Modélisation de la mortalité par cause de décès

Samuel Piveteau

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de :

l'Université Claude Bernard Lyon 1

Ecole Doctorale 486
Sciences Économiques et de Gestion

Spécialité de doctorat : Sciences de Gestion

Soutenue publiquement le 12/07/2021, par :

Samuel Piveteau

Modélisation de la mortalité par cause de décès

Devant le jury composé de :

PLANCHET Frédéric	**Président**
Professeur des Universités, Université Lyon 1	
CAIRNS J.G. Andrew	**Rapporteur**
Professeur, Université Heriot-Watt Edinbourg (Royaume-Uni)	
DEBON AUCEJO Ana Maria	**Rapporteure**
Professeure, Université Polytechnique de Valence (Espagne)	
ARNOLD Séverine	**Examinatrice**
Professeure, Université de Lausanne (Suisse)	
KALLESTRUP-LAMB Malene	**Examinatrice**
Professeure Associée, Université Aarhus (Danemark)	
ROBERT Christian	**Directeur de thèse**
Professeur des Universités, Université Lyon 1	
TRUFIN Julien	**Co-directeur de thèse**
Professeur, Université Libre de Bruxelles (Belgique)	
TOMAS Julien	**Invité**
Actuaire, encadrant de thèse en entreprise chez SCOR	

Résumé

Au cours des deux derniers siècles, l'espérance de vie mondiale à fortement augmenté dans la plupart des pays. Cette baisse de la mortalité a transformé en profondeur la démographie et les relations sociales dans de nombreuses sociétés et s'est accompagnée de défis de taille comme l'accroissement des dépenses de santé, la question de la redistribution intergénérationnelle et depuis peu la gestion du phénomène de dépendance chez les personnes âgées. La compréhension du vieillissement des populations passe par la modélisation de la durée de vie humaine. Nous proposons dans cette thèse un travail sur la mortalité par cause de décès. L'étude de la mortalité discriminée selon les causes de décès présente de nombreux défis spécifiques par rapport à l'étude de la mortalité générale. Récemment, la mortalité par cause est devenue un sujet plus important au sein des entreprises d'assurance, en ce qu'elle permet d'introduire des expertises médicales fournies par les divers médecins spécialistes. Ce travail a été réalisé au cours des quatre dernières années au sein de l'entreprise de réassurance SCOR et répond systématiquement à des préoccupations pratiques rencontrées au quotidien par les équipes techniques actuarielles. Nous traitons par la suite de trois aspects liés à la modélisation de la mortalité par cause. Ces sujets sont abordés sous forme d'articles de recherche. La première étude porte sur l'extrapolation de la mortalité par cause aux âges avancés. Il s'agit d'un sujet crucial tant le manque de données pour certaines causes de décès se fait sentir pour les âges extrêmes. Afin de fournir des avis d'experts ou de projeter la mortalité par cause, des tables de mortalité exhaustives pour chacune des causes se révèlent une nécessité. La deuxième étude traite de la construction de groupes à partir d'une base de données de mortalité temporelle. Les séries de mortalité sont caractérisées par la cause du décès et l'âge, pour ensuite être regroupées au sein de groupes homogènes qui se distinguent entre eux par leur dynamique spécifique. Le troisième article traite de la projection de la mortalité par cause dans le temps. La méthode proposée permet d'inclure des dynamiques souples et des relations de dépendance temporelle entre causes de décès. L'ensemble de ces articles s'articulent les uns avec les autres et permettent d'aborder l'étude de la mortalité par cause sous un angle nouveau.

Summary

Over the last two centuries, global life expectancy has increased significantly in most countries. This decline in mortality has profoundly transformed demographics and social relations in many societies and has been accompanied by major challenges such as the increase in health care expenditures, the question of intergenerational redistribution and, more recently, the management of the long term care for the elderly. The understanding of population aging requires the modeling of human lifespan. In this thesis, we propose a work on mortality by cause of death. The study of cause-specific mortality presents many challenges that are not involved in the study of all-causes mortality. Recently, cause-specific mortality has become a more important topic within insurance companies, in that it allows the introduction of medical expertise provided by various medical specialists.
This work has been carried out over the last four years within the reinsurance company SCOR and systematically responds to practical concerns encountered on a daily basis by the actuarial technical teams. We discuss three aspects related to the modeling of mortality by cause. These topics are treated in the form of research papers. The first study deals with the extrapolation of cause-specific mortality to older ages. This is a crucial topic due to the lack of data for some causes of death at older ages. In order to provide expert advices or to project cause-specific mortality, comprehensive life tables for each cause are needed. The second study concerns the construction of clusters from a time series mortality database. Mortality series are characterized by cause of death and age and then clustered into homogeneous groups that are distinguished from each other by their specific dynamics. The third article discusses the projection of cause-specific mortality over time. The proposed method allows for the inclusion of flexible dynamics and time-dependent relationships between causes of death. All of these articles are interrelated and provide a new approach to the study of cause-specific mortality.

Remerciements

Ce travail de thèse n'aurait jamais pu se concrétiser sans la participation et l'encadrement d'un certain nombre de personnes. Je tiens ainsi à remercier tout spécialement mes directeurs de thèse Christian Robert et Julien Trufin pour leur patience face à mon attitude pas toujours facile, leurs conseils avisés tant sur un plan technique qu'organisationnel et le temps accordé pour de longues sessions de relecture. J'adresse une pensée reconnaissante aux professeurs Andrew Cairns, Ana Maria Debon Aucejo, Séverine Arnold, Malene Kallestrup-Lamb et Frédéric Planchet, qui ont bien voulu faire partie du jury de thèse. Je remercie également l'entreprise SCOR et ses équipes pour m'avoir donné l'occasion de produire ce travail. J'ai une pensée toute particulière pour Julien Tomas, qui m'a encadré au sein de l'entreprise et que je remercie pour tous les moments de travail et rire passés ensemble. Je tiens également à adresser mes remerciements à quelques-uns collègues de SCOR qui m'ont beaucoup apporté et avec lesquels j'ai pu nouer de véritables liens d'amitié : Agné Ulcinaite, Jonathan Bastien, Nesrine Mtibaa, Tiziana Torri et Razvan Ionescu. Je tiens enfin à adresser mes remerciements à celle qui m'a accompagné dans l'ombre durant tout ce travail, supportant tant bien que mal mes humeurs fluctuantes et qui a toujours tâché de rendre aussi légers que possible les moments difficiles, Sara.

Table des matières

Chapitre 1

Introduction

1.1 Les risques liés à la durée de vie

Les risques afférents à la durée de vie humaine peuvent être de plusieurs natures : le risque de longévité, le risque de mortalité, le risque sanitaire et le risque de dépendance. Par risque de longévité, nous entendons le risque que la prolongation d'une vie humaine au-delà d'un certain âge aboutisse à l'épuisement des ressources financières. Le risque de mortalité quant à lui se comprend comme le risque économique encouru par les membres d'un ménage à la suite du décès prématuré d'un des principaux pourvoyeurs de revenus. Les risques de santé et de dépendance ne sont pas directement liés à la durée de vie humaine, mais leur probabilité de survenance s'accroît avec le vieillissement. Le risque de santé se réalise lorsque la situation financière des individus ne leur permet pas de faire face à un aléa médical. Enfin le risque de dépendance recouvre les difficultés financières subséquentes à une perte d'autonomie, conduisant par exemple l'individu à intégrer un établissement d'hébergement pour personnes âgées dépendantes (EHPAD).

La couverture des risques de longévité, mortalité, santé et dépendance s'est considérablement développée durant le XXe siècle (voir à ce sujet Day (2009)[34]), pour devenir quasi-systématique dans les pays occidentaux. La participation des états au système d'assurance varie selon les pays, laissant plus ou moins de place au secteur privé. Ainsi en France, le risque de longévité est principalement couvert par le système de retraite piloté par le gouvernement et les partenaires sociaux, tandis que le Royaume-Uni, conformément au système béveridgien, ne propose qu'un filet de survie. C'est pourquoi le marché de l'assurance longévité est bien plus développé au Royaume-Uni qu'en France (voir Rusconi (2008)[120]). À l'élargissement de la couverture de risques s'ajoute le phénomène de vieillissement des

populations dans les pays développés. La Figure 1.1 illustre l'accroissement de la durée de vie humaine au sein des pays de l'OCDE.

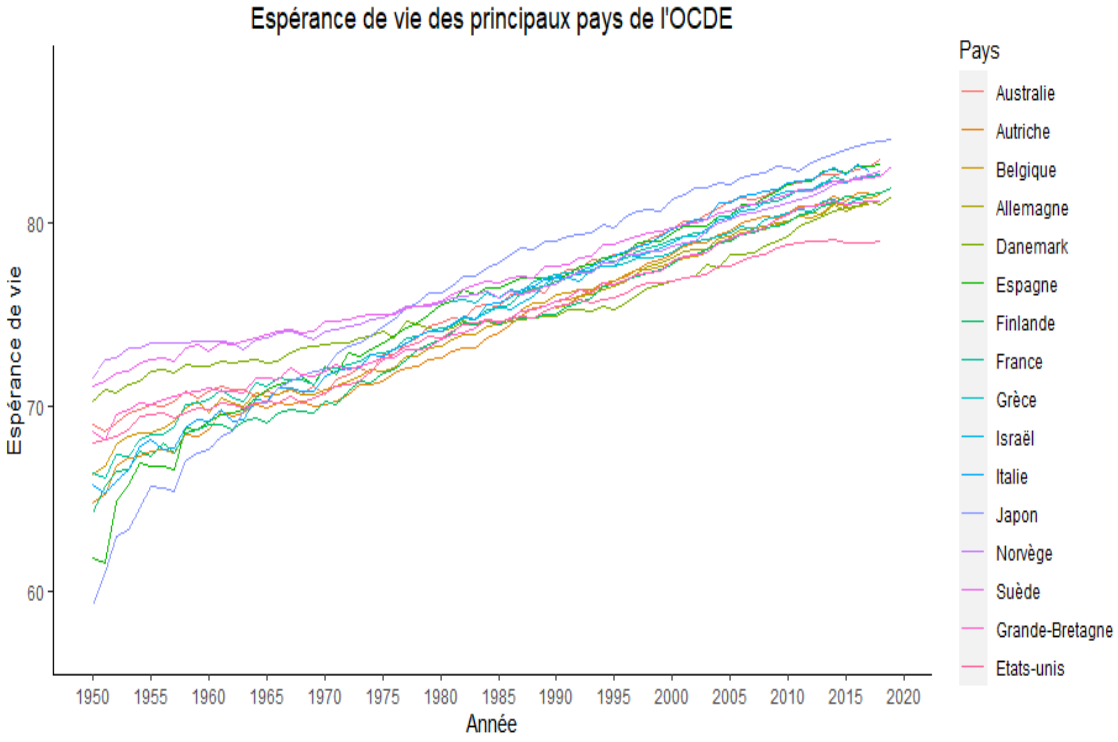


FIGURE 1.1 – *Évolution de l'espérance de vie des principaux pays de l'OCDE de 1950 à 2019 issus des données HMD*

Ces phénomènes de généralisation assurantielle et de vieillissement des populations impliquent des montants considérables pour chacun des risques évoqués. Ainsi rien que sur le marché de l'assurance longévité Michaelson et Mulholland (2014)[100] estiment le montant des actifs à plus de 60 000 milliards de dollars (voir Blake et al. (2018)[15] pour une analyse détaillée concernant les annuités et le rapport de l'OCDE sur la santé (2019)[105] pour un panorama global des dépenses de santé par pays en 2019). Leur gestion représente par conséquent un enjeu économique et social extrêmement important. Ces dépenses d'assurance dépendent de variables aléatoires telles que la durée de vie, la fragilité des personnes, le niveau de croissance économique, lesquelles peuvent s'avérer difficiles à prédire sur le long terme. Afin d'assurer une bonne gestion de ces risques, la modélisation mathématique peut se révéler d'une aide précieuse, simplifiant les aspects trop complexes et permettant de se concentrer sur l'essentiel. Si aujourd'hui l'étude de la mortalité humaine s'accomplit principalement au moyen de modèles, ces dernières années, un regain d'intérêt pour les jugements d'expert a pu être observé. Dans certaines circonstances, la modélisation ne permet pas de prendre en consi-

dération des éléments connus des experts. Il est permis de penser que l'inclusion de ces avis permettrait d'améliorer les prévisions sur court, moyen ou long terme. La modélisation de la mortalité par cause de décès ou selon des caractéristiques individuelles se prêtent de façon adéquate aux jugements d'expert. En ce qui concerne les causes de décès, le personnel médical est souvent considéré comme plus clairvoyant au sujet des innovations sanitaires, et par conséquent plus à même de déterminer une direction pour la mortalité d'une cause particulière. Les assureurs peuvent quant à eux émettre des avis sur les caractéristiques de leurs assurés, prédisant la mortalité à partir de modèles sophistiqués calibrés sur d'autres populations.

