Cours écrit du MOOC</b>	<b>15</b>
<b>4</b>	<b>Introduction</b>	<b>17</b>
4.1	Historique et présentation des tables de mortalité de référence (TH-0002 et TF-0002) . . . . .	17
4.2	Notions de base des modèles de durée . . . . .	19
4.3	Premiers développements de modèles pour les taux de mortalité . . . . .	24

<b>5</b>	<b>Description des données du problème</b>	<b>26</b>
5.1	Choix de nos données . . . . .	26
5.2	Calculs d'exposition et calcul des taux de mortalité .	27
<b>6</b>	<b>Méthodes d'ajustement</b>	<b>31</b>
6.1	Modèles paramétriques . . . . .	31
6.1.1	Gompertz-Makeham . . . . .	31
6.1.2	Lee-Carter . . . . .	32
6.2	Modèles semi-paramétriques . . . . .	32
6.2.1	Cox . . . . .	32
6.2.2	Brass . . . . .	34
6.3	Modèle non-paramétrique . . . . .	34
6.3.1	Kaplan-Meier . . . . .	34
<b>7</b>	<b>Le modèle de Lee-Carter</b>	<b>37</b>
7.1	Présentation et mise en place du modèle . . . . .	37
7.2	Projection du modèle . . . . .	42
7.3	Analyse des résidus et limites du modèle . . . . .	48
<b>8</b>	<b>Conclusion</b>	<b>51</b>
<b>A</b>	<b>Code R du chapitre 5</b>	<b>55</b>
<b>B</b>	<b>Code R du chapitre 7</b>	<b>60</b>

# Remerciements

Nous tenions à remercier les principaux acteurs nous ayant conseillés tout au long de notre projet.

Nous remercions Jérémy Gorris (actuaire chez Axa) et Guillaume Metge (actuaire chez Allianz) pour leur implication, leur énergie et leur disponibilité. En effet, ils ont toujours été disponibles lorsque nous avons des questions ou des doutes à propos du développement de notre projet. Nous remercions également Dominique Abgrall, actuaire chez Allianz, qui nous a fourni deux bases de données simulées sans lesquelles un des chapitres de notre cours n'aurait pas pu être développé correctement.

Nous tenons également à remercier les deux professeurs nous ayant encadrés tout au long de notre projet, Franck Vermet, directeur des études, et Anthony Nahelou, actuaire chez Sterenn Actuariat, qui nous ont conseillés sur la mise en forme de nos cours.

**Première partie**  
**Gestion de projet**



# Chapitre 1

## Introduction

Notre Bureau d'Etude a consisté à construire un MOOC sur l'élaboration d'une table de mortalité d'expérience. Le terme "MOOC" est un sigle anglophone signifiant "Massive Open Online Courses". On peut le traduire en français par "Formation en Ligne Massive Ouverte à Tous". Ce projet a été proposé par Jérémie Goris et Guillaume Metge le 12 septembre 2016.

Guillaume et Jérémie nous ont d'abord proposé d'étudier la bibliographie propre aux tables de mortalité. Cette bibliographie devait porter sur des méthodes que nous étions en mesure d'implémenter et de présenter lors de nos cours.

Le projet a commencé par la compréhension et la visualisation des données brutes, ressemblant le plus possible à des données trouvées dans des bases de compagnie d'assurance. Nous devons ensuite traiter ces données afin de calculer des taux de mortalité bruts. Nous devons ensuite les lisser en proposant des modèles paramétriques, semi paramétriques ou non paramétriques. Nous avons choisi d'implémenter la méthode Lee-Carter qui est un modèle paramétrique. Pour finir, nous verrons les limites de ce modèle.

Nous parlerons dans une première partie du déroulement de notre projet, c'est-à-dire l'agenda que nous avons suivi, la manière dont nous nous sommes répartis le travail et les problèmes que nous avons rencontrés. Nous finirons par une conclusion de notre tra-

vail et discuterons des apports de ce projet.  
Dans une deuxième partie, nous vous présenterons la version écrite de notre MOOC.



## Chapitre 2

# Déroulement du projet

### 2.1 Agenda de l'élaboration du MOOC

Nous avons travaillé pendant le premier semestre, de Septembre à Janvier, sur le fond de notre MOOC, le temps de découvrir notre sujet et de nous en imprégner. Dans ce but, nous avons étudié les cours de Frédéric Planchet, professeur à l'Institut de Science Financière et d'Assurance (ISFA) et Jean-David Fermanian, professeur à l'Ecole Nationale de la Statistique et de l'Administration (ENSAE). Ces cours sont disponibles sur internet, les liens étant donnés en fin de rapport, dans la section "Bibliographie".

De Janvier à Février, nous avons examiné la manière dont nous allions présenter nos cours. En effet, nous avons dû mener une réflexion sur le nombre de vidéos que le MOOC devait contenir ainsi que sur leur durée.

Nous avons également identifié la cible de ce MOOC, car ce n'était pas spécifié dans la feuille de route. Ces cours sont destinés aux élèves de Master 2 de notre école d'Actuariat. Dès lors, nous avons considéré qu'il n'était pas nécessaire de revenir sur les acquis des années précédentes. En Janvier, nous avons également codé la méthode de Lee-Carter. Pour ce faire, nous avons besoin de données. Nous avons récupéré des données extraites des tables de mortalité TH0002 et TF0002 par Dominique Abgrall, ac-

tuaire chez Allianz Vie. Nous avons choisi d'utiliser ces données car leur présentation est sensiblement la même que dans une compagnie d'assurance.

Enfin, nous avons décidé de ne pas présenter l'ensemble des modèles possibles pour la modélisation des taux de mortalité mais seulement quelques-uns, ciblés par "famille" : modèles paramétriques, semi paramétriques et non paramétriques.

Le but de notre cours est d'apprendre à construire une table de mortalité en pratique en utilisant un modèle de durée.

Le plan du MOOC terminé (il a été modifié de manière mineure jusqu'à la fin du projet), nous avons pu commencer l'enregistrement des vidéos.

## **2.2 Répartition du travail et gestion du projet**

### **2.2.1 Site Trello**

Afin d'atteindre le but commun que nous nous sommes fixés, c'est-à-dire la production d'un MOOC, nous nous sommes appuyés sur un outil informatique permettant une meilleure organisation du travail collectif et collaboratif : TRELLO.

Coopératif car chaque individu du groupe a effectué des tâches individuelles participant à la construction commune du MOOC.

Collaboratif car plusieurs travaux ont nécessité d'agir en commun, dans le but d'une amélioration continue de l'ensemble du projet.

Pour ce faire, Trello nous a été d'une très grande utilité. Il s'agit d'une application de gestion de projet en ligne simple et efficace.

L'application se présente sous la forme d'un tableau de bord permettant de coordonner le travail d'équipe et de mettre en partage les différents fichiers du projet. Il est organisé en colonnes, comportant des cartes à l'intérieur desquelles se trouvent les fichiers et documents correspondants à l'intitulé.

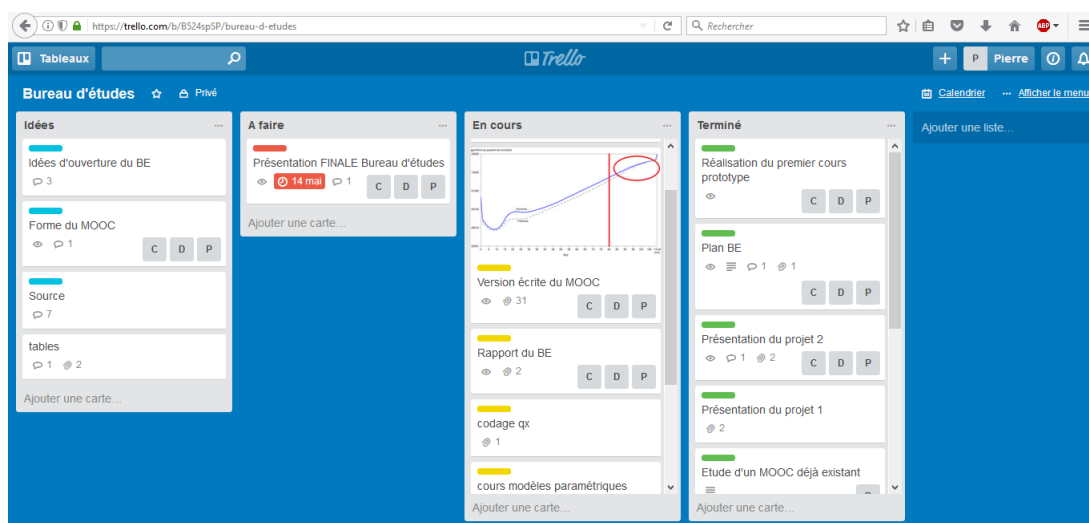


FIGURE 2.1 – Vision globale du projet sur Trello

Ce tableau répartit la liste des tâches sous forme de colonnes "idées" (colonne bleue), "à faire" (colonne rouge), "en cours" (colonne jaune) et "fait" (colonne verte). Ces tâches sont assignées aux différents membres du projet. Au fil de leur exécution, il suffit de glisser-déposer les cartes correspondantes d'une colonne à l'autre. Nous avons aussi utilisé des check-lists, des dates limites et des notifications afin de baliser clairement chaque étape du projet.

Cet outil est avantageux car il nous permet de connaître l'état d'avancement du projet d'un seul coup d'œil, de voir quelles modifications ont été apportées et par quel membre du groupe ainsi que de stocker et de mettre en commun les différents fichiers et documents que nous produisons ou dont nous avons besoin.

### 2.2.2 VSDC

Pour la production du contenu vidéo du MOOC, nous avons utilisé le logiciel gratuit VSDC.

Nous avons fait ce choix car une fois la phase d'apprentissage terminée, VSDC est simple d'utilisation. Il fonctionne sur le principe

du « mille feuilles ». La zone de montage est constituée d'autant de pistes que nécessaires. Sur chaque piste, on avance en concaténant image après image ou alors seconde après seconde. De prime abord, le fait d'ajouter un certain nombre de pistes peut sembler complexe et lourd à gérer. Cependant, l'utilisation assez intuitive des commandes nous facilite la tâche.

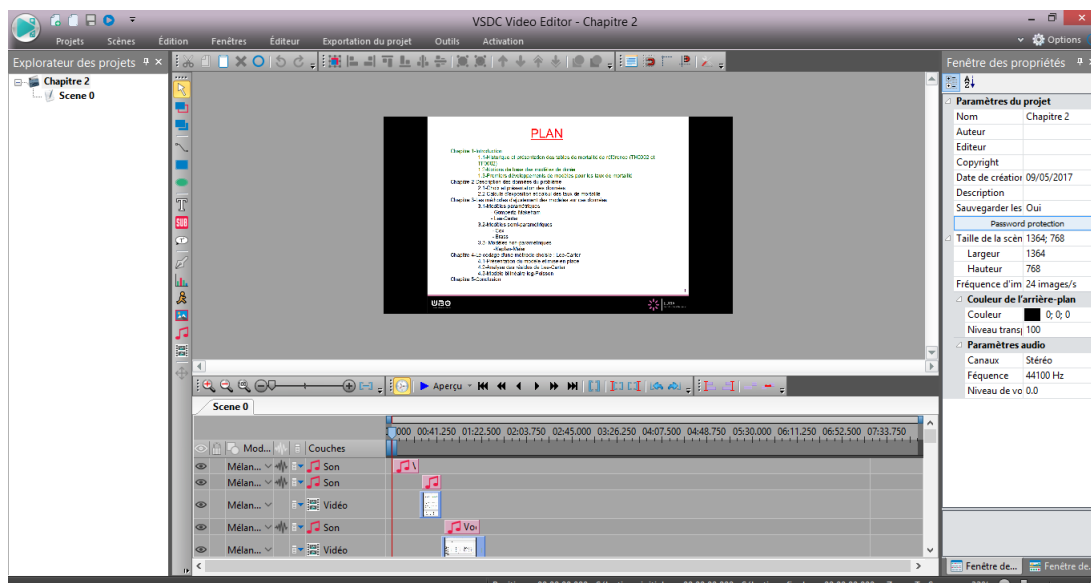


FIGURE 2.2 – Création d'une vidéo avec VSDC

Nous avons surtout gardé à l'esprit que notre projet ne sera pas figé dans le temps et que nous (ou d'autres personnes souhaitant travailler dessus) serons sûrement amenés à le faire évoluer et à l'améliorer. Nous avons bien conscience que le MOOC produit peut être amélioré sur bien des points, sur la forme comme sur le fond. L'aspect « mille feuilles » sera justement très appréciable car chacun des éléments (pistes audio, slides, etc.) constituant les vidéos pourra être modifié, supprimé ou remplacé, et que d'autres pourront être rajoutés. Nous nous inscrivons dès lors dans une démarche d'amélioration continue de notre projet et espérons pour l'avenir en présenter une version qui nous satisferait entièrement.

### 2.2.3 Création de QCM

En complément de nos cours vidéos, nous avons préparé des QCM sur les différents sujets afin de vérifier l'apprentissage du contenu des vidéos. Nous avons réalisé ces QCM sur le site quizyourself.com.

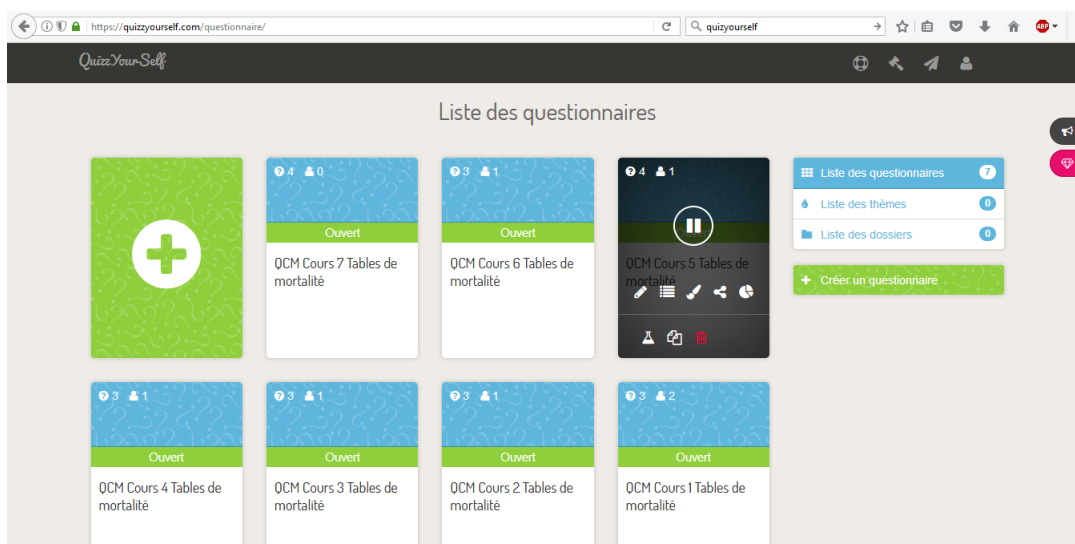


FIGURE 2.3 – Création d'un qcm sur quizyourself

## 2.3 Problèmes rencontrés

Aucun de nous trois n'ayant réalisé de montage auparavant, nous avons dû apprendre à maîtriser rapidement ce logiciel. Cependant, même en le maîtrisant, le montage des vidéos prend beaucoup de temps.

De plus, réaliser des enregistrements audios clairs et de bonne qualité est un tâche chronophage et qui nécessite un matériel adapté. Nous avons donc utilisé un casque avec microphone afin d'optimiser le rendu final.

## Chapitre 3

### Conclusion

De part son originalité, il s'agit d'un projet que nous avons trouvé très intéressant et formateur. En effet, un actuaire doit pouvoir transmettre ses connaissances à des membres provenant de toutes les branches de l'entreprise. C'est pourquoi réaliser un MOOC nous a non seulement permis d'étudier en détail la construction de tables de mortalité et ainsi en connaître davantage sur le domaine de l'assurance-vie, mais également de présenter nos études. Nous devons prochainement être en mesure de savoir expliquer de manière claire et concise des études parfois complexes à un public non initié.

**Deuxième partie**

**Cours écrit du MOOC**





## Chapitre 4

# Introduction

Nous réalisons ce MOOC dans le cadre du bureau d'étude de Master 1 à l'EURIA.

L'objectif de ce cours est de comprendre la construction de fichiers incontournables en assurance-vie : les tables de mortalité.

### **4.1 Historique et présentation des tables de mortalité de référence (TH-0002 et TF-0002)**

On peut considérer que les premiers travaux concernant l'étude de la mortalité des humains ont été effectués par GRAUNT au 17<sup>e</sup> siècle afin de modéliser la propagation des épidémies de peste noire qui ont ravagé des villes comme Londres à cette époque. Cependant, mis à part ce pionnier et quelques autres comme PETTY ou DE WITT, c'est au 18<sup>e</sup> siècle que l'on connaît un véritable essor de l'étude approfondie des tables de mortalité, avec MALTHUS, KERSEBOOM ou encore DEPARCIEUX.

On se rend compte à cette époque que la démographie joue un rôle prépondérant dans des domaines aussi variés que les domaines sanitaires, militaires ou encore théologiques. Malgré leurs efforts, ces précurseurs de l'étude des tables de mortalité se sont confrontés à des problèmes tels que le manque de données ou les difficultés d'un langage technique encore balbutiant en la matière.

En ce sens, les grands traités de démographie ratifiés à la fin du 18e siècle joueront un rôle fondamental car ils synthétiseront et rassembleront tous les savoirs démographiques de l'époque.

Nous avons ci-dessous l'exemple de la table TH 0002 avec les  $L_x$  représentant le nombre de survivants à un âge donné (compris entre 0 et 112ans). Comme on peut le voir, la population initiale est constituée de 100 000 individus. Il reste par exemple 99 274 individus âgés de 20 ans. Les  $q_x$  sont les taux de mortalité pour chaque âge. Par exemple, la probabilité de mourir entre 20 et 21 ans est de 0,0035. Ces tables réglementaires françaises (TH 002 pour les hommes, TF0002 pour les femmes) sont utilisées en cas de vie et non pour les calculs de rentes viagères.

Age	$L_x$	$q_x$
0	100 000	0.00384
1	99 616	0.00033
2	99 583	0.00021
3	99 562	0.00017
4	99 545	0.00014
5	99 531	0.00012
6	99 519	0.00011
7	99 508	0.00010
8	99 498	0.00010
9	99 488	0.00010
10	99 478	0.00011
11	99 467	0.00011
12	99 456	0.00012
13	99 444	0.00013
14	99 431	0.00016
15	99 415	0.00020
16	99 395	0.00024
17	99 371	0.00029
18	99 342	0.00033
19	99 309	0.00035
20	99 274	0.00035
21	99 239	0.00034
22	99 205	0.00034
23	99 171	0.00034
24	99 137	0.00034
25	99 103	0.00035
26	99 068	0.00035
27	99 033	0.00036
28	98 997	0.00037
29	98 960	0.00039
30	98 921	0.00042
31	98 879	0.00047
32	98 833	0.00052
33	98 782	0.00058
34	98 725	0.00064
35	98 662	0.00070
36	98 593	0.00076
37	98 518	0.00084
38	98 435	0.00093
39	98 343	0.00103
40	98 242	0.00114
41	98 130	0.00125
42	98 007	0.00138
43	97 872	0.00151
44	97 724	0.00165
45	97 563	0.00180
46	97 387	0.00195
47	97 197	0.00210
48	96 993	0.00224
49	96 776	0.00238
50	96 546	0.00251
51	96 304	0.00265
52	96 049	0.00282
53	95 778	0.00302
54	95 489	0.00324
55	95 180	0.00346
56	94 851	0.00369
57	94 501	0.00392
58	94 131	0.00414
59	93 741	0.00440
60	93 329	0.00468
61	92 892	0.00503
62	92 425	0.00543
63	91 923	0.00589
64	91 382	0.00640
65	90 797	0.00697
66	90 164	0.00763
67	89 476	0.00838
68	88 726	0.00923
69	87 907	0.01020
70	87 010	0.01133
71	86 024	0.01259
72	84 941	0.01401
73	83 751	0.01563
74	82 442	0.01752
75	80 998	0.01970
76	79 402	0.02228
77	77 633	0.02527
78	75 671	0.02874
79	73 496	0.03276
80	71 088	0.03749
81	68 423	0.04304
82	65 478	0.04956
83	62 233	0.05709
84	58 680	0.06564
85	54 828	0.07518
86	50 706	0.08567
87	46 362	0.09693
88	41 868	0.10865
89	37 319	0.12053
90	32 821	0.13260
91	28 469	0.14546
92	24 328	0.15965
93	20 444	0.17531
94	16 860	0.19229
95	13 618	0.21060
96	10 750	0.23005
97	8 277	0.25045
98	6 204	0.27208
99	4 516	0.29473
100	3 185	0.31837
101	2 171	0.34316
102	1 426	0.36886
103	900	0.39556
104	544	0.42279
105	314	0.45223
106	172	0.48256
107	89	0.50562
108	44	0.54545
109	20	0.55000
110	9	0.55556
111	4	0.75000
112	1	1.00000

FIGURE 4.1 – Table TH0002

La table CIMA est utilisée dans bon nombre de pays d'Afrique et fait suite à une initiative de 2009 visant à établir de nouvelles tables, les dernières étant trop anciennes. La lecture peut en être faite de manière similaire avec une colonne  $d_x$  en plus, représen-

tant le nombre de morts d'un âge  $n$  à un âge  $n+1$ . Par exemple, entre 0 et 1 an, 5 368 personnes sont décédées donc le  $q_x$  est de  $5\,368 / 1\,000\,000$ .

Age	Lx (CIMA H)	qx (CIMA H)	dx (CIMA H)	Age	Lx (CIMA H)	qx (CIMA H)	dx (CIMA H)	Age	Lx (CIMA H)	qx (CIMA H)	dx (CIMA H)	Age	Lx (CIMA H)	qx (CIMA H)	dx (CIMA H)
0	1 000 000	0,5368%	5 368	30	966 201	0,1993%	1 926	60	815 120	1,3526%	11 025	89	183 727	13,2356%	24 317
1	994 632	0,0730%	726	31	964 354	0,2037%	1 964	61	804 094	1,4816%	11 914	90	159 409	14,2255%	22 677
2	993 906	0,0659%	555	32	962 390	0,2108%	2 028	62	792 181	1,6146%	12 791	91	136 733	16,0604%	21 960
3	993 351	0,0476%	473	33	960 362	0,2205%	2 116	63	779 390	1,7517%	13 653	92	114 773	19,1167%	21 943
4	992 878	0,0407%	404	34	958 244	0,2342%	2 244	64	765 737	1,9128%	14 647	93	92 830	22,6765%	21 051
5	992 474	0,0371%	368	35	956 000	0,2477%	2 368	65	751 090	2,0784%	15 611	94	71 779	26,7793%	19 222
6	992 105	0,0334%	332	36	953 632	0,2638%	2 515	66	735 479	2,2540%	16 578	95	52 557	31,4613%	16 535
7	991 773	0,0316%	313	37	951 116	0,2809%	2 672	67	718 901	2,4163%	17 371	96	36 022	36,7363%	13 233
8	991 460	0,0316%	313	38	948 445	0,3004%	2 849	68	701 530	2,5912%	18 178	97	22 789	42,5865%	9 705
9	991 147	0,0297%	294	39	945 596	0,3233%	3 057	69	683 352	2,7789%	18 989	98	13 083	48,9604%	6 406
10	990 853	0,0316%	313	40	942 539	0,3497%	3 296	70	664 363	2,9820%	19 811	99	6 678	55,7429%	3 722
11	990 541	0,0316%	313	41	939 243	0,3816%	3 584	71	644 551	3,2014%	20 635	100	2 955	62,7667%	1 855
12	990 228	0,0372%	368	42	935 699	0,4143%	3 877	72	623 917	3,4342%	21 426	101	1 100	69,8028%	768
13	989 860	0,0442%	438	43	931 782	0,4501%	4 194	73	602 490	3,6827%	22 188	102	332	76,5727%	254
14	989 422	0,0543%	538	44	927 588	0,4877%	4 524	74	580 303	3,9498%	22 921	103	78	82,7772%	64
15	988 884	0,0716%	708	45	923 065	0,5271%	4 866	75	557 382	4,2415%	23 641	104	13	88,1389%	12
16	988 176	0,0907%	896	46	918 199	0,5683%	5 199	76	533 740	4,5624%	24 351	105	2	100,0000%	2
17	987 279	0,1128%	1 113	47	912 999	0,6041%	5 515	77	509 389	4,9122%	25 022	106	0	100,0000%	0
18	986 166	0,1337%	1 319	48	907 484	0,6408%	5 815	78	484 367	5,3001%	25 672	107	0	100,0000%	0
19	984 848	0,1489%	1 466	49	901 669	0,6774%	6 108	79	458 695	5,7439%	26 347	108	0	100,0000%	0
20	983 382	0,1578%	1 551	50	895 561	0,7141%	6 395	80	432 348	6,2563%	27 049	109	0	100,0000%	0
21	981 830	0,1633%	1 603	51	889 168	0,7540%	6 705	81	405 299	6,8398%	27 722	110	0	100,0000%	0
22	980 227	0,1665%	1 632	52	882 462	0,7962%	7 026	82	377 577	7,4844%	28 259				
23	978 596	0,1689%	1 653	53	875 435	0,8406%	7 359	83	349 317	8,1750%	28 557				
24	976 943	0,1721%	1 681	54	868 076	0,8884%	7 712	84	320 761	8,8986%	28 543				
25	975 262	0,1753%	1 710	55	860 365	0,9375%	8 066	85	292 218	9,6665%	28 247				
26	973 552	0,1800%	1 753	56	852 299	0,9885%	8 425	86	263 970	10,4879%	27 685				
27	971 800	0,1847%	1 795	57	843 874	1,0565%	8 941	87	236 285	11,3595%	26 841				
28	970 004	0,1895%	1 838	58	834 933	1,1416%	9 531	88	209 445	12,2792%	25 718				

FIGURE 4.2 – Table CIMA

## 4.2 Notions de base des modèles de durée

Un modèle de durée décrit le temps qui s'écoule (durée de vie d'un individu, durée entre la demande d'un prêt et une défaillance de remboursement). Ces modèles sont utilisés pour l'établissement de tables de mortalité, mais aussi en finance (par exemple, pour modéliser les risques de défaillance).

Comme on l'a vu lors du précédent cours, les premières études ont été réalisées au 18e siècle mais c'est seulement au 19e siècle que se développent réellement les premiers travaux techniques permettant de décrire mathématiquement les taux de mortalité. Par exemple, en 1825, Gompertz utilise un modèle de décrois-

sance géométrique pour décrire le nombre de survivants. Puis, le modèle de Weibull en 1952 considère que le taux de mortalité croît de manière géométrique. En 1972, celui de Cox prend en compte la dépendance des variables.

Dans la suite de ce cours, on définit une durée de survie par une variable aléatoire positive absolument continue  $T \in (\Omega, \mathcal{F}, \mathbb{P})$  de fonction de densité  $f$  et de fonction de répartition  $F$ . On caractérise souvent les lois des temps de survie par les fonctions de survie et de hasard définies ci-dessous.

**Définition 4.2.1. Fonction de survie :**

Soit  $t$  un réel positif. La fonction de survie  $S$  de  $T$  est définie par  $S(t) = \mathbb{P}(T > t)$ . Cette fonction correspond à la probabilité de survie jusqu'au temps  $t$  de l'individu observé.