Nous proposons dans cette introduction un aperçu des modèles de mortalité utilisés dans les disciplines médicale, actuarielle et démographique. La Section 1.2 offre une perspective historique de la modélisation de la durée de vie humaine pour une population. La Section 1.3 introduit la modélisation de la dynamique de mortalité à travers le temps. Dans la Section 1.4 nous abordons le thème de l'hétérogénéité de la mortalité à travers les facteurs explicatifs et les causes de décès. La Section 1.5 fournit un aperçu des données utilisées pour ce travail de thèse. Enfin la Section 1.6 conclut cette introduction par la présentation des trois principaux chapitres qui composent cette thèse. Le but principale de cette thèse est d'offrir un éventail d'outils de modélisation permettant aux actuaires de traiter de la mortalité par cause de décès. Aussi les applications sont elles toutes tournées dans ce sens. Trois aspects nous ont paru fondamentaux pour mieux appréhender les causes de décès à savoir l'incertitude aux âges avancés, la segmentation du risque au sein d'une base de mortalité comportant plusieurs causes et la projection temporelle de la mortalité par cause de décès. Si ces problématiques peuvent également se poser au sujet de la mortalité générale, certaines particularités émergent dès lors que de multiples causes sont à l'oeuvre que nous tâcherons d'identifier tout au long de ce travail de thèse.

1.2 Les modèles de mortalité

La compréhension des variables relatives au décès suppose la modélisation mathématique de la durée de vie comme variable aléatoire. Les démographes, actuaires, et chercheurs en santé publique se sont beaucoup concentrés sur cette problématique au cours des trois derniers siècles. La première tentative de modélisation de la durée de vie remonte à Graunt (1662)[67]. Ce dernier a estimé la mortalité de la population londonienne au moyen des statistiques publiques. Il est apparu à la lumière de cette étude que la mortalité, malgré son caractère aléa-

toire, présentait une certaine dépendance par rapport à l'âge et au groupe sociale. Halley (1693)[69] a proposé une méthode pour construire des tables de mortalité fiables à partir des statistiques de naissances et de décès. Les travaux de De Moivre (1725)[40] constituent une étape décisive ; pour la première fois, la fonction de survie est modélisée comme une fonction continue par rapport à l'âge. Suite à des observations plus précises sur la durée de vie humaine, Gompertz (1825)[64] a proposé d'utiliser la notion de force de mortalité et de lui attribuer une forme paramétrique.

La force de mortalité à l'âge x est définie comme

$$\mu_x = \lim_{s \to 0_+} \frac{\mathbb{P}(x < X \leq x + s | x < X)}{s}.$$

Il en résulte que la fonction de survie $S(x) = \mathbb{P}(X > x)$ s'écrit comme :

$$S(x) = e^{-\int_0^x \mu_s ds}.$$

Observant la mortalité à l'âge adulte, Gompertz constate que la force de mortalité semble avoir une allure exponentielle par rapport à l'âge. Il propose alors la forme suivante :

$$\mu_x = ae^{bx}.$$

Ce modèle à deux paramètres est repris et amélioré par Makeham (1867)[95], lequel ajoute au modèle de Gompertz, un paramètre modélisant la part de mortalité indépendante de l'âge.

$$\mu_x = ae^{bx} + c.$$

Le modèle du Gompertz-Makeham est toujours grandement utilisé aujourd'hui pour modéliser la mortalité adulte (voir Gavrilov et Gavrilova (2011)[59], Kirkwood (2015)[81]). Ces travaux ont donné une impulsion majeure à la modélisation de la durée de vie au moyen de modèles paramétriques. Ces derniers ont été au fil des années sophistiqués pour inclure davantage d'informations telle la mortalité aux âges jeunes qui semble ne pas suivre la dynamique exponentielle observée à l'âge adulte. Heligman et Pollard (1980)[70] proposent un modèle à huit paramètres à même de capturer la mortalité à tous les âges, les paramètre se divisant en trois sous-groupes ayant respectivement pour fonction de capturer la mortalité aux âges jeunes, à l'âge adulte et aux âges avancés. Thatcher et al. (1998)[124] ont testé plusieurs modèles afin de capturer la mortalité aux âges supérieurs à 80 ans. Le modèle de Kannisto (1994)[77] de forme logistique s'est révélé particulièrement adapté pour capturer la mortalité aux âges très avancés. Il se présente sous la forme suivante :

$$\mu_x(\alpha_1, \alpha_2) = \frac{\alpha_1 exp(\alpha_2 x)}{1 + \alpha_1 exp(\alpha_2 x)}.$$

D'autres méthodes d'estimation de la mortalité ont émergé à la suite des travaux sur les martingales et les processus de comptage. Ces derniers permettent d'estimer de façon non-paramétrique la distribution de mortalité au sein d'une cohorte. Ces travaux ont servi de base aux actuaires et démographes pour l'estimation des forces de mortalité par âge pour une population donnée. Selon les besoins du chercheur ou de l'entreprise, les estimateurs ainsi obtenus sont lissés ou ajustés par rapport à un modèle paramétrique. Deux estimateurs non-paramétriques de la mortalité sont couramment utilisés : Kaplan-Meier (1958)[79] et Nelson-Aalen (d'après les articles Nelson (1969)[103] et Aalen (1978)[1]). L'intérêt de ces estimateurs réside dans leur faculté à capturer la mortalité de façon non-paramétrique, malgré la présence partielle de certains individus au cours de la période d'observation (phénomènes de censure et de troncature). Soient D_x le nombre de décès observés à l'âge x et N_x le nombre de personnes exposées au risque à l'âge x, l'estimateur de Kaplan-Meier s'écrit :

$$\widehat{S}(t) = \prod_{x \leq t} \frac{N_x - D_x}{N_x}.$$

L'estimateur de Nelson-Aalen mesure la force de mortalité cumulée et se calcule de la façon suivante :

$$\widehat{H}(t) = \int_0^t \mu_x dx = \sum_{x \leq t} \frac{D_x}{N_x}.$$

Dans le cas où les observations ne sont possibles qu'aux âges entiers, et sous l'hypothèse de force de mortalité constante entre deux âges entiers, une variante de l'estimateur de Nelson-Aalen est proposée qui aboutit à la force de mortalité suivante :

$$\mu_x = \frac{D_x}{E_x},$$

où E_x désigne l'exposition au risque observée entre les âges x et $x + 1$.

Les méthodes d'estimation et modèles dont nous avons parlé jusqu'à présent ont été élaborés afin d'estimer la mortalité d'une cohorte ou d'une population à une période déterminée. Or la mortalité peut, sous l'influence de certains facteurs, subir des variations. À partir du XIXe siècle, l'espérance de vie s'est mise à augmenter dans les pays occidentaux. Cette dynamique a pour effet de rendre caduque les estimations réalisées sur la base de la mortalité observée. La nécessité d'inclure la

dynamique de mortalité dans le temps devient alors une nécessité si l'on souhaite obtenir des projections fiables de la mortalité future.

1.3 Les modèles de mortalité stochastiques

Les premières prévisions de mortalité à intégrer la dimension temporelle remontent à Cannan (1895)[23], qui étudie la dynamique de mortalité par cohorte. Les approches adoptées pour prévoir les forces de mortalité futures sont restées rudimentaires jusqu'au développement de méthodes statistiques pour les séries temporelles dans les années 1970 Box et Jenkins (1970)[18]. Saboia (1974)[121] est le premier à faire usage de ces méthodes pour la prévision de séries temporelles démographiques. Ces méthodes présentent de nombreux avantages dont le plus important est sans doute le cadre probabiliste dans lequel elles s'insèrent. Ces méthodes ont non seulement permis de quantifier l'espérance de la mortalité future, mais aussi sa distribution, grâce à laquelle la construction d'intervalles de confiance est rendue possible. Jusqu'au début des années 1990, les méthodes de projection de mortalité ne se différenciaient pas des approches adoptées pour les autres séries. Après l'estimation des espérances de vie ou des forces de mortalité sur un historique, les forces de mortalité ou les paramètres du modèle étaient projetées comme des séries temporelles usuelles (voir Pollard (1987) [110]). À partir des années 90, un certain nombre de modèles de mortalité stochastiques sont apparus. L'objectif de ces modèles était de simplifier le processus d'estimation et de fournir des résultats cohérents. En effet les projections des forces de mortalité pour chaque âge pouvaient s'avérer incohérentes du fait de la projection séparée des différents indicateurs. Ces nouvelles approches ont été développées par Alho (1991)[4], Alho et Spencer (1991)[5] ; McNown et Rogers (1989[97], 1992[98]) ; Bell et Monsell (1991)[13], et Lee et Carter (1992)[114]. Parmi ces modèles, le plus connu et utilisé est sans conteste le modèle de Lee-Carter (LC). Nous donnons ici une description détaillée de ce dernier en raison de l'usage intensif que nous en ferons dans la suite de cette thèse. Soient x un âge pouvant aller de 1 à x_{max} et t une année comprise entre 1 à T, soit $\mu_{x,t}$ désignant la force de mortalité à l'âge x et à l'année t, alors le modèle de LC s'écrit :

$$log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t},$$

où α_x est le terme de niveau, κ_t représente la tendance générale de la mortalité, β_x représente la sensibilité de la mortalité par âge à la dynamique de κ_t et $\epsilon_{x,t}$ représente le terme d'erreur.

Dans le modèle initial, les termes d'erreur sont indépendants et identiquement distribués selon une loi Gaussienne, tandis que la dynamique des κ_t est modélisée comme une marche aléatoire avec pas, i.e.

$$\kappa_{t+1} = \kappa_t + \rho + \eta_t,$$

où ρ désigne le pas de la marche aléatoire et η_t les termes d'erreur indépendants et identiquement distribués selon une loi Gaussienne. La calibration du modèle s'effectue en deux temps. En première instance, les paramètres $((\alpha_x)_x, (\beta_x)_x, (\kappa_t)_t)$ sont estimés au moyen d'une minimisation des termes d'erreurs quadratiques. Comme les auteurs l'ont fait remarquer dans l'article pionnier de 1992, l'absence de contraintes sur les paramètres rend le modèle non identifiable. En effet, soit un ensemble de paramètres $((\alpha_x)_x, (\beta_x)_x, (\kappa_t)_t)$ minimisant la moyenne des erreurs quadratiques du modèle de LC et C une constante non nulle, alors les paramètres suivants :

$$\kappa_t^\star = \kappa_t + C,$$
$$\alpha_x^\star = \alpha_x - C\beta_x, \quad \beta_x^\star = \beta_x,$$

minimisent également la somme des erreurs quadratiques. Une contrainte est donc nécessaire sur les paramètres afin d'identifier le modèle. La contrainte la plus courante est $\sum_x \beta_x = 1$ et $\sum_t \kappa_t = 0$. Elle aboutit à une minimisation des erreurs quadratiques par une décomposition en valeurs singulières de la matrice de log-mortalité centrée. Cette contrainte n'est pas la seule à garantir l'identifiabilité du modèle et nous verrons dans les chapitres 3 et 4 de la thèse que des contraintes différentes peuvent s'avérer plus utiles dans des contextes spécifiques. Enfin les paramètres de la dynamique des κ_t sont considérés comme une marche aléatoire avec tendance, dont le paramètre ρ est estimé par $\widehat{\rho} = \frac{\kappa_T - \kappa_1}{T}$.

Le modèle de LC a connu un tel succès que de nombreuses variantes de celui-ci ont été proposées au cours des dernières années. Brouhns et al. (2002)[20] proposent d'intégrer le modèle de LC en considérant les décès comme des réalisations d'une loi de Poisson. Soient $D_{x,t}$ le nombre de décès et $E_{x,t}$ l'exposition au risque pris à l'âge x et à l'année t, le modèle LC-Poisson s'écrit :

$$D_{x,t} \sim \mathscr{P}\big(E_{x,t}\, exp\,(\alpha_x + \beta_x \kappa_t)\big),$$

avec des contraintes sur les paramètres similaires à celles du modèle de LC standard.

Booth et al. (2006)[17] proposent de comparer 5 variantes du modèle de LC

sur 10 pays différents, en sélectionnant de façon optimal la plage temporelle sur laquelle calculer la dynamique du paramètre de temps κ_t. Li et Lee (2005)[89] modifient le modèle de LC afin de prendre en compte de la dynamique de plusieurs populations simultanément. De Jong et Tickle (2006)[36] proposent quant à eux un modèle Éspace-État afin de réduire le nombre de paramètres à estimer, de lisser les résultats et de calibrer simultanément les paramètres structurels et la dynamique du modèle. Renshaw et Haberman(2006)[116] ajoutent un effet cohorte, permettant ainsi d'inclure les comportements générationnels fréquemment observés. Yang et al. (2010)[137] proposent d'exploiter les autres composantes principales dans la décomposition en valeurs singulières afin d'affiner les résultats. D'autres modèles de mortalité stochastiques non directement dérivés du modèle de LC ont également été proposés, notamment le modèle CBD (voir Cairns et al. (2006)[21]). Les modèles les plus utilisés ont été recensés dans Cairns et al. (2009)[22]. Currie (2016)[32] propose un cadre unifié de calibration des modèles de mortalité stochastiques usuels en les intégrant dans un cadre GLM (voir Mccullagh et Nelder (1989)[96]). Ces modèles bien que pertinents pour mesurer et prévoir la mortalité générale ne sont pas toujours adaptés pour prendre en compte des phénomènes d'hétérogénéité dans la mortalité. Nous étudions cet aspect dans la section suivante.

1.4 L'hétérogénéité de mortalité

La thématique d'hétérogénéité de la mortalité recouvre plusieurs sens selon la façon dont on appréhende la mortalité. Il semble important de distinguer deux sens. Lorsque l'on traite des différences de mortalités individuelles, on sous-entend par-là que des variables inégalement distribuées parmi la population génèrent des risques de décès différents. Il a été observé en effet que les probabilités de décès varient considérablement entre les individus. Prendre en considération un ensemble de variables caractéristiques des individus observés, dans l'estimation de la mortalité revient à accomplir une régression ou une classification de la mortalité par rapport aux variables observées.

La seconde distinction qui peut être faite lorsque l'on parle d'hétérogénéité de la mortalité concerne la variabilité des causes entraînant le décès. Si l'on admet une distinction de la mortalité selon les causes de décès, il peut être observé de grandes disparités tant dans les probabilités de décès par âge que dans leur dynamique. La majeure partie de cette thèse étant orientée autour de l'analyse de la mortalité par cause, nous nous arrêtons plus longuement sur cette composante de l'hétérogénéité dans la mortalité.

1.4.1 Hétérogénéité des facteurs de décès

La mortalité peut dépendre d'un certain nombre de variables. Selon l'individu observé, la probabilité de décès n'est pas identique. Plusieurs facteurs peuvent expliquer cela. Les facteurs biologiques semblent être des déterminants majeurs de la longévité d'une grande partie des animaux dont l'être humain. Jazwinski (1996)[74] étudie la relation entre gènes et longévité. Gravina (2009)[68] observe l'impact des facteurs épigénétiques sur la durée de vie. Ainsi, idéalement la séquence ADN des individus pourrait être prise en considération dans la détermination des probabilités de décès. Pour ce faire, un champs théorique appelé la statistique génétique a émergé, proposant des techniques adaptés aux données particulières que constituent les gènes (voir Kempthorne (1953)[80] ou plus récemment Evans et al. (2002)[52]).

D'autres caractéristiques individuelles sont également à prendre en considération lorsque l'on souhaite estimer au mieux la structure de longévité d'un individu. Les caractéristiques peuvent être médicales (opérations passées, maladies, biomarqueurs), socio-économiques (catégorie socioprofessionnel, niveau d'éducation, revenus) ou environnementales (exposition à la pollution, pratique sportive, alimentation). Pour étudier la relation entre la durée de vie et ces variables individuelles, des outils de régression statistique ont été développés. Le plus connu de ces modèles est sans conteste le modèle de régression de Cox (Cox (1974)[30]) aussi connu sous le nom de modèle à risque proportionnel. Zethelius et al. (2008)[138] ont par exemple utilisé avec succès le modèle de Cox pour estimer la mortalité liée aux maladies cardiovasculaires sur la base de marqueurs biométriques tels que l'indice de masse corporel, le taux de cholestérol et la pression artérielle. La montée en puissance de calcul des ordinateurs au cours des récentes décennies a permis l'introduction de méthodes d'apprentissage statistique plus avancées. Gordon et Olshen (1985)[65] ont proposé un modèle d'arbre de régression adapté aux données de survie. La méthode des réseaux de neurones développée par Rosenblatt (1958)[118] a été adaptée aux données de survie par Faraggi et Simon (1995)[53], en remplaçant le régresseur linéaire du modèle de Cox par des réseaux de neurones.

Lorsque les données individuelles ne sont pas disponibles, l'analyse de bases de données agrégées par caractéristiques (économiques, sociales, sexe, géographique, etc) peut être envisagée afin de mieux cerner certaines dynamiques de mortalité. La situation en termes d'espérance de vie à travers le monde est très hétérogène (voir Vaupel (2011)[129]), bien que la situation semble s'être améliorée en faveur

d'une croissance de l'espérance de vie au sein des pays les plus pauvres au cours du XXe siècle (voir Permanyer et al. (2019)[109]). Ce processus d'homogénéisation serait dû en partie à un rattrapage économique des pays pauvres sur les pays riches. Néanmoins, la relation qui lie l'espérance de vie au développement économique est complexe (voir Cervellati et Sunde (2005)[26]) et ne peut se rapporter à une simple dépendance de la durée de vie au PIB. Le développement recouvre en effet plusieurs aspects : biens matériels, éducation, inégalités, pollution, accès aux soins pour n'en citer que quelques-uns. Ainsi les États-Unis, l'une des plus importantes économies au monde, affichait en 2013 une espérance de vie moyenne similaire à celle de Cuba, pays longtemps considéré comme très pauvre (voir Roser et al. (2013)[119]). Les distinctions dans les structures de mortalité se font ici au regard de macro-critères tels que l'appartenance à un pays spécifique, le niveau de richesse ou le développement du dit-pays. Pour ce type d'analyse, les données sont agrégées, ce qui conduit à des estimations et projections de mortalité par groupe. Ainsi Villegas et Haberman (2014)[130] analysent la dynamique de mortalité par catégories socioprofessionnelles à partir d'extensions du modèle de LC multi-population, Alkema et al. (2016)[6] projettent la mortalité par pays et régions à partir de scénarios de mortalité. Cette approche de projection de la mortalité par groupe suppose que les groupes soient déjà constitués, ce qui ne va pas de soi. Le questionnement de Lee (2000)[85] sur la nécessité ou non de distinguer les sexes avant d'appliquer le modèle de LC en est une parfaite illustration. La question du groupement de populations a été traitée dans le cadre des structures et dynamique de mortalité par pays au moyen de méthodes de classification. Ainsi Mesle et Vallin (2002)[99] proposent une classification hiérarchique sur la mortalité des pays européens pour les années allant de 1965 à 1995. Cette analyse a permis de constituer 4 groupes aux structures de mortalité différentes : les pays du bassin méditerranéen, les pays du nord, les pays d'Europe centrale et enfin les anciens pays de l'URSS. Cossman et al. (2007)[29] proposent une classification des comtés des États-Unis, afin d'observer les différences géographiques de mortalité au sein du pays. Les deux travaux précédemment cités proposaient des classifications sur des niveaux de mortalités. Carracedo et al. (2018)[24] proposent une classification des dynamiques de mortalités au sein des pays européens. Si ces classifications permettent d'identifier les pays dont les dynamiques et distributions de mortalité sont analogues, aucune ne s'intéresse spécifiquement à la question du groupement des dynamiques de mortalité par ensemble de caractéristiques autre que l'appartenance géographique. Une telle approche permettrait de constituer des groupes de dynamique cohérente à partir d'une base de mortalité, lesquels pourraient ensuite être extrapolés au moyen de modèles de mortalité stochastique. Le chapitre

3 de cette thèse s'inscrit dans cette démarche de classification. Il propose une méthode de classification à même d'identifier des groupes aux tendances de mortalité similaires du point de vue d'un modèle de LC.

1.4.2 Hétérogénéité des causes de décès

Une cause de décès est la reconnaissance par le personnel médical de la pathologie ayant entraîné la mort. Cette définition bien que simple en apparence soulève plusieurs problèmes. L'arbitraire des nomenclatures médicales en est un. La liste des pathologies évolue dans le temps, certaines émergent tandis que d'autres se scindent en plusieurs sous-pathologies. De plus, la considération que porte le corps médical aux différentes pathologies est sujette à changement. Certaines maladies non identifiées par le passé ont acquis une reconnaissance comme causes de décès au fil du temps, altérant la distribution des décès par âge pour de nombreuses pathologies. Il en résulte une difficulté parfois importante dans l'identification des causes de décès sur de longues périodes. Un autre problème couramment avancé, est celui des causes de décès multiples (voir par exemple Wall et al. (2005)[133]). Dans les certificats de décès, plusieurs causes sont recensées. Le corps médical identifie la cause principale, celle qu'il estime être la cause directe du décès ou avoir généré la chaîne de pathologies ayant entraîné la mort. Il est néanmoins difficile de distinguer clairement une cause principale lorsqu'un individu est sujet à plusieurs maladies potentiellement mortelles, a fortiori lorsque ce sont les effets cumulés de celles-ci qui ont entraîné le décès.

Il est donc plus complexe qu'il n'y parait d'identifier de façon stable dans le temps les causes de décès. Afin que ces difficultés n'empêchent pas l'analyse de la mortalité par cause, l'approche la plus fréquente consiste à considérer des groupes suffisamment larges pour ne pas être sujets à des évolutions artificielles au cours du temps, ainsi qu'à ne pas inclure d'âges trop avancés dans les analyses. En effet, aux âges extrêmes, la présence d'un grand nombre de pathologies chez un même individu rend ardue la tâche de déterminer la cause principale. Le chapitre 2 de cette thèse propose une méthode permettant d'extrapoler la mortalité aux grands âges par cause de décès afin de pallier cet écueil. Malgré ces précautions, il n'en demeure pas moins qu'une part d'erreur réside dans les données relatives aux causes de décès, part avec laquelle il nous faut composer. La modélisation des causes de décès intéresse plusieurs parties : le corps médical, les organismes de santé publique, les assureurs et les démographes. Du point de vue de la modélisation, les causes de décès présentent également un défi intéressant.