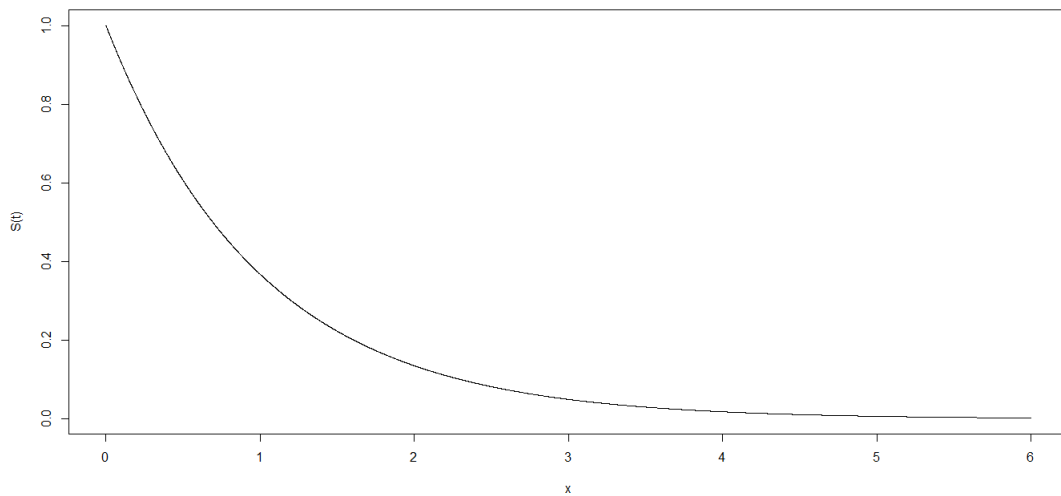


FIGURE 4.3 – Fonction de Survie d'une variable T suivant une loi exponentielle de paramètre 1

Par exemple, dans l'étude de la longévité, pour un âge  $x$ ,  $L_x$  est le nombre de survivant et donc  $S(x) = \frac{L_x}{L_0}$  représente bien la probabilité d'être en vie.

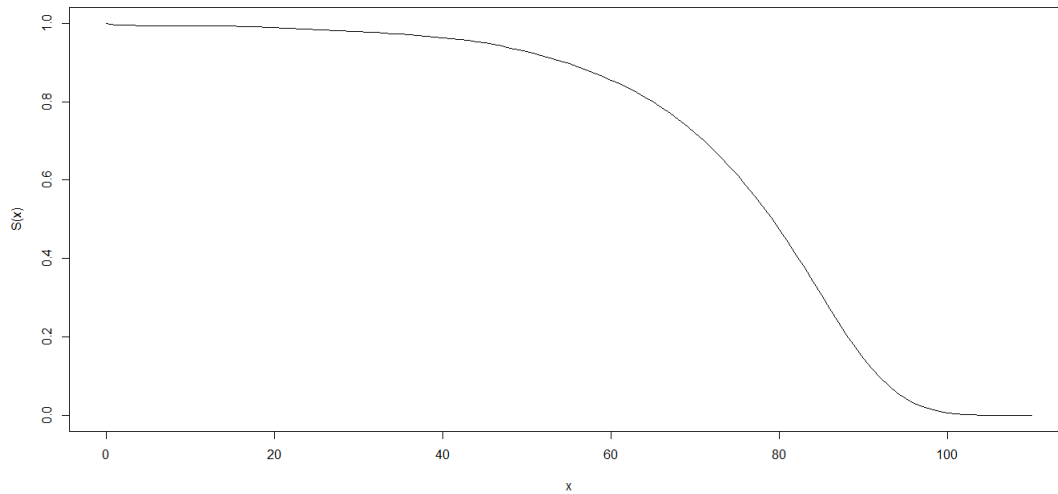


FIGURE 4.4 – Fonction de Survie d'individu de la table TH0002

**Définition 4.2.2.** *Fonction de hasard :*

Soit  $t$  un réel positif. La fonction de hasard  $S$  de  $T$  est définie par  $h(t) = \frac{f(t)}{S(t)}$ .

Cette fonction correspond à la probabilité de décès dans un temps infinitésimalement petit sachant que l'individu observé est un vie au temps  $t$ .

Dans l'étude de la longévité, on note  $\mu_x := h(x)$  "le taux instantané de mortalité".

Cette terminologie est justifiée par la proposition suivante :

**Proposition 4.2.1.**

Soit un réel positif  $t$ ,  $h(t) = \lim_{u \rightarrow 0} \frac{P(T < t+u | T > t)}{u}$

**Preuve :**

$$h(t) = \frac{f(t)}{S(t)}$$

$$\text{et } f(t) = F'(t) = -S'(t)$$

$$\text{donc } h(t) = -\frac{S'(t)}{S(t)} = -\frac{d \ln(S(t))}{dt}$$

$$h(t) = \frac{1}{S(t)} \lim_{u \rightarrow 0} \frac{S(t) - S(t+u)}{u}$$

$$h(t) = \frac{1}{P(T > t)} \lim_{u \rightarrow 0} \frac{P(T > t) - P(T > t+u)}{u}$$

$$h(t) = \lim_{u \rightarrow 0} \frac{P(t < T < t+u)}{u P(T > t)}$$

$$h(t) = \lim_{u \rightarrow 0} \frac{P(T < t+u | T > t)}{u}$$

Les modèles de durée sont caractérisés par des observations à information incomplète. On modélise cette perte d'information par des censures et des troncatures.

**Définition d'une censure :** Soit un entier naturel  $n$ .

Soit :

- Un échantillon de  $n$  temps de survie réels  $(T_1, \dots, T_n)$
- Une censure de taille  $n$   $(C_1, \dots, C_n)$

On note l'échantillon censuré  $((X_1, D_1), \dots, (X_n, D_n))$  où pour tout  $i$  entre 1 et  $n$   $X_i$  est le temps de survie observé et  $D_i$  représente la présence ou l'absence de censure sur l'individu  $i$ .

**Exemple :**

Si on étudie la durée de vie d'un individu qui dispose d'un contrat d'assurance-vie, on peut observer la vie de l'individu entre la date de sa souscription et la date de son rachat total. De plus, on sait qu'il est en vie le jour de la souscription et qu'il n'est pas mort lors

de son rachat total.

Prenons le cas des censures à droite. Il existe principalement trois types de censures.

**Définition 4.2.3. Censure de type I :**

Soit  $C$  un réel positif, dans ce cas  $(C_1, \dots, C_n) = (C, \dots, C)$ . On observe donc la durée de vie de l'individu  $i$  que lorsque  $T_i < C$ .

$$X_i = \min(X_i, C) \text{ et } D_i = \mathbb{1}_{\{T_i < C\}}.$$

**Définition 4.2.4. Censure de type II :**

On range les durées dans l'ordre croissant  $(T_{(1)}, \dots, T_{(n)})$ .

On observe donc la durée de vie de l'individu  $i$  jusqu'à ce que le  $k^{eme}$  meurt.

$$\text{On a donc } X_i = \min(T_i, T_{(k)}) \text{ et } D_i = \mathbb{1}_{\{T_i < T_{(k)}\}}.$$

**Définition 4.2.5. Censure de type III :**

$(C_1, \dots, C_n)$  sont des variables aléatoires indépendantes identiquement distribuées.

$$X_i = \min(T_i, C_i) \text{ et } D_i = \mathbb{1}_{\{T_i, C_i\}}.$$

Cette censure est couramment observée pour l'étude de la longévité dans un contrat d'assurance.

De même, on peut définir des censures à gauche et des censures par intervalles.

**Définition 4.2.6. Troncature**

On définit une troncature lorsque la durée de survie  $T_i$  n'est observable que dans un certain sous-ensemble de réels positifs.

Contrairement aux censures, nous ne possédons aucune information sur cette durée de survie lorsqu'elle est en-dehors de son ensemble d'observation.

### 4.3 Premiers développements de modèles pour les taux de mortalité

Comme on l'a vu, les tables de mortalité sont des tableaux attribuant pour un âge donné, un nombre de survivants en comptant une cohorte initiale de 100 000 individus par exemple.

La fonction de hasard est apparentée au taux de mortalité et la fonction de survie représente le nombre de survivants divisé par le nombre d'individus dans l'échantillon de départ.

En notant  $\mu_x$  le taux instantané de mortalité et  $l_x$  le taux de survivants à l'âge  $x$ , on a la relation suivante liant les taux aux nombre de survivants. On obtient la formule suivante :  $\mu_x = -\frac{d \ln(l_x)}{dx}$ .

Cette relation est démontrée dans la section précédente.

On peut ainsi obtenir par intégration le taux de survivants en connaissant les taux instantanés de mortalité.

#### Exemple de loi de mortalité :

— *Loi exponentielle de paramètre  $a < 1$  :*

On suppose dans ce cas que  $l_x = a^x$ , c'est-à-dire que le nombre de survivants décroît de façon géométrique ; plus  $a$  est proche de 0, plus la décroissance est rapide. Ce modèle suppose donc un taux de mortalité instantané constant, ce qui ne reflète pas la réalité. En effet, les personnes âgées ont clairement un taux de mortalité plus important que les personnes d'une vingtaine d'année. La nécessité d'avoir un modèle plus fin s'impose alors.

— *1825 : Modèle de Gompertz de paramètre  $(B,c)$  :*

Ce modèle propose de corriger le problème soulevé précédemment en supposant que les taux instantanés de mortalité



augmentent de façon exponentiel avec l'âge :  $\mu_x = Bc^x$

$$l_x = \exp\left(-\int_0^x (\mu_s) ds\right) = \exp\left(-\frac{B}{\ln(c)}(c^x - 1)\right)$$

Le nombre de survivants décroît donc de façon double exponentielle en fonction de l'âge.

Même si ce modèle semble plus réaliste que le précédent, il n'explique pas la part de la mortalité qui n'est pas liée au vieillissement de la population. Ainsi, le modèle de Gompertz-Makeham expliqué ci-dessous est le premier à avoir expliciter un autre facteur de mortalité.

- 1860 : *Modèle de Gompertz-Makeham de paramètre (A,B,c)* : Dans ce modèle la constante  $A$  vise à expliquer la mortalité accidentelle :  $\mu_x = A + Bc^x$   
Les autres paramètres peuvent s'interpréter de la même manière que pour le modèle de Gompertz.

## Chapitre 5

# Description des données du problème

### 5.1 Choix de nos données

Les données que nous allons vous présenter sont proches de celles que nous pouvons retrouver dans une compagnie d'assurance-vie. Ces données ont été simulées à partir des tables de mortalité TH0002 et TF0002. Nous allons utiliser ces données afin de calculer et d'ajuster des modèles sur les taux de mortalité. Les codes R associés aux différentes étapes de chapitre seront présentés en annexe A.

1	Cnt	Pays	TypFisc	Reseau	TxGaranti	Fonds	CSP	Sexe	Civilite	DatNaiss	DatEffetCnt	DatEffet	Typo	Montant	
2	71942509	FRA	STD	Salarie	0.5	EUR	NR	M	M	16/12/1953	18/03/2005	29/10/2006	DC	282.467139297896	
3	34338140	FRA	STD	Salarie		2 EUR	inactif divers	F	MME	11/12/1956	15/07/2003	15/07/2003	FraisSous	130.05	
4	87477248	FRA	STD	Salarie		0 EUR	cadre foncti	M	M	12/11/1918	05/09/2007	05/09/2007	FraisSous	135.6	

1	Cnt	Pays	ChargementGestion	CSous	CommSous	FreqChgts	TypFisc	Reseau	TxGaranti	Fonds	CSP	Sexe	Civilite	DatNaiss	DatEffet	DatDC	CodSitu	PM
2	66063878	FRA	0.99		4	0	4 STD	Salarie		0 EUR	NR	M	M	18/10/2011	12/04/2012		En Cours	131.16
3	27158756	FRA	0.99		3.5		4 STD	Salarie	2.5	EUR	employ fonc	M	M	31/10/1950	07/11/2006		En Cours	414.94
4	49320745	FRA	0.99		4	3.75	4 STD	Courtier	0.5	UC	employ adm	M	M	26/06/1957	30/12/2003		En Cours	3700.21

FIGURE 5.1 – Premières lignes des base de mouvements et base de contrat

La base de mouvements présente les différentes actions réalisées sur les contrats, comme par exemple la souscription, le rachat total, les règlements de frais de souscription ou les décès. La nature de ces actions est renseignée dans la colonne "typo". Les autres colonnes qui nous intéressent sont : "DatNaiss" qui indique

la date de naissance du souscripteur, "DatEffet" qui donne la date à laquelle l'action a été effectuée et "Cnt" qui donne l'identifiant lié au contrat. D'autres variables sont des informations complémentaires de la base comme la catégorie socioprofessionnelle du souscripteur.

La base de contrats nous indique la date d'ouverture des différents contrats ainsi que la date de naissance des assurés pour ces contrats.

Afin de calculer les taux de mortalités  $q_x$  à l'aide de ces tables, il faut préparer les données. Dans un premier temps, on ne garde que les variables qui nous intéressent dans les 2 bases, à savoir Cnt, DateEffet et Typo (+ 'RT') dans base mouvements et Cnt, DatNaiss, DatEffet, DatDC dans base contrats. Pour finir, on fusionne les 2 tables par rapport aux numéros de contrats dans une table que nous nommerons Base.

## 5.2 Calculs d'exposition et calcul des taux de mortalité

On évalue l'âge, l'exposition et le décès ou non de chaque individu pour chaque année "an" à partir de la Base afin de calculer les nombres de décès de  $d_x$  et les nombres de personne exposées  $l_x$ . On suppose que chaque assuré a le même âge qu'au 1<sup>er</sup> Janvier pendant toute l'année "an".

Pour calculer l'exposition par individu, on commence par créer des vecteurs de booléens contenant :

- Décès ou non pendant l'année l'année an
- Mort ou non avant l'année an

- Rachat total ou non avant l'année an
- Rachat total ou non pendant l'année an
- Contrat souscrit ou non avant l'année an
- Contrat souscrit ou non pendant l'année an

On calcule ensuite le nombre de jours :

- Entre la date de rachat et le premier jour de l'année
- Entre la date de décès et le premier jour de l'année
- Entre le dernier jour de l'année et la date de souscription
- Entre la date de rachat et la date de souscription
- Entre la date de décès et la date de souscription

Finalement, l'exposition individuelle pour l'année an est calculée comme ceci :

- Si le contrat est actif pendant toute l'année an :  
Ce ratio est de 1
- Si le contrat est actif pendant une partie de l'année an :  
Ce ratio est le prorata temporis durant lequel le contrat est actif
- Si le contrat n'est pas actif pendant l'année an :  
Ce ratio est nul

Nous obtenons alors une matrice contenant l'âge, l'exposition et le décès ou non de chaque individu pour chaque année "an"  
 Nous effectuons maintenant la somme des expositions, ainsi que la somme des décès ou non par âge afin d'obtenir le nombre de personne exposées  $l_x$  et le nombres de décès de  $d_x$  de l'année n.  
 Nous calculons les  $q_x$  pour l'année an grâce à l'équation  $q_x = l_x / d_x$ .  
 A présent, nous pouvons construire une matrice de  $q_{xt}$  avec  $x$  l'âge des assurés et  $t$  l'année afin d'ajuster des modèles sur ces valeurs.

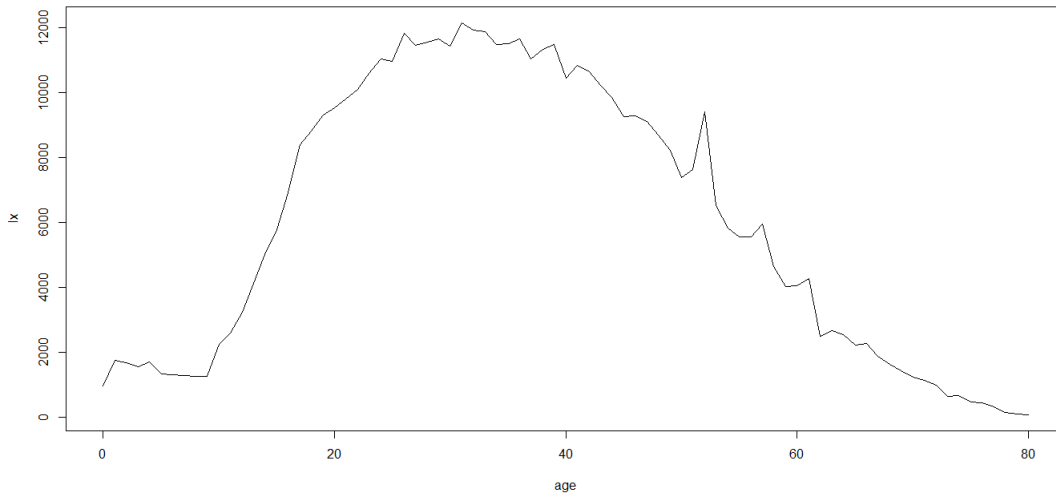


FIGURE 5.2 – Population exposée en fonction de l'âge en 2006

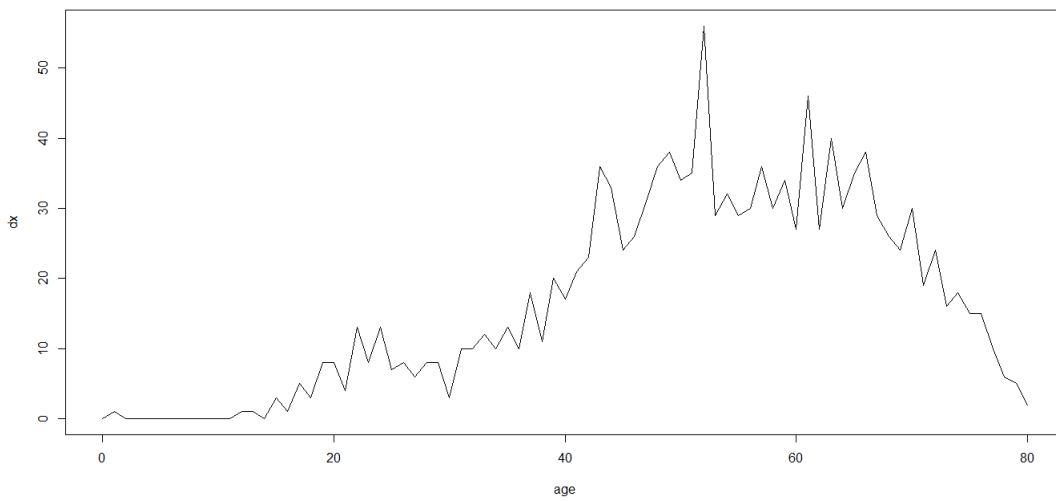


FIGURE 5.3 – Nombre de décès en fonction de l'âge en 2006

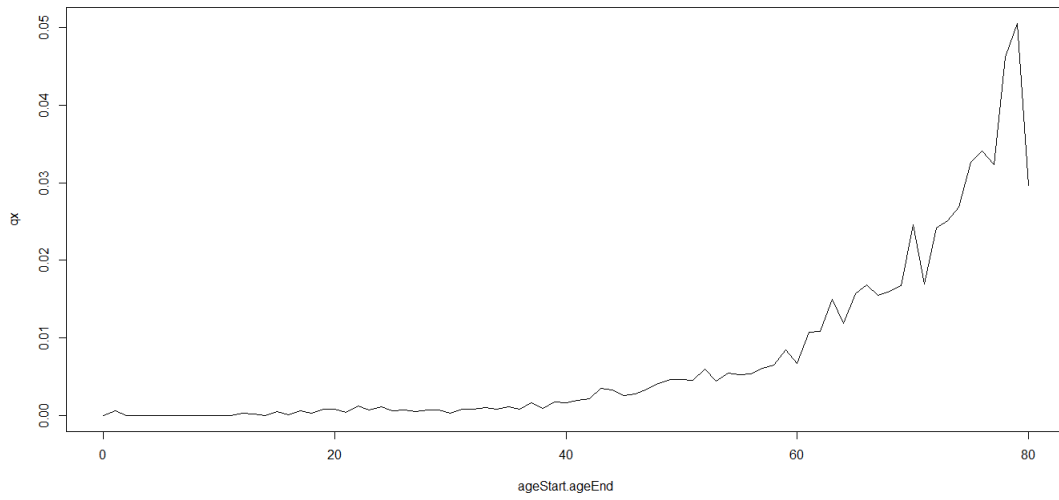


FIGURE 5.4 – Quotient de mortalité en fonction de l'âge en 2006

## Chapitre 6

# Méthodes d'ajustement

Nous donnerons une présentation générale de différents modèles simples couramment utilisés pour la construction de tables de mortalité. Cependant, le modèle de Lee-Carter sera présenté plus en détail dans le chapitre suivant.

### 6.1 Modèles paramétriques

Ces modèles paramétriques s'ajustent sur le taux de mortalité instantané  $m_{u_{xt}}$ . Cependant, dans le modèle de Gompertz-Makeham, on suppose que ce taux ne dépend que de l'âge dans la population.

Le modèle de Gompertz-Makeham est donc déterministe par rapport au temps et le modèle de Lee-Carter permet de construire des tables de mortalités prospectives, c'est-à-dire des tables qui prédisent la mortalité future.

#### 6.1.1 Gompertz-Makeham

Le modèle de Gompertz-Makeham est un des modèles de référence pour la construction de tables de mortalité. Il vise à estimer les taux de mortalité  $\mu_x$  pour l'âge  $x$  par  $\mu_x = A + Bc^x$

Dans ce modèle la constante  $A$  appelée terme de Makeham vise à expliquer la mortalité accidentelle et le terme de Gompertz  $Bc^x$

dépend du temps et croit de manière exponentielle. Ce modèle donne de bons résultats pour les âges compris entre 30 et 80 ans. Cependant, il a tendance à surestimer la mortalité pour les âges les plus avancés. Estimation des paramètres pour les données :

### 6.1.2 Lee-Carter

Ce modèle vise à estimer le logarithme des taux instantanés de mortalité  $\mu_{xt}$  à l'année  $t$  pour l'âge  $x$  en fonction de deux vecteurs dépendants de l'âge  $\alpha$  et  $\beta$  et d'un vecteur temporel  $\kappa$  sous la forme :

$$\ln(\mu_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

- $\alpha_x$  peut être interprété comme la moyenne temporelle du logarithme du taux de mortalité de l'âge  $x$ .
- $\kappa_t$  explique l'évolution temporelle de la mortalité de la population étudiée.
- $\beta_x$  représente la sensibilité pour chaque âge  $x$  en fonction de  $\kappa_t$ .
- $\epsilon_{xt}$  est le terme résiduel gaussien pour l'âge  $x$  et l'année  $t$ , les résidus sont supposés gaussiens et identiquement distribués.

Ce modèle sera expliqué plus en détail dans le chapitre suivant.

## 6.2 Modèles semi-paramétriques

### 6.2.1 Cox

Ce modèle fait partie d'une classe de modèles semi-paramétriques appelés modèles à hasard proportionnel.

Posons  $T_\theta$  la variable aléatoire des durées observées,  $\theta$  étant un



paramètre influant sur  $T$ . Notons  $f_\theta$  sa densité si elle existe, et  $h_\theta$  sa fonction de hasard.

On note  $S(t)$  une fonction de survie connue,  $f(t)$  la densité associée et  $h(t)$  la fonction de hasard associée.

La fonction de survie de la variable  $T_\theta$  est  $S(t)^\theta$ . En dérivant la fonction de survie de  $T_\theta$ , on a  $f_\theta(t) = \theta S(t)^{\theta-1} f(t)$  et on obtient la fonction de hasard  $h_\theta$  de  $T_\theta$  telle que, pour  $t$  réel positif, on a :

$$h_\theta = \frac{\theta f(t)}{S(t)} = \theta h(t)$$

Le modèle de Cox suppose que  $\theta = \exp(Xi'\beta)$  pour un individu  $i$ .

Dans le cas de l'étude de la mortalité, ce modèle vise à exprimer le taux instantané de mortalité  $\mu_x$  d'un individu  $i$  à l'âge  $x$  connaissant les caractéristiques modélisées par le vecteur de covariables  $Xi$  de cet individu :

$$\mu_{x|Xi} = \mu_x^0 * \exp(Xi'\beta)$$

- $\mu_x^0$  correspond au risque instantané de décès de base. C'est celui que l'on obtient lorsque toutes les covariables sont nulles. C'est le seul terme qui dépend de l'âge de l'individu.
- $\exp(Xi'\beta)$  ne dépend que des covariables et le vecteur  $\beta$  mesure l'influence des covariables sur le taux instantané de mortalité  $\mu_x$ .

On remarque que les quotients des taux instantanés de mortalité de deux individus différents ne dépendent pas de l'âge mais seulement des caractéristiques de ces individus. Ce modèle est donc utile pour comparer la mortalité de deux groupes différents.

### 6.2.2 Brass

Le modèle de Brass est un modèle de régression linéaire sur le logit du taux instantané de mortalité  $\text{logit}(q_x)$  par le logit du taux instantané de mortalité d'une table de mortalité de référence  $\text{logit}(q_x^{\text{ref}})$  :

$$\text{logit}(q_x) = \beta_1 + \beta_2 * \text{logit}(q_x^{\text{ref}})$$

On rappelle que la fonction logit est définie par :

Pour tout  $x$  dans l'intervalle  $]0;1[$   $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$

Ce modèle peut être utile pour ajuster une table de mortalité sur des données avec peu d'observations en la comparant avec une table de mortalité de référence.

## 6.3 Modèle non-paramétrique

### 6.3.1 Kaplan-Meier

On estime en général la fonction de survie par la fonction de survie empirique, comme on le fait avec la fonction de répartition. On a comme estimation de la fonction de survie :

$$S_n(x) = \frac{1}{n} \sum_{i=1}^n 1(T_i > x)$$

Cependant, cet estimateur ne prend pas en compte les censures. C'est pourquoi on va trouver un nouvel estimateur.

En notant  $t_1 < t_2 < \dots < t_K$  les instants d'observation et  $I_j = ]t_{j-1}, K - t_{j,K}]$  les intervalles d'observation, on pose :

- $m_{j,K}$  le nombre de décès dans  $I_j$
- $n_{j,K}$  le nombre d'individus dans  $I_j$
- $c_{j,K}$  le nombre de censures dans  $I_j$

On notera que le nombre de décès, d'individus et de censures dépendent du nombre  $K$  d'intervalles. On obtient alors une estimation des taux de mortalité pour l'intervalle  $I_j$  :

$$q_j^1 = \frac{m_{j,K}}{n_{j,K} - c_{j,K}}$$

En exprimant la fonction de survie en fonction des taux de mortalité, on obtient l'estimation de la fonction de survie en  $t$  :

$$S^1(t) = \prod_{j|t_j < t} (1 - q_j^1)$$

En effet :

$$S(t_j) = P(T > t_j) = \prod_{i=1}^j P(T > t_i | T > t_{i-1}) = \prod_{i=1}^j (1 - q_i)$$

avec  $q_i$  le taux de mortalité pour l'intervalle  $i$ .

On a alors un premier estimateur de la fonction de survie :

$$S^1(x) = \prod_{j|t_j < x} \left(1 - \frac{m_{j,K}}{n_{j,K} - c_{j,K}}\right)$$

Kaplan et Meier ont alors réfléchi à faire tendre la taille des intervalles vers 0 lorsque le nombre d'intervalles tend vers l'infini.

En notant :

- $X(1) < X(2) < \dots < X(i)$  les observations ordonnées
- $m_i$  le nombre de décès à la date  $X(i)$
- $n_i$  le nombre d'individus "à risque" à la date  $X(i)$

On fait tendre la taille des intervalles vers 0 lorsque le nombre  $K$  d'intervalles tend vers l'infini. Pour cela, il suffit que :

$$\sup_{1 \leq j \leq K} |t_{j,K} - t_{j-1,K}| \rightarrow 0$$

Lorsque  $K \rightarrow \infty$ , on a alors :

$$S^1(t) \rightarrow \prod_{i=1}^k \left(1 - \frac{m_i}{n_i}\right)^{1(X(i) \leq t)}$$

La limite obtenue est l'estimateur de Kaplan-Meier qui est aussi appelé "estimateur produit-limite".

## Chapitre 7

# Le modèle de Lee-Carter

Table de mortalité prospective :

Dans certains cas, les assureurs utilisent une extrapolation des valeurs passées des taux instantanés de mortalité pour calculer des tables prospectives, c'est-à-dire des tables qui prennent en compte les tendances temporelles de la mortalité.