Mortalité par cause : des taux brutes à la fonction de survie

L'un des aspects les plus singuliers de la modélisation des décès par cause est la notion de risques concurrents. Le décès dû à une cause donnée empêche de connaître l'âge au décès d'un individu si celui-ci était mort d'une autre cause. Considérons I causes de décès, soit X_i la durée de vie d'un individu avant la survenu du décès par cause i. La variable X_i n'est jamais observée sauf si elle est inférieure à toute variable X_j pour $j \neq i$. Nous observons la durée de vie de l'individu X comme

$$X = \min(X_1, X_2, ..., X_I).$$

Afin de travailler sur la mortalité par cause, certaines hypothèses sont nécessaires quant à la relation qui existe entre ces variables. Chiang (1968)[27] a proposé de considérer les durées comme indépendantes, ce qui simplifie grandement les estimations et s'avère par conséquent très populaire (voir Prentice et al. (1978)[112], Wilmoth (1995)[135] et Putter et al. (2007)[113]).

Notons $\mu_x^{i,n}$ la force de mortalité nette pour la cause i prise à l'âge x.

$$\mu_x^{i,n} = \lim_{dx \to 0_+} \frac{\mathbb{P}(x < X^i \leq x + dx | X^i > x)}{dx}.$$

L'estimation de cette force de mortalité nette est impossible sans hypothèse supplémentaire, et ce en raison de l'impossibilité d'identifier la structure de dépendance en présence de risques compétitifs (voir Tsiatis (1975)[126]). Tout ce qu'il nous est permis de connaître, c'est la force de moralité brute μ_x^i définie comme :

$$\mu_x^i = \lim_{dx \to 0_+} \frac{\mathbb{P}(x < X^i \leq x + dx | X > x)}{dx},$$

dont l'estimation, sous l'hypothèse de force de mortalité constante entre deux âges et d'observation à âge entier, nous est donnée par :

$$\mu_x^i = \frac{D_x^i}{E_x},$$

avec D_x^i le nombre de décès observés pour la cause i à l'âge x et E_x l'exposition définie comme dans la Section 1.2. Sous l'hypothèse d'indépendance des causes, les forces de mortalité brutes et nettes sont égales. D'autres choix de modélisation existent et conduisent à des relations différentes entre forces brutes et nettes. Carriere (1994)[25], Zheng et Klein (1995)[139], Kaishev et al. (2007)[75] et Dimitrova et al. (2013)[44] proposent tous des approches à base de copules, afin de modéliser

la relation de dépendance entre les causes de décès :

$$S(x) = C_\theta \left(S^1(x), S^2(x), ..., S^I(x) \right).$$

La fonction $C_\theta(.)$ fait la jonction entre la fonction de survie totale et les fonctions de survie brutes $S^i(x)$ définies par la formule suivante :

$$S^i(x) = e^{-\int_0^x \mu_s^i ds}.$$

Lorsque l'on fait l'hypothèse d'indépendance entre les causes de décès, la fonction C_θ est implicitement définie comme la fonction produit :

$$\begin{aligned} C_\theta \left(S^1(x), S^2(x), ..., S^I(x) \right) = \prod_{i=1}^{I} S^i(x) &= e^{-\int_0^x \sum_{i=1}^I \mu_s^i ds} \\ &= e^{-\int_0^x \mu_s ds} = S(x), \end{aligned}$$

ce qui conduit bien à $\mu_x = \sum_i \mu_x^i$.

Afin d'inclure la dynamique de mortalité par cause de façon cohérente, Li et Lu (2019)[88] combinent les projections sur la mortalité brute par cause de décès à un modèle de copules archimédiennes. Après avoir estimé les forces de mortalités brutes par cause, ces dernières sont projetées au moyen d'un modèle de Lee-Carter. Les prévisions servent ensuite à reconstruire la mortalité par cause grâce à la fonction de copule. Si l'approche peut paraître séduisante, elle ne dit rien de la meilleure façon de projeter les forces de mortalité causales brutes, étape nécessaire pour conduire cette approche et qui présente certaines difficultés.

Projection des forces de mortalité brutes par cause

Assez tôt, des tentatives de modélisation des causes de décès motivées par un éventuel gain de précision dans les prévisions ont vu le jour (voir Crimmins (1981)[31]). On pensait alors que les prévisions des décès par cause offriraient une perspective plus fine pour l'estimation de la mortalité future. McNown et Rogers (1992)[98] proposent une approche permettant de prédire les taux de décès par cause au moyen de séries temporelles. La projection des causes n'avait apporté aucun gain notable à la projection de la mortalité, par rapport à une approche fondée sur le taux de décès toutes causes confondues. Girosi et al. (2007)[62] appliquent avec succès le modèle de LC séparément sur chacune des causes de décès :

$$log(\mu_{x,t}^i) = \alpha_x^i + \beta_x^i \kappa_t^i + \epsilon_{x,t}^i,$$

où les paramètres sont définis comme dans la Section 1.3 et où i désigne la cause sur laquelle le modèle est appliqué.

Malgré certains succès, le modèle de LC appliqué aux causes de décès présente certains défauts. Wilmoth (1995)[135] démontre que pour certains types de modèles, dont le modèle de LC, les prévisions de la mortalité par agrégation des causes de décès seront toujours plus hautes à long terme que les prévisions fondées sur la mortalité toutes causes. Ce problème d'incohérence entre les prévisions des résultats agrégés et des sous-composantes de la série principale est traité dans la littérature statistique à travers la réconciliation des séries hiérarchiques (voir Hyndman et al. (2011)[73]). Li et al. (2019)[87] proposent d'appliquer les méthodes de réconciliation au problème des causes de décès extrapolées par le modèle de LC. Hélas, l'hypothèse d'absence de biais n'est pas vérifiée dans ce cadre, ce qui peut conduire à une inconsistance par rapport au cadre classique de réconciliation optimale. Nous traitons ce problème dans le chapitre 4 de cette thèse, en introduisant un modèle de LC-Poisson non biaisé. Le chapitre 2 traite également du problème de réconciliation entre l'extrapolation de la mortalité toutes causes aux grands âges et l'extrapolation de la mortalité par cause au moyen d'une approche top-down adapté à la mortalité (pour plus de détails sur les approches top-down, voir Schwarzkopf et al. (1988)[122] et Athanasopoulos et al. (2009)[12]).

La modélisation des causes de décès représente également un défi en termes d'analyse des dépendances temporelles entre les causes, sujet essentiel pour l'étude de l'élimination des causes de décès introduite par Elandt-Johnson (1976)[51], situation hypothétique dans laquelle une ou plusieurs causes de décès viendraient à disparaître. Cette disparition peut impacter les autres causes pour au moins deux raisons : d'une part les autres causes n'étant plus en concurrence avec les causes éliminées vont pouvoir s'exprimer entraînant par la même une augmentation des décès qui leur sont imputables, d'autre part l'élimination des causes peut être due à la disparition ou apparition d'un facteur extérieur influençant également les autres causes. Ainsi la disparition ou diminution d'une cause de décès n'est pas sans conséquences sur les autres causes de décès et a fortiori sur la mortalité générale. Si l'approche par copule évoquée ci-dessus peut constituer une réponse à ce problème, elle ne permet pas nécessairement de prendre en considération l'évolution des décès dans le temps. Dès lors d'autres modèles fondés sur l'observation des dynamiques de mortalité ont été proposés. Afin de modéliser les dépendances par cause, Arnold et Sherris (2013)[10] analysent l'évolution temporelle des forces de mortalité par cause au moyen d'un modèle VECM (Vector Error Correction Model) permettant d'identifier un équilibre de long terme dans la dynamique des

séries temporelles. Plus précisément, si l'on s'intéresse à l'évolution temporelle d'indicateurs de mortalité par cause $(y_t^i)_i$, les auteurs proposent de déterminer des équilibres de long termes tels que :

$$\sum_i b_i y_t^i = z_t,$$

où z_t est une série temporelle stationnaire. L'analyse a été conduite sur un ensemble de pays et le nombre de relations de cointégration est amener à dépendre tant de l'origine géographique que de l'indicateur de mortalité utilisé. Cette approche a été étendue dans Arnold et Sherris (2015)[9] et Arnold et Sherris (2016)[11]. Appliqué aux forces de mortalité par cause, ce modèle permettrait de mieux cerner la dynamique de décès et ainsi d'améliorer les estimations. Deux extensions paraissent pour cela nécessaires : intégrer la cointégration dans des modèles de mortalité usuels et procéder à son estimation en parallèle de l'estimation des paramètres du modèle. Si la première étape reste à définir, la seconde étape pourrait s'inspirer de Wagner (2010)[131].

Dans le même esprit, Lyu et al. (2020)[93] proposent d'appliquer le modèle multi-population de Li et Lee (2005)[89] aux causes de décès afin de prendre en compte la structure de dépendance et assurer la cohérence de long terme des prévisions :

$$log(\mu_{x,t}^i) = \alpha_x^i + \beta_x^i \kappa_t^i + B_x K_t + \epsilon_{x,t}^i,$$

où α_x^i, β_x^i et κ_t^i s'interprètent comme dans le modèle de LC par cause, tandis que B_x et K_t sont des paramètres additionnels communs à toutes les causes. La dynamique de mortalité de K_t est ensuite modélisée comme une marche aléatoire avec drift tandis que les dynamiques des κ_t^i sont modélisées comme un AR(1). Ainsi à long terme, la mortalité par cause est portée majoritairement par les paramètres K_t qui assurent la relation de dépendance entre les causes. Le chapitre 4 de cette thèse propose d'introduire la structure de dépendance temporelle en adaptant le modèle Poisson-LC (voir Brouhns et al. (2002)[20]) dans le cadre d'un modèle Espace-État. La structure de dépendance est intégrée à travers une variable latente multivariée. Exploitant le récent développement des méthodes d'apprentissage statistique, Deprez et al. (2017)[43] ont proposé une approche fondée sur le modèle Regression Tree Boosting pour attester la qualité d'ajustement dans les modèles de mortalité et pour modéliser la mortalité par cause de décès en Suisse.

Alai et al. (2015)[3] modélisent la dynamique de mortalité au moyen d'un mo-

dèle de régression logistique multinomiale.

$$log\left(\frac{q^i_{x,t}}{p_{x,t}}\right) = a^i + b^i_x + c^i_x t,$$

où $q^i_{x,t}$ désigne la probabilité de décès par cause i entre l'âge x et $x+1$ à l'année t sachant qu'on a survécu jusqu'à l'âge x et $p_{x,t}$ la probabilité de survie au-delà de l'âge $x+1$ sachant qu'on a survécu jusqu'à l'âge x. Les paramètres a^i, b^i_x et c^i_x servent à capturer la structure de mortalité par âge ainsi que sa dynamique temporelle. Dans ce modèle, la structure de dépendance temporelle est imposée par la forme multi-logistique qui lie entre elles les causes de décès à âge donné. Kjaergaard et al. (2019)[82] étendent la structure de dépendance à tous les âges en proposant de modéliser les densités de décès $d^i_{x,t} = \mathbb{P}\left(x < X^i \leq x+1\right)$ avec des modèles CoDA (Compositional Data Analysis, voir Aitchison (1986)[2] pour une introduction générale ou Oeppen (2008)[106] et Bergeron et al. (2017)[14] pour une application aux prévisions de mortalité). Parmi les modèles CoDA proposés, le plus efficace sur les populations étudiées est le 2S-CoDA défini de la façon suivante :

$$clr(d^i_{x,t}) = \alpha^i_x + \beta^i_x \kappa^i_t + B^i_x K_t + \epsilon^i_{x,t}, \tag{4.1}$$

où K_t permet de mesurer la dynamique commune des causes tandis que κ^i_t capture la mortalité spécifique. La fonction clr désigne la transformation centred log-ratio sur les données compositionnelles. On remarque là encore que la dépendance des structures de mortalité est permise par l'association d'un modèle de mortalité à une transformation de certains coefficients de mortalité. Nous montrons dans le chapitre 2 qu'une relation existe entre les modèles CoDA et le modèle multinomial, conduisant à définir des variantes du modèle multinomial usuel. Nous appliquons ces variantes à l'extrapolation de la mortalité par causes aux âges avancés.