Nous avons choisi d'expliquer le modèle de Lee-Carter, ainsi que sa variante le modèle de log-Poisson, plus en détail car ce sont des modèles stochastiques classiques.

Les codes R d'ajustement de ces modèles sur les données HMD France seront disponibles en annexe B de ce cours.

### 7.1 Présentation et mise en place du modèle

Ce modèle vise à estimer le logarithme des taux instantanés de mortalité  $\mu_{xt}$  à l'année  $t$  pour l'âge  $x$  en fonction de deux vecteurs dépendants de l'âge  $\alpha$  et  $\beta$  et d'un vecteur temporel  $\kappa$  sous la forme :

$$\ln(\mu_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}$$

- $\alpha_x$  peut être interprété comme la moyenne temporelle du logarithme du taux de mortalité de l'âge  $x$ .

- $\kappa_t$  explique l'évolution temporelle de la mortalité de la population étudiée.
- $\beta_x$  représente la sensibilité pour chaque âge  $x$  en fonction de  $\kappa_t$ .
- $\epsilon_{xt}$  est le terme résiduel pour l'âge  $x$  et l'année  $t$ .

**Mise en place du modèle :**

On notera  $t_{min}$  et  $t_{max}$  les années d'observation minimale et maximale. On notera de même  $x_{min}$  et  $x_{max}$ .

En utilisant la méthode des moindres carrés, on cherche le vecteur :

$$(\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t) = \underset{\alpha_x, \beta_x, \kappa_t}{\operatorname{argmin}} \left( \sum_{x,t} \ln(\mu_{xt}) - \alpha_x - \beta_x \kappa_t \right)^2$$

Afin de rendre la solution unique, on impose les contraintes :

$$\sum_{x=x_{min}}^{x_{max}} \beta_x = 1 \qquad \sum_{t=t_{min}}^{t_{max}} \kappa_t = 0$$

On obtient avec ces contraintes :

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

On construit une matrice  $Z = (z_{xt})$  tel que :

$$z_{xt} = \ln(\mu_{xt}) - \hat{\alpha}_x$$

Pour estimer les paramètres  $\beta_x$  et  $\kappa_t$ , on utilise la décomposition SVD de la matrice  $Z$  :

$$Z = U \Sigma V'$$

où :

- $\Sigma$  est la matrice diagonale contenant les valeurs singulières au carré  $(\lambda_1, \dots, \lambda_n)$  de  $Z$  positives rangées dans l'ordre décroissant

- $U$  est la matrice orthogonale des vecteurs singuliers à gauche de  $Z$
- $V$  est la matrice orthogonale des vecteurs singuliers à droite de  $Z$

Le paramètre  $\beta_x$  est estimé par  $\frac{u_1}{\sum_x u_{x1}}$  et  $\kappa_t$  par  $\sqrt{\lambda_1} * \sum_x u_{x1} * v_1$

**Application aux données HMD de la France entre 1950 et 2014 :**

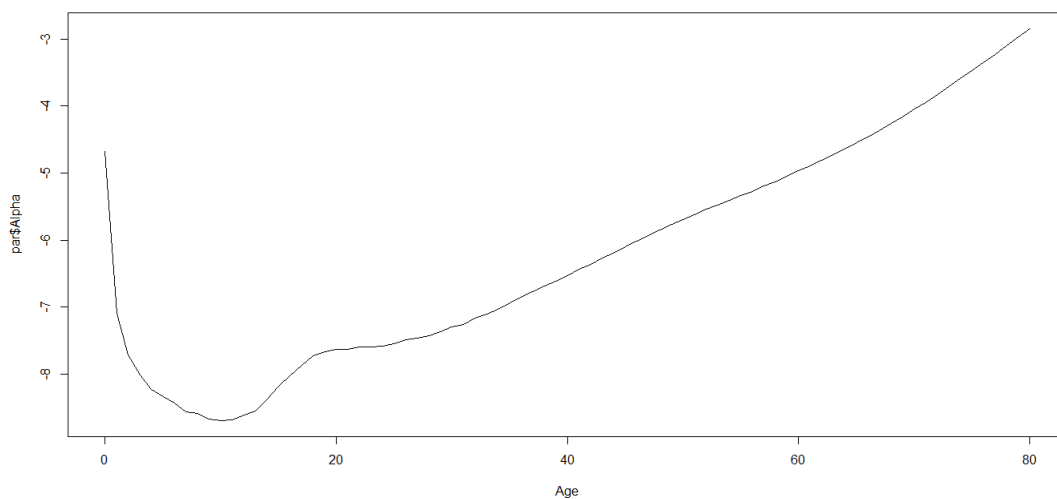


FIGURE 7.1 – Paramètre  $\alpha$  du modèle de Lee-Carter

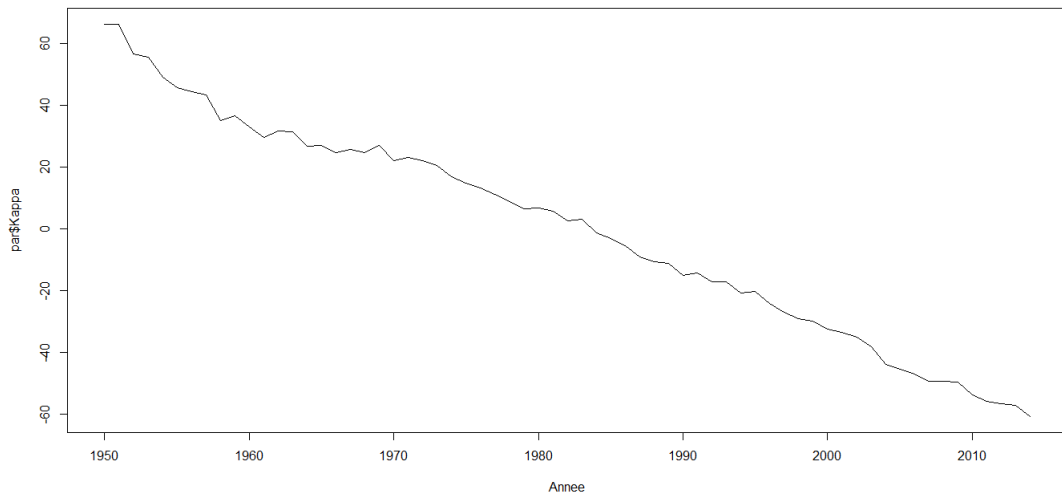


FIGURE 7.2 – Paramètre  $\kappa$  du modèle de Lee-Carter

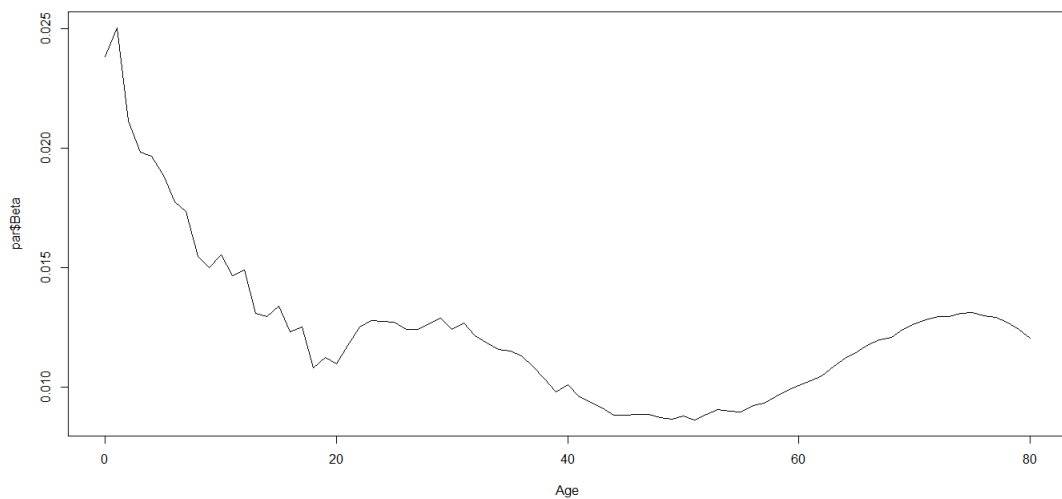


FIGURE 7.3 – Paramètre  $\beta$  du modèle de Lee-Carter

*Interprétation des résultats :*

- En observant le paramètre  $\alpha$  (moyenne temporelle du logarithme du taux de mortalité par âge), on remarque un pic de mortalité infantile à moins de 4 ans puis une augmentation de cette mortalité de 10 à 20 ans. Entre environ 20 et 30 ans, la mortalité reste à peu près constante, ce qui correspond à une surmortalité accidentelle au début de la troisième décennie



de vie. Enfin, la mortalité augmente de manière exponentielle jusqu'à 90 ans.

- Le paramètre  $\kappa$  nous indique l'évolution générale de la mortalité au fil des années. On remarque donc que la mortalité est décroissante entre 1950 et 2014.
- En analysant le paramètre  $\beta$ , on remarque que la mortalité diminue de manière plus importante pour la population de moins de 20 ans que pour le reste de la population. Ce phénomène se traduit par une orthogonalisation des tables de mortalités.

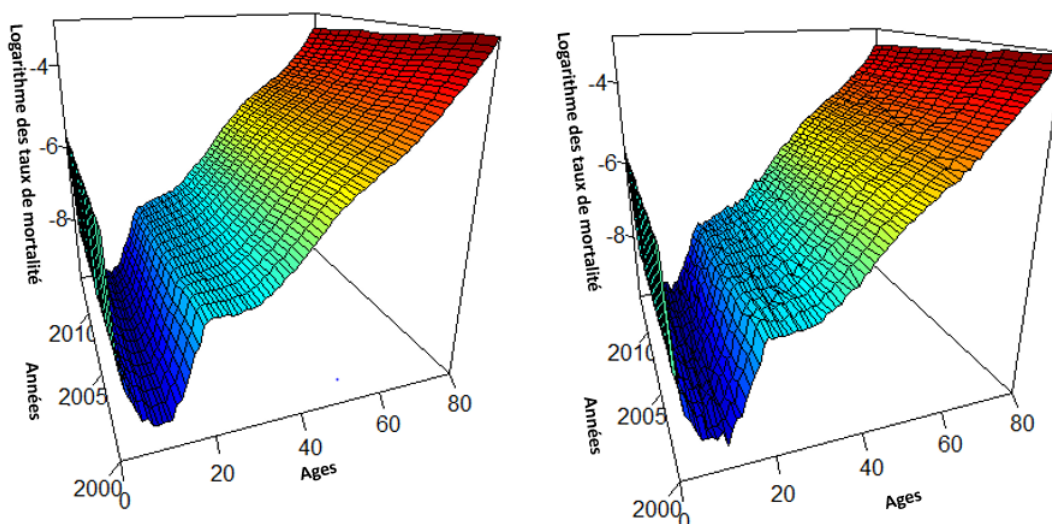


FIGURE 7.4 – Taux instantanés de mortalité issus du modèle de Lee-Carter et réels entre 2000 et 2014

En observant le logarithme des taux de mortalités en fonction de l'âge et du temps obtenu par le modèle de Lee-Carter (figure de gauche), on remarque que l'on a lissé ces taux par rapport à ceux que l'on avait observés (figure de droite), ce qui permet de réduire les variations au cours du temps et de l'âge de cette valeur. On voit aussi que pour toute année comprise entre 2000 et 2014,

la forme générale de la mortalité suit bien celle décrite par le paramètre  $\alpha$  et que, comme nous le décrit le paramètre  $\kappa$ , la mortalité est bien décroissante par rapport au temps.

## 7.2 Projection du modèle

Projection du modèle :

Les paramètres que nous avons obtenus sont ceux des âges allant de 0 à 80 ans.

Pour projeter les taux instantanés de mortalité, on suppose que les paramètres  $\alpha$  et  $\beta$  ne dépendent pas de l'année. On cherche donc à projeter le paramètre  $\kappa$  dans le temps.

**Régression linéaire** Depuis 1950, le paramètre Kappa semble avoir un comportement linéaire en fonction de l'année. Nous allons donc ajuster le modèle linéaire suivant pour toute année  $t$  :

$$\kappa_t = a + bt + \epsilon_t$$

- $a$  et  $b$  sont des constantes réelles
- $\epsilon_t$  est le terme résiduel pour l'année  $t$ , les résidus sont supposés identiquement distribués de loi  $\mathcal{N}(0, \sigma^2)$ .

On estime les paramètres  $a$  et  $b$  par la méthode des moindres carrés. Les intervalles de confiance de niveau 95% pour ces paramètres sont :

$$\left[ \hat{a} - t_{n-2;97,5\%} \sqrt{\frac{\hat{\sigma}^2 \sum_t t^2}{n \sum_t (t - \bar{t})^2}}; \hat{a} + t_{n-2;97,5\%} \sqrt{\frac{\hat{\sigma}^2 \sum_t t^2}{n \sum_t (t - \bar{t})^2}} \right]$$

$$\left[ \hat{b} - t_{n-2;97,5\%} \sqrt{\frac{\hat{\sigma}^2}{\sum_t (t - \bar{t})^2}}; \hat{b} + t_{n-2;97,5\%} \sqrt{\frac{\hat{\sigma}^2}{\sum_t (t - \bar{t})^2}} \right]$$

Avec :

- $\hat{a}$ ,  $\hat{b}$  et  $\hat{\sigma}$  les estimateurs de a, b et  $\sigma$
- $\bar{t}$  l'année moyenne
- $n$  le nombre d'années sur lesquelles on réalise la régression.
- $t_{n-2;97,5\%}$  le quantile d'ordre 97,5% de la loi de Student à n-2 degrés de liberté

Pour projeter le paramètre  $\kappa_{t_0}$  dans le temps pour l'année  $t_0$ , on estime  $\hat{\kappa}_{t_0} = \hat{a} + \hat{b}t_0$  avec un intervalle de prédiction à 95% :

$$\left[ \hat{\kappa}_{t_0} - t_{n-2;97,5\%} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{\sum_t (t - \bar{t})^2}}; \hat{\kappa}_{t_0} + t_{n-2;97,5\%} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{\sum_t (t - \bar{t})^2}} \right]$$

On estime les paramètres a et b pour les années allant de 1950 à 2014 :

On obtient  $\hat{a} = 3615$  (avec un intervalle de confiance à 95% : [3522; 3708]) et  $\hat{b} = -1.824$  (avec intervalle de confiance à 95% : [-1.871; -1.777]).

Le paramètre  $\kappa$  estimé, et donc la mortalité, est bien décroissant dans le temps.

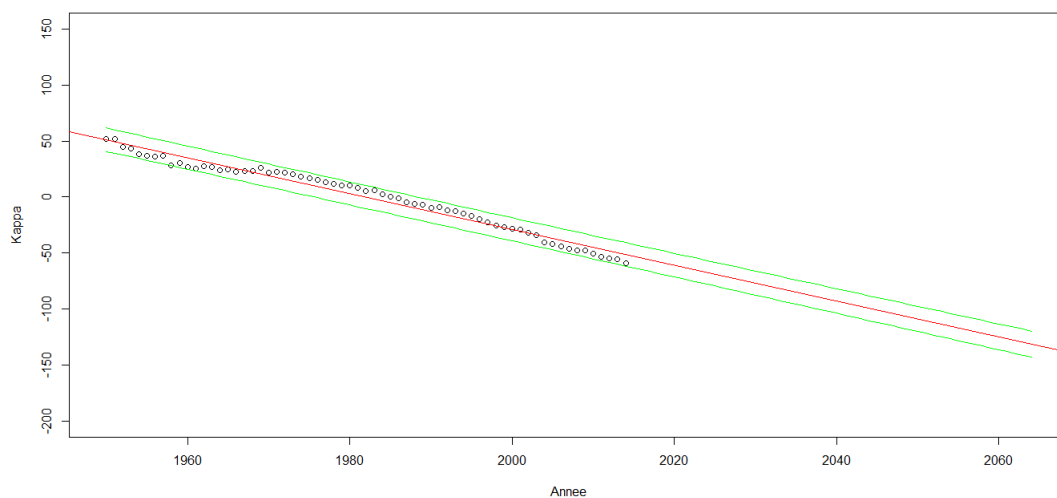


FIGURE 7.5 – Régression linéaire de  $\kappa$  en fonction des années (entre 1950 et 2014) et intervalle de prédiction à 95% jusqu'en 2064

On estime les paramètres  $a$  et  $b$  pour les années allant de 1900 à 2014 :

On obtient  $\hat{a} = 4266$  (avec intervalle de confiance :  $[4103; 4429]$ ) et  $\hat{b} = -2.17$  (avec intervalle de confiance :  $[-2.263; -2.096]$ ). Le paramètre  $\kappa$  estimé décroît plus rapidement que dans le premier modèle.

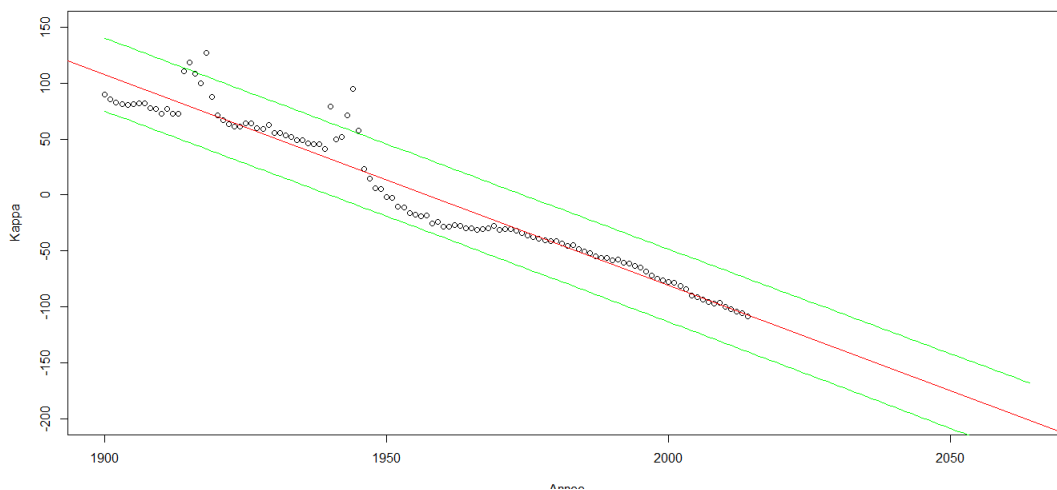


FIGURE 7.6 – Régression linéaire de  $\kappa$  en fonction des années (entre 1900 et 2014) et intervalle de prédiction à 95% jusqu'en 2064

Les valeurs obtenues par des événements extrêmes, comme dans ce cas la guerre, sont éloignées de l'intervalle de prédiction obtenu dans ce modèle.

Ce phénomène nous amène à penser que ce modèle de projection n'est pas forcément le mieux adapté.

**Validation du modèle :** Dans les deux cas, le  $R^2$  qui calcule le pourcentage de variance expliqué par le modèle par rapport à la variance totale est de plus de 95% et les p-values associées aux tests ( $(H_0) : a = 0$  contre  $(H_1) : a \neq 0$  et  $(H_0) : b = 0$  contre  $(H_1) : b \neq 0$ ) sont proches de 0. Les tests sont donc rejetés et le modèle semble adapté.

Cependant, lorsqu'on observe les résidus, ces derniers ne semblent pas gaussiens identiquement distribués dans le temps, particulièrement entre 1900 et 2014. En effet, durant les périodes de guerre, les résidus sont bien trop importants.

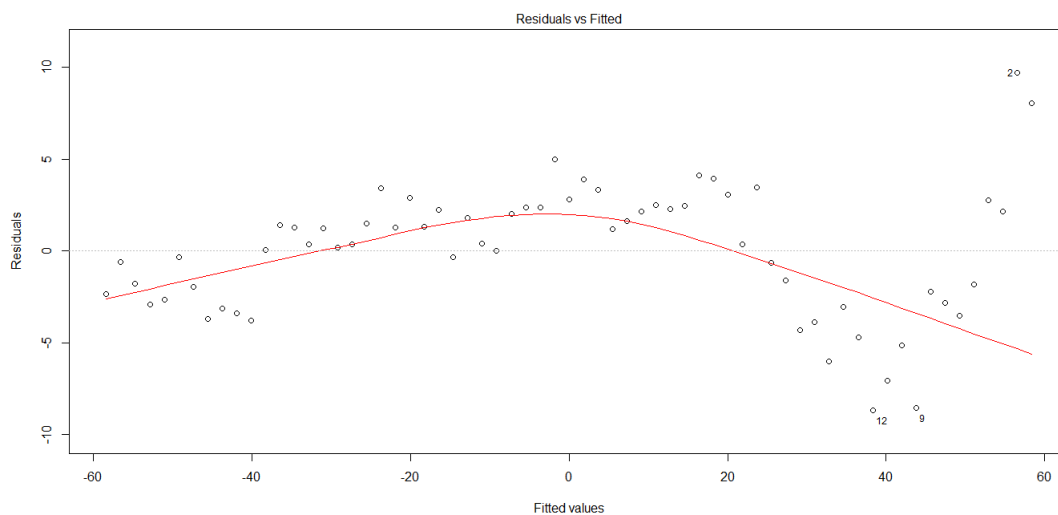


FIGURE 7.7 – Résidus de  $\kappa$  (entre 1950 et 2014)

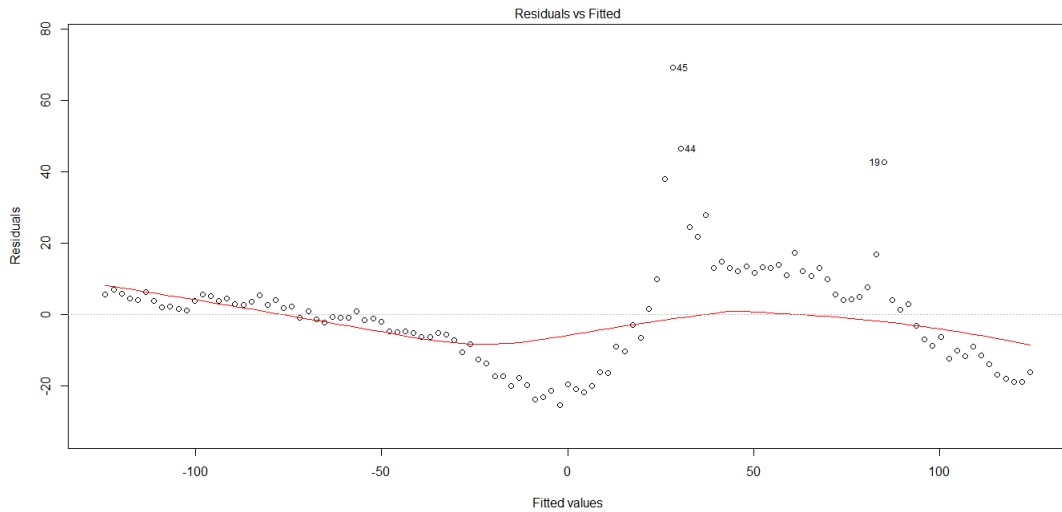


FIGURE 7.8 – Résidus de  $\kappa$  (entre 1900 et 2014)

**Marche aléatoire avec dérive :** Nous allons donc projeter ce paramètre  $\kappa$  avec la méthode originellement adoptée par Lee et Carter. Il s'agit d'une marche aléatoire avec une dérive  $\lambda$ , c'est-à-dire un modèle ARIMA(0,1,0), pour toute année  $t$  :

$$\kappa_t = \lambda + \kappa_{t-1} + \epsilon_t$$

avec  $\epsilon_t$  qui suit indépendamment une loi normale de paramètre  $(0, \sigma^2)$

Estimation des paramètres de 1950 à 2014 :  $\hat{\lambda} = -1.99$  et  $\hat{\sigma} = 4.861$   
 et de 1900 à 2014 :  $\hat{\lambda} = -1.99$  et  $\hat{\sigma} = 53.76$ .

Le processus est en moyenne décroissant de la même manière si l'on considère les deux périodes différentes. Cependant, si on considère une période prenant en compte les deux guerres mondiales, la variance du processus augmente grandement car il y a eu un choc sur les taux de mortalité durant ces guerres.

La trajectoire moyenne du paramètre  $\kappa$  est ici un scénario moyen

dans lequel aucun choc de mortalité à la hausse (épidémie, guerre...) ou à la baisse (avancé importante de la médecine,...) n'est pris en compte.

A partir de ces  $\kappa_t$  projetés, on en déduit les taux instantanés de mortalité projetés et on peut obtenir des tables prospectives transversales de mortalité, c'est-à-dire des tables de mortalité pour une année donnée, ainsi que des tables générationnelles de mortalité. Grâce à cela, on peut par exemple calculer des probabilités de suivi entre deux années futures ainsi que des espérances de vie résiduelles.

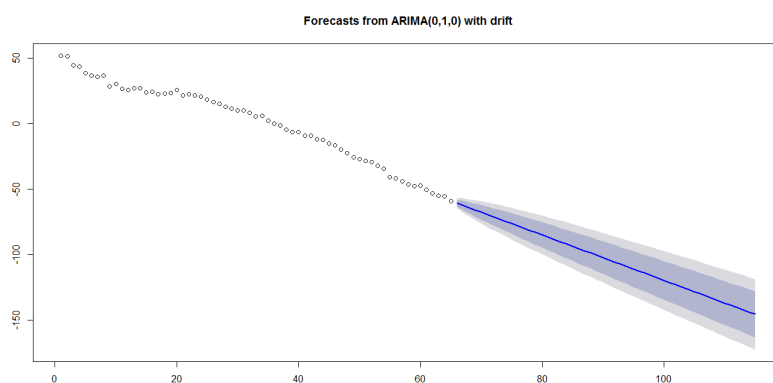


FIGURE 7.9 – Trajectoires de  $\kappa$  entre 1950 et 2014 modélisées par une marche aléatoire (80% en bleu foncé, 95% en bleu clair)

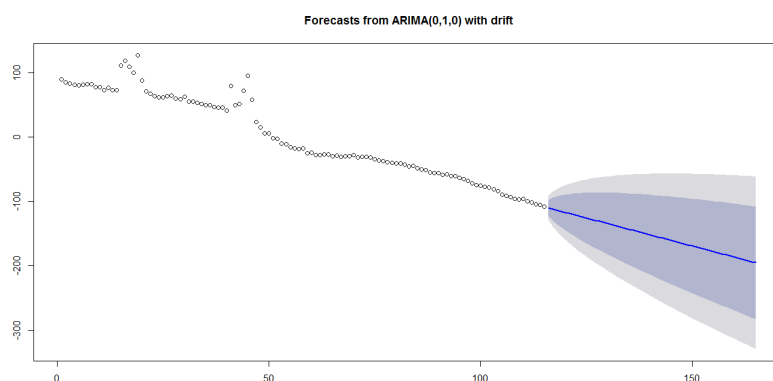


FIGURE 7.10 – Trajectoires de  $\kappa$  entre 1900 et 2014 modélisées par une marche aléatoire

### 7.3 Analyse des résidus et limites du modèle

#### Analyse des résidus :

Le modèle de Lee-Carter suppose que les résidus du logarithme des taux instantanés de mortalité sont gaussiens, indépendants et identiquement distribués. Or, on observe que ces résidus semblent plus importants pour les âges les plus jeunes, l'hypothèse d'homoscédasticité semble donc refusée.