Finalement, Redondo Lourés et Cairns (2021)[115] ont proposé d'inclure des effets cohortes à la mortalité par cause afin d'affiner les prévisions. Ils ont notamment identifié des effets cohortes plus marqués pour les causes de décès à forte dépendance comportementale, comme les causes liées au tabagisme.

1.5 Données

1.5.1 Origine des données

La plupart des applications de ce travail de thèse concernent la population féminine américaine. La mesure des risques biométriques de cette population re-

présente un enjeu important pour les compagnies d'assurance, puisque de fortes incertitudes existent tant à propos de la mortalité aux grands âges par cause que des prévisions de cette mortalité future. Nous utilisons également les données de mortalité de la population masculine américaine dans le chapitre 3. Pour ces populations les données de mortalité sont disponibles par cause de décès pour tous les âges jusqu'à 120 ans, bien que nous n'ayons jamais utilisé l'intégralité des âges dans notre travail, et ce pour des raisons de fiabilité évoquées précédemment. La période d'observation des décès court de 1979 à 2017. Nous avons considéré 6 grandes causes de décès, choisies pour leur importance du point de vue médicale.

(1) Les cancers sont des maladies provoquées par prolifération anormale des cellules. Les plus fréquents aux États-Unis sont les cancers du sein, le cancer colorectal, le cancer de la prostate et le cancer des poumons.

(2) Les maladies métaboliques concernent le fonctionnement anormal du métabolisme dû à une altération chimique du métabolisme cellulaire. Parmi ces maladies, la plus fréquente est le diabète.

(3) Les maladies mentales et neurologiques recouvrent deux pathologies distinctes. Les maladies neurologiques ont trait au système nerveux et incluent la maladie d'Alzheimer et la maladie de Parkinson. Les maladies mentales sont des troubles affectant le comportement des individus. Ces maladies ne tuent pas directement, mais peuvent induire des comportements à risques aux conséquences fatales.

(4) Les maladies du système circulatoire regroupent l'ensemble des pathologies affectant le système de circulation sanguine. On y compte par exemple les maladies coronariennes, cérébrovasculaires et les thromboses veineuses.

(5) Les maladies respiratoires affectent l'appareil respiratoire des individus, altérant les échanges gazeux entre l'individu et l'extérieur. Les pneumonies et l'asthme sont les deux maladies respiratoires les plus connues.

(6) Les causes externes regroupent l'ensemble des causes de décès n'ayant pas pour origine une altération de la santé individuelle. Elles incluent, entre autres, les accidents de la route, les suicides non imputables aux maladies mentales, les chutes mortelles et les homicides.

Afin d'être exhaustif dans le décompte des décès, nous incluons une dernière catégorie appelée "Autres causes", regroupant l'ensemble des décès non imputables aux causes évoquées ci-dessus.

Les variables $D^i_{x,t}$, représentant le nombre des décès dus à la cause i à l'âge x et survenus à l'année t, sont issues de la base de données CDC (Centers for Disease Control)[56]. Les variables d'exposition $E_{x,t}$ représentent le temps cumulé

d'observation de la population à l'âge x durant l'année t. Les données d'exposition proviennent de la base HMD[71] (Human Mortality Database). À chaque décès, un médecin enregistre les causes du décès et les classe selon leur ordre d'importance dans la survenue du décès. Le référencement de la pathologie se fait au moyen de la nomenclature International Classification of Diseases (ICD). Cette classification a été amenée à évoluer au cours des années afin de répondre aux avancées médicales. C'est la raison pour laquelle plusieurs ICD sont recensées, chacune propre à une période. Par exemple, la classification ICD 9 fut en usage entre les années 1979 et 1998, avant d'être remplacée par ICD 10. Comme nous considérons de grandes causes de décès, l'effet des changements de nomenclature est mineur, sans toutefois être sans conséquence comme nous le verrons ultérieurement. Nous fournissons dans le tableau 1.1 les codes employés pour construire les 7 causes de décès décrites plus haut.

Cause de décès	1979 à 1998 (ICD 9 : UCOD 282)	1999 à 2017 (ICD 10 : UCOD 358)
Cancer	04600 – 11800	69 - 146
Métabolique	11900 – 13200	156-173
Mental et nerveux	14100 – 15900	174-194
Circulatoire	16000 – 20300	197-246
Respiratoire	20400 – 22700	247-278
Externes	29900 – 35800	381-456
Autres	les autres	les autres

Table 1.1 – *Classification des causes de décès selon les nomenclatures ICD 9 et 10.*

1.5.2 Visualisation des données

Nous fournissons ici deux graphiques permettant d'illustrer quelques faits stylisés concernant la mortalité par cause. La Figure 1.2 illustre l'évolution des taux de décès de la population féminine américaine par groupe d'âges et causes de décès sur la période allant de 1979 à 2017.

Le groupe d'âges 0-25 ans correspond à la période de vie qualifiée de juvénile. Nous observons que la hiérarchie des causes de décès pour cette tranche d'âge est restée relativement stable au cours des dernières décennies. La cause majeure de décès est externe : les accidents de la route sont une source importante de décès des populations jeunes, de même que les suicides, et accidents divers. Suivent les autres causes et les cancers, auxquels on associe souvent les maladies infantiles et cancers juvéniles. Enfin arrivent les maladies circulatoires, mentales et nerveuses,

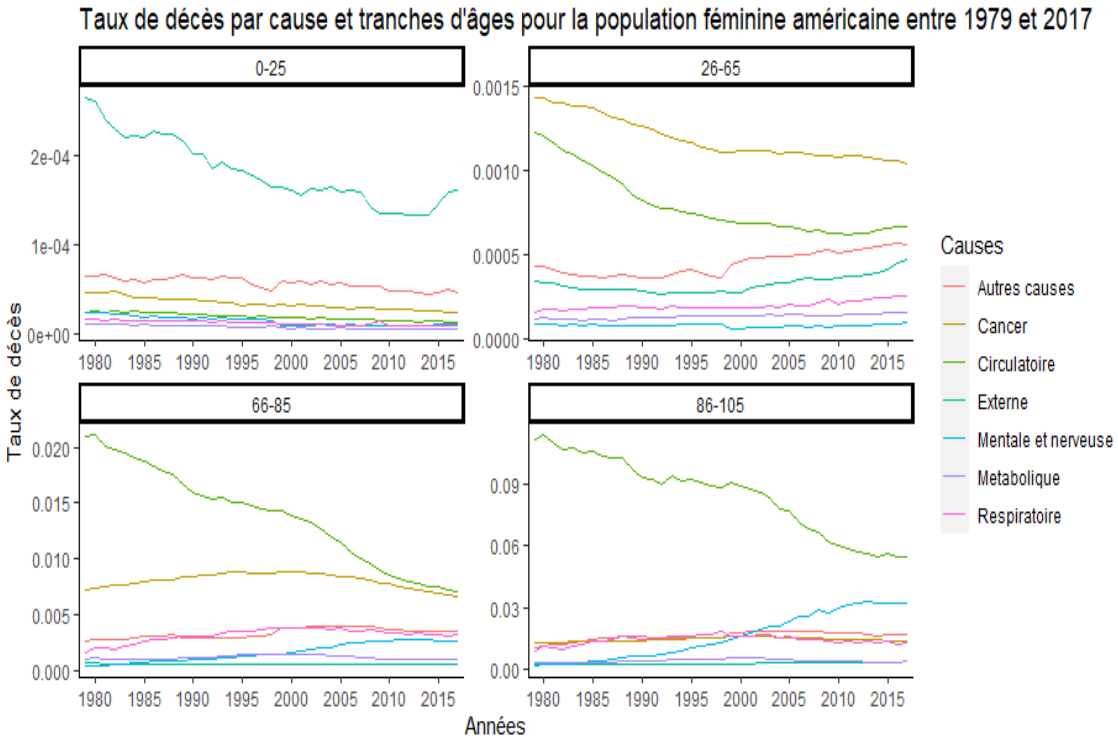


FIGURE 1.2 – *Tendances de mortalité par groupe d'âges et causes de décès*

respiratoires et métaboliques, maladies rarement mortelles ou peu fréquentes pour cette tranche d'âge. Du point de vue de la dynamique, on observe que les causes externes semblent avoir décru de façon importante jusqu'en 2014, avant de connaître une augmentation sur les récentes années. Les autres causes ont soit légèrement diminué, soit stagné. Globalement, la mortalité est faible pour cette période de la vie.

La tranche d'âge 26-65 correspond à la période de vie active des individus. Sur la période d'étude, la hiérarchie des causes de décès est demeurée la même. Les deux principales causes de décès pour cette tranche d'âge sont les cancers et les maladies circulatoires, suivies par les autres causes et les causes externes, encore assez présentes à cette période de la vie, notamment au début. Enfin les maladies respiratoires, métaboliques, mentales et nerveuses sont peu responsables de décès à ces âges. Du point de vue de la dynamique, la période est caractérisée par une baisse des taux de décès par cancers et maladies circulatoires. Cette diminution est, pour chacune des causes, découpée en deux phases, l'une de décroissance rapide, l'autre de décroissance plus faible. Sur les dernières années, une légère hausse des décès imputables aux maladies circulatoires a même été observée. Les autres causes et les causes externes de décès se sont accrus au cours de la période, tandis que les mortalités dues aux causes respiratoires, métaboliques, mentales et nerveuses

semblent avoir stagné ou légèrement augmenté.

La groupe d'âges 66-85 correspond à la première période de retrait de la vie active. Durant la période courant de 1979 à 2017, les principales causes de décès sont les maladies circulatoires et le cancer, suivies des autres causes et maladies respiratoires. Enfin les maladies mentales, nerveuses, métaboliques et externes arrivent en dernier. Sur la dynamique de mortalité, les causes circulatoires ont grandement diminué, rejoignant les cancers, qui après une phase de croissance ont connu après les années 2000 une légère baisse. Les décès par maladies mentales et nerveuses ont augmenté significativement pour cette tranche d'âge, arrivant au niveau des maladies respiratoires et autres causes en 2017.

Enfin le groupe 86-105 correspond aux âges extrêmes. Les maladies circulatoires sont ici encore la principale cause de décès. Viennent ensuite les maladies mentales, nerveuses, autres causes, cancers et maladies respiratoires. Les causes externes et pathologies métaboliques génèrent peu de décès pour ces âges. La dynamique des causes de décès sur cette période est caractérisée par deux mouvements opposés. D'un côté on observe une baisse importante de la mortalité circulatoire, de l'autre la croissance importante des maladies mentales et nerveuses semble compenser en partie cette baisse. Sur les plus récentes années, on constate que ces deux causes stagnent. Les autres causes n'ont pas connu de grands changements.