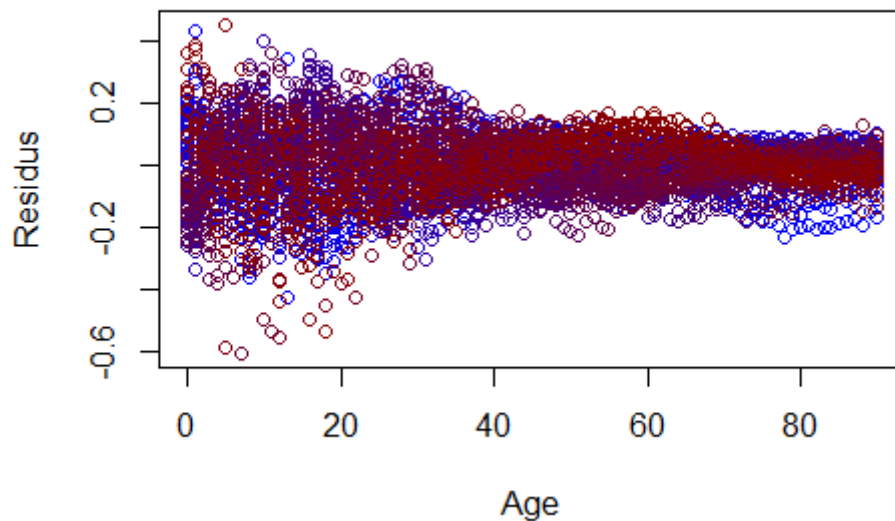


FIGURE 7.11 – Résidus entre 1950 (bleu) et 2014 (rouge)



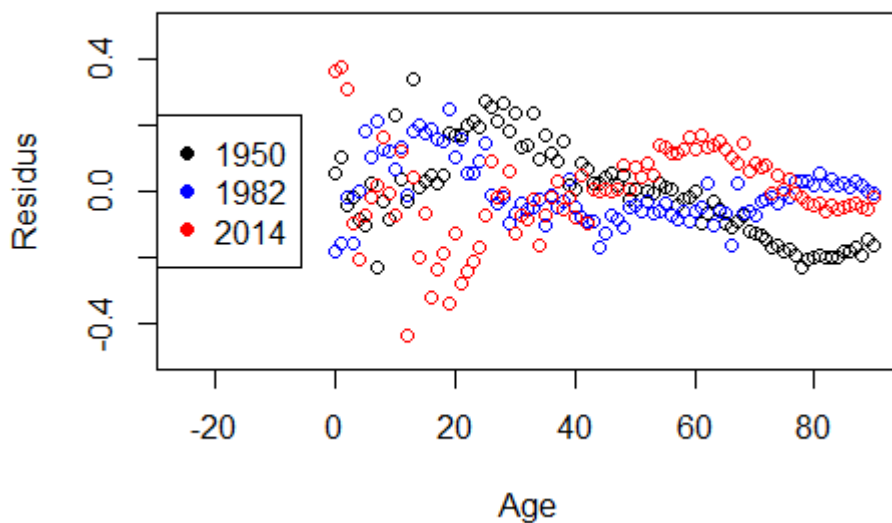


FIGURE 7.12 – Résidus en 1950, 1982 et 2014

**Limites de la méthode :**

L'homoscédasticité du modèle est supposée vraie dans la méthode des moindres carrés. Or, les résidus dans la méthode de Lee-Carter ne sont pas gaussiens indépendants identiquement distribués avec l'âge, l'hypothèse ne semble donc pas réaliste.

De plus, l'approche étant purement extrapolative, elle ne prend pas en compte les avancées de la médecine ou d'autres facteurs qui pourraient impacter de façon importante la mortalité dans le futur.

Le modèle bilinéaire log-poisson, qui est une adaptation du modèle de Lee-Carter, permet d'effacer une partie de ces problèmes et possède de meilleures propriétés probabilistes grâce à une méthode utilisant le maximum de vraisemblance.

**Modèle bilinéaire log-Poisson :**

Le nombre de décès  $d_{xt}$  suit une loi de Poisson de paramètre  $I_{xt}\mu_{xt}$  avec  $I_{xt}$  le nombre de personnes exposées à l'âge  $x$  et au temps  $t$  et :

$$\ln(\mu_{xt}) = \alpha_x + \beta_x \kappa_t$$

Les paramètres du modèle s'interprètent de la même façon que pour le modèle de Lee-Carter et sont estimés par une méthode de maximum de vraisemblance. Dans la pratique, on peut utiliser le package "StMoMo" dans R pour obtenir ces paramètres.

On compare les logarithmes des taux instantanés de mortalité obtenus par les deux derniers modèles allant de 1950 à 2014 pour des âges compris entre 0 à 80 ans avec les logarithmes des taux instantanés de mortalité réels :

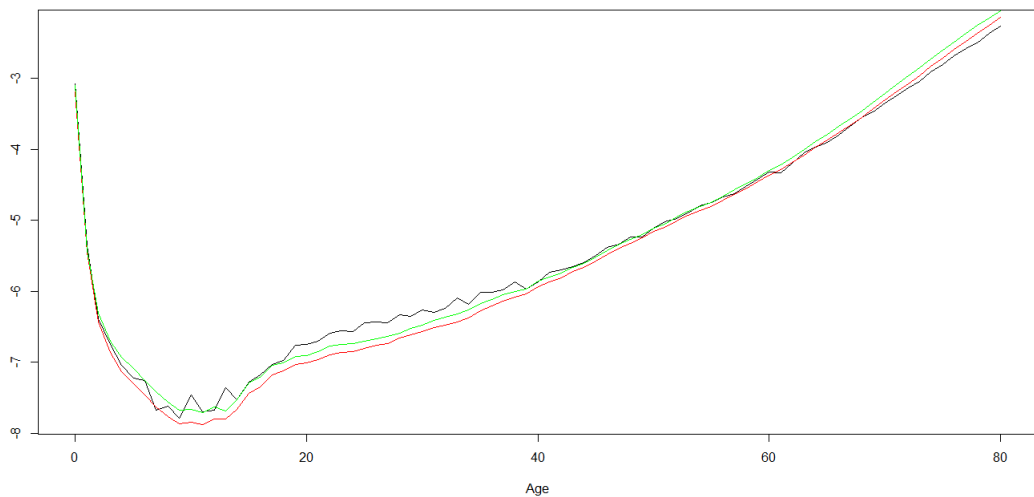


FIGURE 7.13 – Logarithmes des taux instantanés de mortalité en fonction de l'âge en 1950

Ces estimations sont bien dans le même ordre de grandeur.

## Chapitre 8

### Conclusion

En conclusion, nous allons discuter de quelques méthodes permettant d'obtenir de meilleurs résultats dans l'obtention de nos taux de mortalité finaux. On va d'abord décrire comment prendre en compte la forte variabilité des taux pour les âges élevés, puis on expliquera comment lisser les taux de mortalité bruts avant l'ajustement d'un modèle.

Dans une population très âgée, on a en général une forte hétérogénéité des taux de mortalité. En effet, pour des âges élevés, les effectifs dont on dispose sont très faibles, ce qui entraîne une forte volatilité sur nos taux de mortalité bruts et, de fait, sur les taux que l'on aura ajustés grâce à un modèle. Cela entraîne alors un biais entre les taux bruts et les taux ajustés.

On peut alors penser à compléter les taux de mortalité pour les âges élevés. En ayant obtenu nos taux bruts, on ajuste un modèle (par exemple Lee-Carter). Pour les taux au-delà d'un certain âge, on calcule par extrapolation d'autres taux de mortalité que ceux obtenus grâce au modèle : c'est la méthode de Coale et Kisker. La méthode générale va alors être de calculer les taux bruts, d'ajuster un modèle sur ces taux bruts, et d'utiliser les taux obtenus afin de reconstituer les taux pour les âges supérieurs à 80 ans. Coale et Kisker utilisent la notion de taux moyen de croissance qui diminue de manière linéaire après 80 ans. Ce phénomène est illustré sur

la figure ci-dessous.

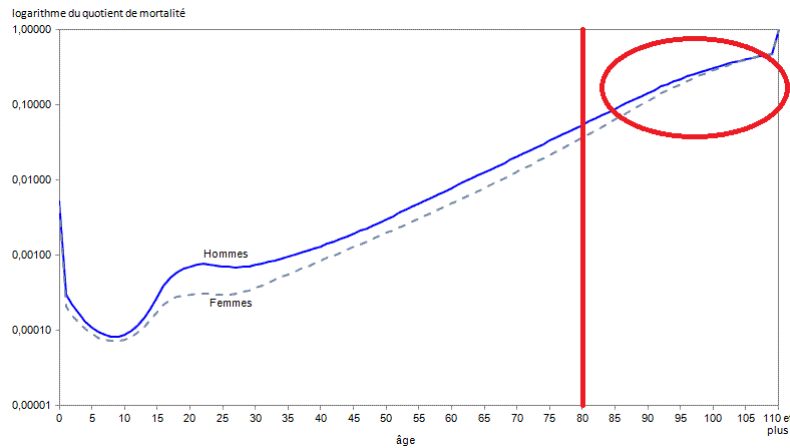


FIGURE 8.1 – Logarithme des taux de mortalité au Canada

On notera qu'on peut aussi utiliser d'autres méthodes que celle-ci afin de palier au problème du biais pour les âges élevés en ajustant d'autres modèles au lieu d'extrapoler les taux qu'on a déjà ajustés. Par exemple, il se peut qu'un modèle de Brass soit plus approprié pour les taux au delà de 80 ans. Dans ce cas, on ajuste un modèle de Lee-Carter jusqu'à 80ans et un modèle de Brass ensuite. C'est ce qu'on appelle l'agrégation de modèles.

On va maintenant voir une autre notion qui est le lissage des taux qui s'effectue sur les taux bruts, et qui s'effectue donc avant d'ajuster un quelconque modèle. On rappelle le déroulement des étapes pour un modèle prospectif : d'abord on calcule les taux bruts, ensuite on ajuste un modèle, enfin on réalise des projections de ces taux. Le lissage des taux intervient donc entre les étapes 1 et 2. A cause notamment des conditions d'observation, les taux bruts peuvent avoir une certaine irrégularité, d'où l'intérêt de faire un lissage qui permet de minimiser l'écart entre les taux bruts et les taux ajustés par un modèle, et ainsi obtenir une courbe des taux

plus lisse qui va devoir respecter deux conditions : la fidélité aux taux bruts et la régularité (il ne faut pas que les taux lissés aient une amplitude de variation trop grande). On a deux catégories de lissages : les lissages paramétriques et non paramétriques. On va présenter un exemple pour chacune de ces catégories.

Le lissage par splines cubiques est un exemple de lissage paramétrique. On suppose qu'il suffit de diviser la plage des taux bruts en deux : les plages  $[x_0; x_1]$  et  $[x_1; x_2]$ , où par exemple  $[x_1; x_2]$  va constituer la plages des valeurs pour les âges élevés. Comme il s'agit de splines cubiques, on a deux polynômes de degré 3,  $p_0(x)$  et  $p_1(x)$  modélisant les taux bruts respectivement pour les plages  $[x_0; x_1]$  et  $[x_1; x_2]$ . On impose les contraintes aux bords suivantes ;  $p_0(x_1)$  doit être égal à  $p_1(x_1)$  et leurs dérivées premières et secondes en  $x_1$  doivent aussi être égales. On pourra utiliser une méthode des moindres carrés pour déterminer ces contraintes, ce qui nous donnera l'expression complète des polynômes et donc les taux bruts lissés.

La méthode de Whittaker-Henderson est un exemple de lissage non paramétrique. On va poser 2 critères : un critère de régularité  $R$  et un critère de fidélité  $F$ . En effet, on a vu précédemment que les taux lissés devaient satisfaire ces 2 critères. Cette méthode attribue un poids ( $a$  et  $b$  constantes réelles) à chacun de ces critères, puis pose une quantité égale à  $aF + bR$  que l'on minimise. On obtient ainsi les taux lissés grâce à un calcul matriciel.

Récapitulons les étapes décrites dans ce cours afin d'affiner au maximum le calcul des taux de mortalité. Tout d'abord, on détermine nos taux bruts à partir des bases présentes dans notre compagnie d'assurance, il s'agit de nos bases mouvement et contrat. Ensuite, on peut faire un lissage de ces taux bruts grâce à des méthodes paramétriques (exemple : lissage par splines cubiques) ou

non paramétriques (exemple : la méthode Whittaker-Henderson). Puis, on ajuste un modèle comme celui de Lee-Carter sur ces taux lissés. Et, pour finir, si on a une trop forte variabilité pour les âges élevés, on peut supprimer les taux ajustés au-delà d'un certain âge (par exemple 80ans) pour les extrapoler (méthode de Coale et Kisker), ou ajuster un autre modèle (par exemple Brass) grâce à la méthode d'agrégation de modèles.

## Annexe A

### Code R du chapitre 5

```
#Importation
library(readr)
Base_Contrats=read_delim("~/Cours/M1/BE/BDD/Base_Contrats.csv",";",
                          escape_double = FALSE,
                          trim_ws = TRUE)
Base_Mouvements=read_delim("~/Cours/M1/BE/BDD/Base_Mouvements.csv",
                             ";", escape_double = FALSE,
                             trim_ws = TRUE)

#Base de données qui contient les dates de
#rachats totaux et les numéros de contrat des assurés
Cnt=Base_Mouvements$Cnt [Base_Mouvements$Typo=="RT"]
DatRT=Base_Mouvements$DatEffet [Base_Mouvements$Typo=="RT"]
RachatTotal=data.frame(Cnt,DatRT)

#Base de données qui contient les dates de naissance,
#date d'effet, date de décès et les numéros
#de contrat des assurés
ContratDC=data.frame(Cnt=Base_Contrats$Cnt,
                     DatNaiss=Base_Contrats$DatNaiss,
                     DatDC=Base_Contrats$DatDC,
                     DatCnt=Base_Contrats$DatEffet)
```

```

#Fusion des deux tables en utilisant le numéro
#de contrat en tant que clé
Base=merge(ContratDC,RachatTotal,by="Cnt",all=T)

#Différence en nombre de jours entre une date donnée
#et une date de référence
diffjour=function(date){
  return(as.numeric(julian(strptime(date,
                                format="%d/%m/%Y"))))
}

#Renvoie l'année associée à la date
annee=function(date){
  unclass(strptime(date, format="%d/%m/%Y"))$year+1900
}
annee("01/11/2016")

#Transforme les valeurs manquantes en 0
NAen0=function(x){
  x[is.na(x)]=0
  return(x)
}

#On évalue l'âge, l'exposition et le décès ou non de chaque
#individu afin de calculer les Dx et Lx pour chaque individu
Ratio_Individu=function(Base,an){
  age=an-annee(Base$DatNaiss)-1
  #âge au 01/01/an
  DC=is.element(annee(Base$DatDC)==an,TRUE)
  #Décès ou non pendant l'année l'année an
  Mort=is.element(annee(Base$DatDC)<an,TRUE)
  #Mort ou non avant l'année an

```



```

Rachat=is.element(annee(Base$DatRT)<an, TRUE)
#Rachat totale ou non avant l'année an
Rachat_an=is.element(annee(Base$DatRT)==an, TRUE)
#Rachat totale ou non pendant l'année an
Contrat=annee(Base$DatCnt)<an
#Contrat souscrit ou non avant l'année an
Souscription=annee(Base$DatCnt)==an
#Contrat souscrit ou non pendant l'année an
C=Contrat&(!Mort)&(!Rachat)
#Contrat actif pendant toute l'année an
CRT=Contrat&Rachat_an&(!Mort)
#Contrat actif au début de l'année an puis racheté
#pendant l'année an
CDC=Contrat&DC&(!Rachat)
#Contrat actif au début de l'année an puis décès
#pendant l'année an
S=Souscription&(!Mort)&(!Rachat)
#Contrat souscrit pendant l'année an
SRT=Souscription&Rachat_an&(!Mort)
#Contrat souscrit puis décès pendant l'année an
SDC=Souscription&DC&(!Rachat)
#Contrat souscrit puis décès pendant l'année an

JRT=NAen0( diffjour( Base$DatRT)-
            diffjour(paste("01/01",
                           as.character(an), sep="/")) )
#Nombre de jours entre la date de rachat
#et le premier jour de l'année
JDC=NAen0( diffjour(Base$DatDC)-
            diffjour(paste("01/01",
                           as.character(an), sep="/")) )
#Nombre de jours entre la date de rachat

```

```

#et le premier jour de l'année
JS=NAen0( diffjour(paste("01/01",
                          as.character(an+1),sep="/")) -
          diffjour(Base$DatCnt) )
#Nombre de j entre le dernier jour de l'année
#et la date de souscription
JSRT=NAen0( diffjour(Base$DatRT) - diffjour(Base$DatCnt) )
#Nombre de jour entre la date de rachat
#et la date de souscription
JSDC=NAen0( diffjour(Base$DatDC) - diffjour(Base$DatCnt) )
#Nombre de jour entre la date
#de décès et la date de souscription
Ratio=C+(CRT*JRT+CDC*JDC+S*JS+SRT*JSRT+SDC*JSDC)/365
#Exposition par individu
return(cbind(age,Ratio,DC))
}

Ratio_test=Ratio_Individu(Base,2006)

#Calcul des lx : effectifs exposés par âge
lx=function(Matrice_Ratio,ageStart,ageEnd){
  lx0=aggregate(Ratio ~ age,data=Matrice_Ratio,FUN = sum)
  age=ageStart:ageEnd
  lx0=lx0[ageStart<=lx0$age & ageEnd>=lx0$age,]
  names(lx0)=c("age","lx")
  return(lx0)
}

lx_test=lx(Ratio_test,0,80)
plot(lx_test, type='l')

```

```

#Calcul des dx : nombre de morts par âge
dx=function(Matrice_Ratio,ageStart,ageEnd){
  dx0=aggregate(DC~ age, data=Matrice_Ratio, FUN = sum)
  age=ageStart:ageEnd
  dx0=dx0[ageStart<=dx0$age & ageEnd>=dx0$age,]
  names(dx0)=c("age", "dx")
  return(dx0)
}

dx_test=dx(Ratio_test,0,80)
plot(dx_test,type='l')

#Calcul des coefficients de mortalité
qx=function(dxn,lxn,ageStart,ageEnd){
  qx=dxn$dx/lxn$lx
  return(data.frame(ageStart:ageEnd,qx))
}

plot(qx(dx_test,lx_test,0,80),type='l')

Matrice_qx=function(Base,yearStart,yearEnd){
  Matrice=NULL
  for (n in yearStart:yearEnd){
    expo=Ratio_Individu(Base,n)
    lxn=lx(expo,0,80)
    dxn=dx(expo,0,80)
    qxn=qx(dxn,lxn,0,80)
    Matrice=cbind(Matrice,qxn$qx)
  }
  return(Matrice)
}

```

## Annexe B

# Code R du chapitre 7

```
library(readr)
#Importation des taux instantanés de mortalité
muHMD=read_delim("~/Cours/M1/BE/BDD//Mx_1x1.csv", " ",
                 escape_double = FALSE, trim_ws = TRUE)

#Transformation des données HMD afin d'obtenir une matrice
Matrice_HMD=function(baseHMD,yearStart,yearEnd
                    ,ageStart,ageEnd,Sexe){
  Age=unique(baseHMD$Age)
  Year=unique(baseHMD$Year)
  baseHMD$Age[baseHMD$Age=="110+"] = 110
  baseHMD$Age=as.numeric(baseHMD$Age)
  if (Sexe=="Mixte"){
    x=baseHMD[c(1,2,5)]
  }else{
    if (Sexe=="Female"){
      x=baseHMD[c(1,2,3)]
    }else{
      if(Sexe=="Male"){
        x=baseHMD[c(1,2,4)]
      }
    }
  }
}
```

```

}
Matrice=matrix(as.numeric(as.matrix(x[,3])),
               length(Age),length(Year))
Matrice=Matrice[ageStart:ageEnd+1,yearStart:yearEnd-1815]
return(Matrice)
}

#Fonction donnant les paramètres du modèle
#de Lee Carter à partir de la matrice des
#quotients de mortalité
Lc=function(Matrice_mu){
  Alpha=apply(Matrice_mu,1,mean)
  M2=Matrice_mu-Alpha
  Beta=svd(M2)$u[,1]/sum(svd(M2)$u[,1])
  Kappa=svd(M2)$d[1]*sum(svd(M2)$u[,1])*svd(M2)$v[,1]
  parametre=list(Alpha,Beta,Kappa)
  names(parametre)=c("Alpha","Beta","Kappa")
  return(parametre)
}

#Lee-Carter pour des données HMD
Lc_HMD=function(baseHMD,yearStart,yearEnd
                ,ageStart,ageEnd,Sexe){
  return(Lc(log(Matrice_HMD(baseHMD,yearStart,yearEnd
                            ,ageStart,ageEnd,Sexe))))
}

#Représentation graphique des 3 paramètres de Lee-Carter
plot_Lc_HMD=function(baseHMD,yearStart,yearEnd,
                     ageStart,ageEnd,Sexe){
  par(mfrow=c(1,3))

```

```

par=Lc_HMD(baseHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)
Annee=yearStart:yearEnd
Age=ageStart:ageEnd
plot(Age,par$Alpha,type='l')
plot(Age,par$Beta,type='l')
plot(Annee,par$Kappa,type='l')
}

#Application pour les années allant de 1950 à 2014
#et les âges de 0 à 80 ans
yearStart=1950
yearEnd=2014
ageStart=0
ageEnd=80
Sexe="Mixte"

Lc_HMD(muHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)

plot_Lc_HMD(muHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)

#Fermeture de la fenêtre graphique
dev.off()

#Représentation en 3 dimensions pour les années 2000 à 2014
#et les âges de 0 à 80 ans.
yearStart=2000
yearEnd=2014
ageStart=0
ageEnd=80
Sexe="Mixte"

library(plot3D)

```

```

#Matrice de l'estimation des taux instantanés de mortalité
#en France pour les données HMD
Memp=Matrice_HMD(muHMD,yearStart,yearEnd
                 ,ageStart,ageEnd,Sexe)
Age=ageStart:ageEnd
Annee=yearStart:yearEnd
persp3D(Age,Annee,log(Memp),theta=-20,phi=20,xlab='âge'
        ,ylab='année',zlab='ln(mu)'
        ,ticktype = "detailed",border = "black")

#Matrice de l'estimation des taux instantanés
#de mortalité par Lee-Carter
par=Lc_HMD(muHMD,yearStart,yearEnd
           ,ageStart,ageEnd,Sexe)
Mmodele=exp(par$Alpha+par$Beta%*%t(par$Kappa))
persp3D(Age,Annee,log(Mmodele),theta=-20,phi=20
        ,xlab='âge',ylab='année',zlab='ln(mu)'
        ,ticktype = "detailed",border = "black",title="")

#Projection du paramètre Kappa
#Avec une régression linéaire
yearStart=1950
yearEnd=2014
ageStart=0
ageEnd=80
yearpred=2064

par=Lc_HMD(muHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)

```

```

z=data.frame(Annee=yearStart:yearEnd,Kappa=par$Kappa)
plot(z,xlim=c(yearStart,yearpred),ylim=c(-200,150))
reglin=lm(Kappa~Annee,data=z)
abline(reglin$coefficients[1],reglin$coefficients[2],
       col='red')
Anneepred=data.frame(Annee=yearStart:yearpred)

#Intervalle de prédiction à 95%
predlin=predict(reglin,Anneepred,interval = "prediction")
lines(Anneepred$Annee,predlin[,2],col='green')
lines(Anneepred$Annee,predlin[,3],col='green')

#Avec une marche aléatoire avec drift
library(forecast)
ARIMA=Arima(par$Kappa,order=c(0,1,0),include.drift=TRUE)
plot(forecast(ARIMA,h=yearpred-yearEnd),type="p")

#Résidus dans le modèle de Lee-Carter
yearStart=1950
yearEnd=2014
ageStart=0
ageEnd=90
Age=ageStart:ageEnd
Annee=yearStart:yearEnd

Memp=Matrice_HMD(muHMD,yearStart,yearEnd,
                 ageStart,ageEnd,Sexe)
par=Lc_HMD(muHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)

```



```

Mmodele=exp(par$Alpha+par$Beta%**t(par$Kappa))
Residus=log(Memp)-log(Mmodele)

#Résidus entre 1950 (en bleu) et 2014(en rouge)
couleur=colorRampPalette(c("blue", "darkred"))
plot(rep(Age,length(Annee)),Residus,
      col=couleur(length(Age))[rep(Age-ageStart+1
                                   ,each=length(Annee))]]
      ,xlab="Age")

#Résidus en 1950, 1982 et 2014
plot(Age,Residus[,1],xlab="Age",ylab="Residus"
      ,ylim=c(-0.5,0.5),xlim=c(-25,90))
points(Age,Residus[,floor((yearEnd-yearStart)/2)+1],
        col='blue')
points(Age,Residus[,yearEnd-yearStart+1],col='red')
legend("left",pch=19,legend=c(yearStart
                              ,yearStart+
                              floor((yearEnd-yearStart)/2)
                              ,yearEnd),
        col=c("black","blue","red"))

# Modèle log-Poisson bilinéaire
# Importation des nombres de décès et expositions
l=read_delim("~/Cours/M1/BE/BDD//Mx_1x1.csv"
             ," ", escape_double = FALSE, trim_ws = TRUE)
d=read_delim("~/Cours/M1/BE/BDD//Mx_1x1.csv"
             ," ", escape_double = FALSE, trim_ws = TRUE)

```

```

library(StMoMo)
yearStart=1950
yearEnd=2014
ageStart=0
ageEnd=80

#Application du modèle
Lcmodel=lc(link="log")
LCfit = fit(Lcmodel,Dxt = Matrice_HMD(d,yearStart,yearEnd
                                     ,ageStart,ageEnd,
                                     Sexe) ,
           Ext = Matrice_HMD(1,yearStart,yearEnd
                              ,ageStart,ageEnd,Sexe) ,
           ages = ageStart:ageEnd,
           years = yearStart:yearEnd,
           ages.fit = ageStart:ageEnd)
par=Lc_HMD(muHMD,yearStart,yearEnd,ageStart,ageEnd,Sexe)
Annee=yearStart:yearEnd
Age=ageStart:ageEnd

#Comparaison des paramètres avec ceux de Lee-Carter
plot(Age,par$Alpha,type='l')
lines(Age,LCfit$ax,type='l',col='red')
plot(Age,par$Beta,type='l')
lines(Age,LCfit$bx,type='l',col='red')
plot(Annee,par$Kappa,type='l')
lines(Annee,LCfit$kt[1,],type='l',col='red')

#Comparaison des logarithmes
#des taux instantanés de mortalité
Memp=Matrice_HMD(muHMD,yearStart,yearEnd,
                 ageStart,ageEnd,Sexe)

```

```
Mmodele=exp(LCfit$ax+LCfit$bx**t(LCfit$kt[1,]))
Mmodele2=exp(par$Alpha+par$Beta**t(par$Kappa))
plot(Age,log(Memp[,1]),type='l')
lines(Age,log(Mmodele[,1]),col='red')
lines(Age,log(Mmodele2[,1]),col='green')
```

## Sources et liens externes

- [1] Données française : Human mortality database. <http://www.mortality.org/cgi-bin/hmd/country.php?cntr=FRA&level=2>.
- [2] The lee-carter model. <http://data.princeton.edu/eco572/LeeCarter.pdf>.
- [3] Site internet de trello. <https://trello.com>.
- [4] Christophe Dutang Arthur Charpentier. L'actuariat avec r. [https://cran.r-project.org/doc/contrib/Charpentier\\_Dutang\\_actuariat\\_avec\\_R.pdf](https://cran.r-project.org/doc/contrib/Charpentier_Dutang_actuariat_avec_R.pdf).
- [5] PLANCHET F. Modèles de durée - applications actuarielles. <http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/a5ee867f10dc2d79c1256f130065f05f/893a6130916e6490c1256f130067c453>.
- [6] Jean-David Fermanian. The lee-carter model. [http://www.crest.fr/ckfinder/userfiles/files/Pageperso/fermania/JDF\\_duree3.pdf](http://www.crest.fr/ckfinder/userfiles/files/Pageperso/fermania/JDF_duree3.pdf).
- [7] Candy N. WENDJA. Construction d'une table de mortalité prospective à partir d'un échantillon réduit. [http://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/0/ff3077c1f701b9b0c1257705004878f7/\\$FILE/Namtchueng.pdf](http://www.ressources-actuarielles.net/EXT/ISFA/1226-02.nsf/0/ff3077c1f701b9b0c1257705004878f7/$FILE/Namtchueng.pdf).