On peut retenir de cette analyse, qu'au cours des années allant de 1979 à 2017, certaines causes ont semblé plus dynamique que d'autres. C'est le cas notamment des causes externes pour la période juvénile, des maladies circulatoires, des cancers pour les âges 26-65 et des maladies mentales et nerveuses pour les âges allant de 65 à 105 ans. Cependant, ces causes n'ont pas évolué de façon constante dans le temps. Certaines dynamiques se sont substituées aux précédentes, produisant ce qu'il convient d'appeler un changement de tendance. Ces changements de tendance sont l'une des difficultés auxquelles le chercheur est confronté lorsqu'il aborde l'étude des causes de décès. Afin d'intégrer dans notre analyse de la dynamique de mortalité par cause ces changements de tendance, nous proposons dans le chapitre 4 de cette thèse d'inclure la possibilité d'évolution du paramètre de pas du modèle de LC. Cet ajout permet de prendre en considération les récentes évolutions de la tendance pour l'extrapolation ainsi qu'une meilleure estimation du risque d'évolution de la mortalité future.

La Figure 1.3 représente les probabilités de décès par cause et par âge en 2017,

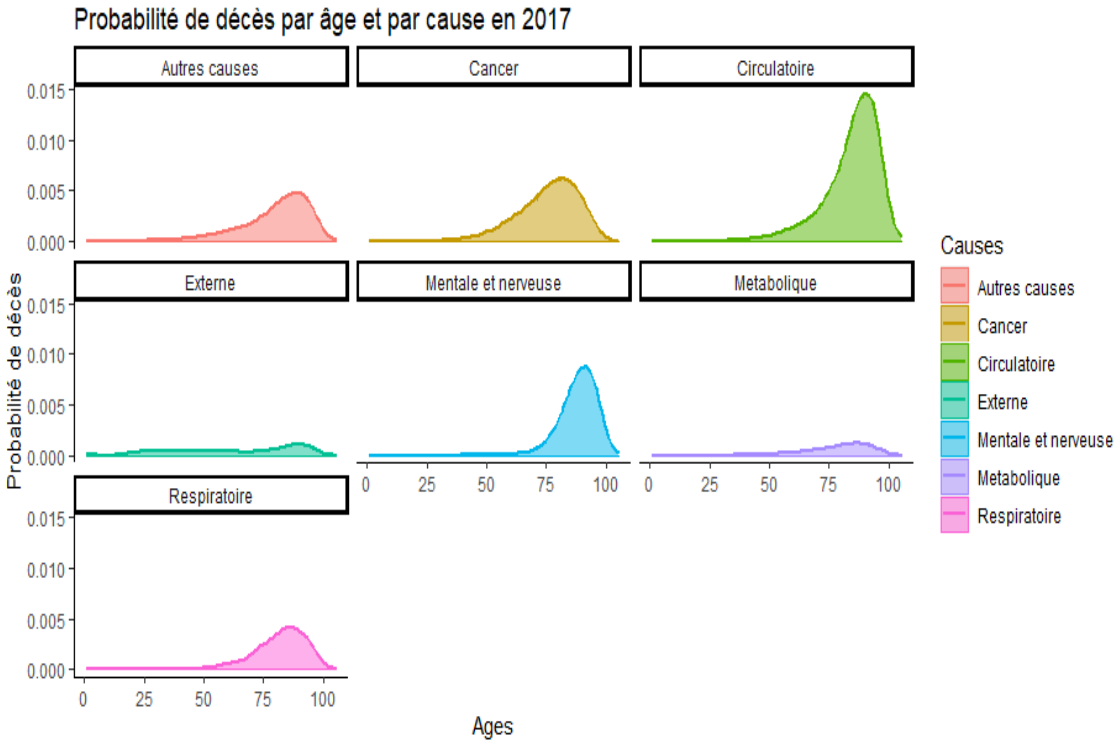


FIGURE 1.3 – *Distribution de la mortalité par âge selon la cause de décès*

sous l'hypothèse d'indépendance des causes de décès. Les courbes ont été lissées afin d'éliminer le caractère erratique des estimations de mortalité aux âges avancés. Nous pouvons constater que la forme des courbes des probabilités est similaire pour 5 des 7 causes étudiées à savoir les maladies circulatoires, les autres causes, les maladies respiratoires, le cancer et les maladies mentales et nerveuses. Pour toutes ces causes, la mortalité croît soudainement après un certain âge, atteignant un pic avant de redescendre. Nous observons cependant que l'âge de croissance de la mortalité ainsi que l'âge auquel la mortalité atteint son maximum diffèrent selon les causes. Ainsi le mode du cancer est à 83 ans tandis que celui des maladies mentales et nerveuses est à 91 ans et celui des maladies circulatoires à 90. La distribution des causes externes et des maladies métaboliques semble plus aplatie. En particulier, les causes externes, prises avant 76 ans semblent presque suivre une loi uniforme, indiquant par là une faible sensibilité de la cause à la variable d'âge.

1.6 Aperçu des chapitres de thèse

1.6.1 Chapitre 2 : Coherent mortality extrapolation by cause of death at high ages

Le deuxième chapitre traite de l'extrapolation de la mortalité par cause de décès aux âges avancés, sujet important comme nous l'avons vu dans la Section 1.4.2 en raison des incertitudes qui pèsent sur les données aux grands âges. L'objectif est d'extrapoler les forces de mortalité par cause μ_x^i aux âges avancés tout en maintenant une cohérence avec les méthodes usuelles d'extrapolation de la mortalité toutes causes aux grands âges. Pour ce faire, nous proposons une approche top-down similaire à celle introduite dans Athanasopoulos et al. (2009)[12] que nous adaptons à la mortalité par cause distribuée comme une variable de Poisson. L'approche top-down est une méthode d'extrapolation des séries temporelles hiérarchiques dans laquelle la série principale est extrapolée puis fragmentée afin de retrouver les sous séries qui la composent. La méthode développée dans Athanasopoulos et al. (2009)[12] consiste à extrapoler la série principale et les contributions des séries qui la composent séparément avant de les recombiner pour obtenir l'intégralité des séries. Ainsi, nous sommes assurés que la somme des séries est égale à la projection de la série principale, chose qui peut s'avérer pertinente si le modèle d'extrapolation de la série principale est jugé important. Dans le cadre de la fermeture des tables de mortalité, certains modèles sont considérés comme des étalons dont il ne faut pas s'éloigner, justifiant ainsi le recours à cette approche top-down pour l'extrapolation de la mortalité par cause aux grands âges. La vraisemblance du modèle de mortalité par cause est scindée en deux parties représentant respectivement la mortalité toutes causes et la contribution des causes de décès à celle-ci.

Nous faisons l'hypothèse que la force de mortalité toutes causes μ_x est une fonction de l'âge x dépendant d'un ensemble de paramètres θ_μ, tandis que la contribution de la cause i à la force de mortalité totale, $S_x^i = \frac{\mu_x^i}{\mu_x}$ dépend d'un autre ensemble de paramètres θ_S. Sous certaines hypothèses, la fonction de vraisemblance s'écrit :

$$L\big(D_x^1, \dots, D_x^I \,;\, \theta_\mu, \theta_S\big) = \prod_x \frac{\big(E_x\ \mu_x(\theta_\mu)\big)^{D_x}}{D_x!} \exp\big(-E_x\ \mu_x(\theta_\mu)\big) \binom{D_x}{D_x^1, \dots, D_x^I} \big(S_x^1(\theta_S)^{D_x^1}, \dots, S_x^I(\theta_S)^{D_x^I}\big).$$

Par conséquent la log-vraisemblance est égale à :

$$l\big(D_x^1,\dots,D_x^I\,;\,\theta_\mu,\theta_S\big) = l_1\big(D_x\,,\,\theta_\mu\big) + l_2\big(D_x^1,\dots,D_x^I\,;\,\theta_S\big) + C,$$

où,

$$l_1\big(D_x\,,\,\theta_\mu\big) = \sum_x D_x\ \log(\mu_x\big(\theta_\mu\big)) - E_x\ \mu_x(\theta_\mu),$$

$$l_2\big(D_x^1,\dots,D_x^I\,;\,\theta_S\big) = \sum_{x,i} D_x^i\ \log(S_x^i\big(\theta_S\big)).$$

L'extrapolation de la mortalité aux âges avancés est obtenue en deux étapes. La première consiste à extrapoler la mortalité toutes causes aux âges avancés en maximisant $l_1\big(D_x\,,\,\theta_\mu\big)$. Cela peut être accompli en utilisant les techniques standards de fermeture de table de mortalité telles que proposées dans Delwarde et Denuit (2005)[39]. Nous avons sélectionné trois approches : le modèle de Gompertz-Makeham[64][95], le modèle de Kannisto (voir Kannisto et al. (1994)[77]) et le modèle Coale-Kisker (1990)[28] revisité sous forme paramétrique ainsi que détaillé dans Flici (2016)[55]. Chacun de ces modèles repose sur des hypothèses de mortalité variées, conduisant à des extrapolations de mortalité sensiblement différentes. La deuxième partie de l'algorithme consiste à extrapoler la contribution des causes de décès à la mortalité générale en maximisant $l_2\big(D_x^1,\dots,D_x^I\,;\,\theta_S\big)$. Dans le cadre qui est le nôtre, une telle extrapolation revient à considérer les décès par cause comme des réalisations de variables multinomiales, dont la somme est égale au nombre de décès toutes causes confondues. Afin d'extrapoler de la façon la plus souple possible, nous proposons d'utiliser un modèle P-splines adapté au modèle multinomial. La méthode des P-splines est adaptée au cadre GLM conformément à Eilers et Marx (1996)[50]. Nous montrons également dans ce chapitre que le modèle GLM multinomial peut être étendu au moyen d'autres fonctions de lien issues de la théorie CoDA (voir Aitchison (1986)[2]). En recombinant les extrapolations de la force de mortalité aux grands âges et des contributions des causes à la mortalité générale, nous obtenons des extrapolations des mortalités par cause cohérentes avec le choix de la méthode de fermeture de table de mortalité. Une telle exigence peut s'avérer nécessaire lorsque la fermeture de la table de mortalité est fondée sur des hypothèses spécifiques que l'on souhaite conserver. Nous proposons une analyse des résultats obtenus sur la population féminine américaine en 2017. Nous constatons que les différents GLMs ont peu d'influence sur le résultat final, contrairement au choix de la méthode de fermeture. Celle-ci impose une cadence de décès plus ou moins rapide à la population, modifiant substantiellement l'importance de certaines causes, sans pour autant bouleverser la hiérarchie des principales causes de décès. Une telle méthode a été développée afin de fournir aux experts médicaux

des bases complètes sur lesquelles appliquer leurs jugements d'experts. Elle peut servir de support à une projection temporelle par cause de décès et par âge telle que proposée dans le chapitre 4.

1.6.2 Chapitre 3 : Mortality clustering with the K-Lee-Carter model

Dans le chapitre 3 de cette thèse, nous proposons un algorithme permettant de constituer des groupes tels que l'ajustement par un modèle de LC sur chaque groupe soit optimal. Le cadre d'étude est le suivant : nous avons à disposition des séries temporelles de forces de mortalité calculées pour un ensemble fini de caractéristiques, par exemple l'âge, la cause de décès ou le pays, et que nous souhaitons ajuster par un modèle de LC. Nous notons $\mu_{a,t}$ la force de mortalité associée à la caractéristique a (par exemple la cause de décès cancer à 60 ans). En raison de la dynamique propre à certains groupes de caractéristiques, il peut être utile de diviser la base de données en plusieurs groupes, sur chacun desquels un modèle de LC est calibré afin d'obtenir un meilleur ajustement et une plus juste appréciation de la dynamique de l'ensemble des séries. Nous désignons par $(C_j)_{j=1,\dots,K}$, les K groupes au sein desquels nous souhaitons partitionner les séries de mortalité $(\mu_{a,t})_t$. Des méthodes de classification de séries temporelles existent déjà dans la littérature statistique, ainsi Paparrizos et Gravano (2015)[107] propose une approche de type K-centroids pour les séries temporelles, permettant de grouper les séries temporelles présentant des formes proches au sein de K groupes différents. Hélas, ce type d'approche n'est pas spécifiquement adaptée aux modèles de mortalité stochastiques, ne fonctionnant que sur des séries standardisées. Pour cette raison, la question demeure ouverte de savoir comment constituer les groupes optimisant l'erreur d'ajustement sur des modèles de LC. Mathématiquement parlant, le problème revient à minimiser la fonction de coût suivante :

$$L\left((C_j)_j, (\beta_a)_a, (\alpha_a)_a, (\kappa_t^j)_{t,j}\right) = \sum_j \sum_{a \in C_j} \sum_t (log(\mu_{a,t}) - \alpha_a - \beta_a \kappa_t^j)^2.$$

avec $(\beta_a)_a$ et $(\alpha_a)_a$ les paramètres du modèle de LC associés à la caractéristique a et κ_t^j le paramètre de dynamique temporelle du groupe j.

Nous proposons un algorithme dérivé des K-centroids (voir Macqueen (1967)[94]) et adapté au modèle de LC que nous nommons les K-LC. Á partir d'un algorithme à l'apparence complexe, nous montrons que, moyennant une modification

des contraintes usuelles sur les paramètres du modèle, la méthode revient à un algorithme des K-centroids avec une fonction de distance particulière :

$$d(x, y) = Var(x)Var(y)(1 - Cor(x, y)).$$

avec x et y deux séries temporelles, Var et Cor les opérateurs de variance et corrélation empiriques.

Nous proposons également deux variantes des K-LC adaptées aux données erratiques : les K-LC contraints, inspirés de Leisch et Grün (2020)[86] et les K-LC pénalisés.

Deux applications sont proposées afin d'illustrer l'algorithme. La première porte sur la division par sexe dans les prédictions de mortalité. Il est usuel (voir Lee (2000)[85]) de diviser la population selon le sexe avant d'appliquer le modèle de LC sur chacune des sous-populations séparément. Nous montrons sur la population américaine entre les années 1979 et 2012 que cette division n'est pas la plus pertinente et qu'un découpage générationnel est plus intéressant pour cette période. La seconde application porte sur le découpage de groupes au sein de la mortalité féminine américaine par âge et causes de décès, pour les mêmes années. Nous montrons que 6 groupes différents mêlant plusieurs causes peuvent être constitués afin d'obtenir un meilleur ajustement. Nous effectuons ensuite des prévisions de mortalité à l'aide d'une marche aléatoire et constatons que les résultats obtenus après découpage par K-LC sont plus proches des réalisations observées sur les années allant de 2013 à 2017 que les résultats obtenus par classification usuelle (sexes séparés et causes séparées). Les deux variantes des K-LC sont également appliquées et permettent d'améliorer les résultats.

1.6.3 Chapitre 4 : A multivariate Poisson state-space model to forecast cause of death mortality

Le chapitre 4 traite des prévisions des taux bruts mortalité par cause. Nous avons proposé un modèle permettant de traiter de trois problèmes fondamentaux dans la projection de la mortalité par cause, à savoir les changements de tendances évoqués dans la section 1.5.2, le problème de la dépendance temporelles des causes de décès évoqués dans la section 1.4.2 et la présence d'un biais dans les prévisions de mortalité. Soit $D^i_{x,t}$ le nombre de décès à l'âge x et à l'année t liés à la cause i, le modèle proposé est le suivant

$$\left(D^i_{x,t}\right)_{x,t,i} \sim \mathscr{P}\left(\lambda^i_{x,t}\right),$$

où $\lambda^i_{x,t} = E_{x,t} exp\left(\alpha^i_x + \beta^i_x \kappa^i_t\right)$ et pour lequel la dynamique s'exprime comme :

$$\kappa_t = \kappa_{t-1} + \rho_{t-1} + \eta_t$$
$$\rho_t = \rho_{t-1} + \zeta_t,$$

où $\kappa_t = \left(\kappa^1_t, \kappa^2_t, ..., \kappa^I_t\right)'$ est le vecteur regroupant le paramètre de dynamique et $\rho_t = \left(\rho^1_t, \rho^2_t, ..., \rho^I_t\right)'$ est le vecteur regroupant les pas de la marche aléatoire des κ^i_t, $\eta_t \sim \mathscr{N}(0, \Sigma_\eta)$ et $\zeta_t \sim \mathscr{N}(0, \Sigma_\zeta)$. De plus $\forall t, s$ les variables η_t, ζ_s sont indépendantes et distribuées selon des lois gaussiennes. Nous montrons que le modèle ainsi défini s'écrit comme un modèle Espace-État non gaussien. Le problème des biais est résolu par l'usage d'un modèle Espace-État dont les variables d'espace ne sont pas les logarithmes des forces de mortalité mais les décès par cause (voir Fung et al. (2017)[58] pour un modèle Espace-État appliqué aux logarithmes des forces de mortalité). Afin de rester dans un cadre connu, nous employons pour chacune des causes le modèle de LC-Poisson tel que présenté dans Brouhns et al. (2002)[20]. Les paramètres temporels du modèle de LC sont considérés comme des variables d'état auxquelles nous attribuons une structure de dépendance multivariée afin d'intégrer la dépendance temporelle entre les causes. Enfin, le problème des changements de tendances des paramètres κ^i_t du modèle est pris en considération par l'ajout d'une dynamique au paramètre de pas ρ^i_t.

La calibration du modèle reprend la méthode fréquentiste proposée par Schön et al. (2011)[123], consistant à estimer les paramètres du modèle au moyen d'un algorithme Expectation-Maximization. Nous adaptons cette méthode au modèle présenté, et montrons que pour un certain nombre de paramètres, le recours à des méthodes d'optimisation numérique n'est pas nécessaire. Une application à la population féminine américaine entre 1979 et 2012 est ensuite proposée. Nous détaillons les structures de dépendance obtenues et nous en mesurons l'impact sur la dépendance entre causes de décès au moyen de simulations. Nous effectuons ensuite des prévisions sur les années 2012 à 2017, que nous comparons avec celles d'un modèle de LC standard appliqué sur chacune des causes séparément. En comparant les résultats issus des modèles aux réalisations de décès observées, nous constatons une très nette supériorité du modèle Espace-État, que nous estimons en grande partie due à sa capacité à prendre en considération les changements de tendances.

Chapitre 2

Coherent mortality extrapolation by cause of death at high ages

Abstract

We propose to introduce a top-down extrapolation method to predict cause of death mortality at advanced ages. The extrapolation process is achieved in two steps. The first one consists in extrapolating the all-cause mortality using standard life table closure methods. In the second step, we project at older ages the contribution by age of each cause to the all-cause mortality. We use for this purpose several variants of the multinomial model with P-splines. The product of the contributions and all-cause mortality results in the extrapolated mortality by cause of death. We apply empirical tests to female causes of death data to understand the impact of each combination of multinomial models and closure methods on the extrapolated trends.

2.1 Introduction

Mortality study is a key topic at the crossroads of several disciplines : demography, actuarial science, gerontology and public health. In order to better understand mortality trends, a current approach consists in decomposing the global mortality regarding the main causes leading to death. Most of the existing work on causes of death focuses either on modelling competing risks or on predicting mortality by cause consistently over time. Arnold and Sherris (2013)[10] study the recent evolution of cause-specific mortality, suggesting a time-series co-integration framework to forecast cause-specific forces of mortality. Alai et al. (2015)[3] examine the cause-specific dynamic using a multi-logistic approach, allowing for a cause of death elimination process. Li et al. (2019)[87] propose an approach to reconcile the prediction of all-cause mortality with the prediction of cause-specific mortality. All these methods propose to forecast mortality by projecting the mortality by cause over time, before aggregating the forces of mortality to retrieve the all-cause mortality. However, it seems that in older ages the cause-specific mortality data suffer from a lack of robustness which can lead to deteriorate the forecasts. The first reason of this poor data quality is that the number of deaths at advanced ages is usually low due to simple fact that few people manage to survive so far. This problem is particularly acute when dealing with cause-specific mortality which split the total amount of death between several subgroups. Another main issue when dealing with causes of death at advanced ages is the reliability of registered cause of death. As identified by Gessert et al. (2002)[61], determining the main cause of death for centenarians is especially difficult because of the presence of multi-pathologies and risk factors related together. For these reasons we propose an extrapolation model adapted to cause-specific mortality which is based on age-pattern mortality dynamic at ages where the data are robust.

The mortality extrapolation for older ages is not an issue specific to causes of death data. The first extrapolating models at high ages have in fact emerged for the overall mortality. These models often rely on parametric methods to extrapolate forces of mortality and are based on specific hypothesis. For instance, Denuit and Goderniaux (2005)[42] is based on the assumption of limit age of life at which the whole population is dead and a quadratic parametric form. The Kannisto model (see Kannisto et al. (1994)[77]) is derived from the hypothesis of an heterogeneous population in which the phenomenon of frail individuals experiencing death before the others leads to a logistic growth of the high ages mortality. These methods were designed for all-cause life table completion and their direct application to

cause-specific forces of mortality is not suitable. In this chapter, we suggest a method allowing extrapolation of mortality by cause while maintaining the specific constraints of the completion methods on the all-cause mortality. To achieve this purpose, we propose a top-down extrapolation approach. This method finds its origin in the literature dealing with the hierarchical time series forecasting (see Shwarzkopf et al. (1988)[122] or Athanasopoulos et al. (2011)[73]). A hierarchical multivariate time series is a structured data composed of time varying series grouped by level, in which the higher level time series are equal to the sum of some inferior level time series. Replacing the time-varying series by age-varying series, the cause-specific deaths and all-cause deaths constitute a very simple hierarchical structure, in which the all-cause force of mortality is equal to the sum of the cause-specific forces of mortality.

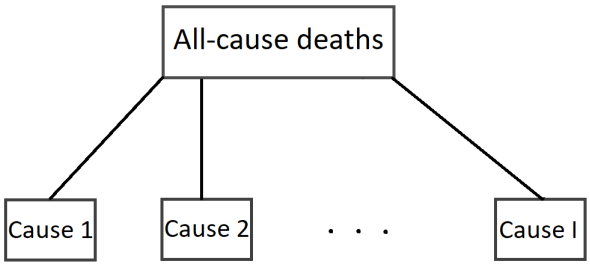


FIGURE 2.1 – *Hierarchical structure of the mortality*

The extrapolation of hierarchical data may be handled using the three approaches presented by Athanasopoulos et al. (2009)[12]. The first one is the bottom-up approach which consists in forecasting the cause-specific mortalities and to sum them up to get the all-cause mortality extrapolation. The second option is the top-down approach which extrapolates the all-cause mortality and disaggregate it to distribute it to the cause-specific mortality. The disaggregation from the top series to its subcomponents can be either based on historical or forecasted contribution of the causes of death to the all-cause mortality. The third method, named optimal forecast reconciliation, proposes to extrapolate all the series and then to reconcile the hierarchical structures, using a linear regression model. All these approaches are treated by Athanasopoulos et al. (2009)[12] for linear state-space time series with additive error terms. Usually, the all-cause mortality is extrapolated at older ages with life table closure techniques relying on specific assumptions which can be either medical or technical. Considering cause-specific mortality, an extrapolation process based on bottom-up or optimal reconciliation methods may lead to inconsistent all-cause mortality in regard to the usual mortality assumptions. Therefore, the only approach among the three ensuring that the all-cause mortality

assumptions are met is the top-down method. We adapt the approach developed for state-space time series to the probabilistic framework of cause of death realizations by age.

The approach is adjusted to fit the mortality in a Poisson framework. In our study, the series are not time-dependent but age-dependent. Assuming a Poisson distribution for the realization of cause-specific deaths, we obtain, under some assumptions developed in Section 2.2, a suitable decomposition for top-down approach. The aggregate mortality is a Poisson realization while the contributions from each cause are multinomial distributed. The first one is extrapolated at advanced ages with usual life table closure methods, while the contributions are extrapolated with the GLM P-splines approach (see Eilers and Marx (1996)[50]). One of the theoretical contributions of this paper is to enlighten two variants of the multinomial model. These models are inspired by the "Compositional Data Analysis" (CoDA) theory, developed by Aitchison (1986)[2]. By combining the extrapolations of contribution and all-cause mortality, we obtain the cause-specific mortality extrapolation.

This chapter is organized as follows. In Section 2.2, we define the theoretical framework for analyzing cause-specific mortality. The assumptions lead to split the likelihood into two parts : on the one hand, the likelihood associated with the general mortality and, on the other hand, the likelihood associated with the contribution of each cause of death to the overall mortality. Section 2.3 concerns life table closure approaches. We present three closure methods commonly used in practice and based on specific assumptions regarding mortality at older ages. Section 2.4 is devoted to the multinomial GLM theory. After a brief review on the multinomial GLM, we show how the different transformations introduced in the Compositional Data Analysis framework can be transposed to the multinomial framework. We also provide details on the association of P-splines and GLMs. Section 2.5 discusses the application on US Female mortality by cause of death. The impact on the extrapolated mortality resulting from each closure method and each multinomial model is addressed. Finally, Section 2.6 concludes this paper.

2.2 Notations and hypothesis

Let x be an integer defined on $\left[x_0, \ldots, x_{\omega_1}\right]$, denoting the age variable and I the number of causes of death. We assume that each death can be assigned to a cause i, and that the causes of death are mutually exclusive, i.e. an individual cannot die

from two different causes of death. The last hypothesis is questionable, especially at older ages for which the determinants leading to the death are manifold or difficult to figure out. Nevertheless, the integration of multiple causes of death makes the modelling more difficult because of the specific models required to capture the multi-aspect dynamic of death (see Moreno-Betancur et al. (2016)[102]). For this reason, we only consider the usual practice consisting in registering the leading cause of death among all the potential ones. For each age x, we have at our disposal information about :

- The number of individuals deceased from cause i at age x is denoted by D_x^i and summarized in the vector $\mathbf{D}_x = \big(D_x^1, \ldots, D_x^I\big)$.
- The all-cause number of death at age x is $D_x = \sum_{i=1}^{I} D_x^i$.
- The total exposure to death, denoted E_x represents the total observed time spent by our individuals between ages x and $x+1$.

We also assume that lifespan is represented by a random variable X. In presence of causes of death, the competing risk framework models the individual life span as $X = \min\big(X_1, X_2, \ldots, X_I\big)$, where X_i represents the age at death from cause i. A convenient assumption in this framework (see Prentice et al. (1978)[112], Wilmoth (1995)[135] and Putter et al. (2007)[113]) is the independence between the cause-specific lifespans X_i which we assume satisfied in the following of this paper. Using these notations, an individual dies from cause i at the integer age x if the associated cause-specific lifespan variables verify $X_i \in [x, x+1[$ and $X_j > X_i$. The force of mortality coefficient for cause of death i at age x is defined as

$$\mu_x^i = \lim_{s \to 0_+} \frac{\mathbb{P}(x < X_i \leq x+s | x < X)}{s},$$

while the all-cause force of mortality is

$$\mu_x = \lim_{s \to 0_+} \frac{\mathbb{P}(x < X \leq x+s | x < X)}{s}.$$

We also suppose that μ_x^i is piecewise constant, i.e. for any integer x,

$$\mu_{x+s}^i = \mu_x^i, \text{ for } s \in [0,1[.$$

The independency hypothesis implies that

$$\mu_x = \sum_i^I \mu_x^i. \tag{2.1}$$

Under the hypotheses of a piecewise force of mortality, independence between

individuals and cause-specific lifespan independency, the likelihood writes (see Andersen et al. (1993)[7] pp. 402-406) :

$$L((D_x^i)_{x,i};(\mu_x^i)_i) = \prod_{x=x_0}^{x_{\omega_1}} \prod_{i=1}^{I} exp(-E_x\mu_x^i)(\mu_x^i)^{D_x^i} \tag{2.2}$$

Regarding μ_x^i we recognize that (2.2) is proportional to a product of independent Poisson distributions of parameter $E_x\mu_x^i$. As a result we may consider that

$$D_x^i \sim \mathscr{P}\big(E_x\ \mu_x^i\big)$$

with $\mathscr{P}\big(\lambda\big)$ the Poisson distribution of parameter λ.

The next part of this section deals with the way to split the likelihood into two independent parts : one relative to the all-cause forces of mortality and the other to the contribution of causes of death to it. In order to achieve this decomposition, we exploit the well-known property (see Fisher (1922)[54]) according to which for $P_1,...P_I$ distributed as independent Poisson with respective parameters $\lambda_1,...\lambda_I$, $P_1,...,P_I|P \sim \mathscr{MN}\left((p_i)_i,P\right)$. where $P=\sum_i P_i$, $p_i=\frac{\lambda_i}{\sum_j\lambda_j}$, and $\mathscr{MN}\left((p_i)_i,n\right)$ is the multinomial law of parameters n and $(p_i)_i$. Because the D_x^i are considered as independent Poisson variables, it follows that :

$$\big(D_x^1,\ldots,D_x^I\big)|D_x \sim \mathscr{MN}\left((S_x^i)_i,D_x\right),$$

where $S_x^i=\frac{\mu_x^i}{\mu_x}$. As a sum of independent Poisson variables $D_x \sim \mathscr{P}\big(E_x\mu_x\big)$, we have $\mathbb{P}\left((D_x^i)_i=(d_x^i)_i \cap D_x=d_x\right)=\mathbb{P}\left((D_x^i)_i=(d_x^i)_i|D_x=d_x\right)\mathbb{P}\left(D_x=d_x\right)$.

Subsequently, we will assume that the force of mortality μ_x depends on the vector of parameters θ_μ and that the relative contribution S_x^i depends on the vector of parameters θ_S, which allows us to split the likelihood into two parts :

$$L\big((d_x^i)_{x,i}\ ;\ \theta_\mu,\theta_S\big) = \prod_x \frac{\big(E_x\ \mu_x(\theta_\mu)\big)^{d_x}}{d_x!}\exp\big(-E_x\ \mu_x(\theta_\mu)\big)\binom{d_x}{d_x^1,\ldots,d_x^I}\big(S_x^1(\theta_S)^{d_x^1},\ldots,S_x^I(\theta_S)^{d_x^I}\big),$$

with $\binom{d_x}{d_x^1,\ldots,d_x^I}=\frac{d_x!}{\prod_{i=1}^I d_x^i!}$

Hence, each part of the log-likelihood depends on a different set of parameters :

$$l\big(d_x^1,\dots,d_x^I\,;\,\theta_\mu,\theta_S\big) = l_1\big(d_x\,,\,\theta_\mu\big) + l_2\big(d_x^1,\dots,d_x^I\,;\,\theta_S\big) + C,$$

where,

$$l_1\big(d_x\,,\,\theta_\mu\big) = \sum_x d_x\ \log(\mu_x\big(\theta_\mu\big)) - E_x\ \mu_x(\theta_\mu),$$

$$l_2\big(d_x^1,\dots,d_x^I\,;\,\theta_S\big) = \sum_{x,i} d_x^i\ \log(S_x^i\big(\theta_S\big)).$$

Consequently, the calibration process can be split in two steps :

(i) the global forces of mortality, μ_x, are estimated using usual parametric methods by maximizing $l_1\big(D_x\,,\,\theta_\mu\big)$,

(ii) the relative contribution, S_x^i, are estimated by maximizing $l_2\big(D_x^1,\dots,D_x^I\,;\,\theta_S\big)$.

If step (i) is a common exercise for actuaries and demographers, step (ii) requires a specific framework which is developed in Section 2.4. Once the model parameters are estimated, we obtain functions $\mu_x(\theta_\mu)$ and $S_x^i\big(\theta_S\big)$, that we extend over for old ages $x > x_{\omega_1}$. For a given age x, the cause-specific forces of mortality are then derived as follows :

$$\mu_x^i(\theta_\mu,\theta_S) = \mu_x(\theta_\mu)S_x^i(\theta_S).$$

2.3 Life table closure

The lack of mortality data for the most advanced ages being a source of uncertainty, demographers and actuaries developed parametric methods to extrapolate the mortality coefficients at older ages. These models capture the shape of the mortality curves at ages where data are reliable and extrapolate the mortality using the calibrated parameters. The specifications of the models determine the subjacent hypotheses made on mortality at advanced ages. Considering the vectors $(D_x)_x$ and $(E_x)_x$ for ages $x = x_0, ..., x_{\omega_1}$, a completion is achieved up to an arbitrary defined limit age x_{ω_2}. Generally, the completion is done up to 120 or 130 years old.

This section covers three different methods often met in practice : Coale and Kisker (1990)[28] (CK), Gompertz-Makeham (1860)[95] (GM) and Kannisto (1994)[77] (Kan). Each of these methods corresponds to specific assumptions made on the advanced ages mortality and results in forces of mortality curves which can differ substantially. The models introduced below are fitted with the Poisson likelihood

as written in Section 2.2. The parametric forms of these models are described in Table 2.1.

Model	Dimension of θ_μ	Force of mortality
Coale-Kisker	3	$\mu_x(\theta_\mu) = exp(\theta_\mu^1 x^2 + \theta_\mu^2 x + \theta_\mu^3)$
Gompertz-Makeham	3	$\mu_x(\theta_\mu) = \theta_\mu^1 exp\left(-\theta_\mu^2\right) + \theta_\mu^3$
Kannisto	2	$\mu_x(\theta_\mu) = \frac{\theta_\mu^1 exp(\theta_\mu^2 x)}{1+\theta_\mu^1 exp(\theta_\mu^2 x)}$

TABLE 2.1 – *Information relative to the life table closure methods*

The original CK (1990)[28] approach is based on a model which does not need particularly complex likelihood information, due to specific hypothesis on the mortality at older ages. Nevertheless, Flici (2016)[55] shows that the CK model can be rewritten into a parametric form which does not require assumptions and appears to be a log-quadratic function which is easier to include in the likelihood framework described in Section 2.2.

The GM model is a very well known method to calibrate the mortality of populations. It has been suggested by Gompertz (1825)[64], and improved by Makeham (1860)[95]. Accurate and parsimonious, the model combines both age independent and age dependent terms to capture the mortality dynamic on a large age range. Although the GM model has been criticized for its low capability to fit the mortality at advanced ages, Gavrilov and Gavrilova (2011)[59] and Gavrilova and Gavrilov (2017)[60] suggest that the logistic shape of mortality is due to poor data quality. Relying on better data results in the fact that the best model to extrapolate the mortality is not necessarily from the logistic function family. This observation suggests that the GM model may be suitable to extrapolate the end life force of mortality. Actually, they noticed that the US mortality at high ages, at least up to 106 years old, seems to be fitted properly with a GM model for cohorts born between 1894 and 1898.

Thatcher et al. (1998)[124] observe for some countries a deceleration of the force of mortality at the most advanced ages. Several models have been tested on different populations and the Kan model was identified as the most appropriate model to catch the shape of the mortality curve at older ages. An explanation of this phenomenon was proposed by Vaupel (1979)[128] according to which the deceleration of mortality at advanced ages is the consequence of the population

heterogeneity. The frail part of the population dying earlier, only the most robust individuals remain. These ones experiencing a lower mortality, a progressive slowdown in the growth of the force of mortality is observed leading to a mortality plateau.

2.4 P-splines Multinomial GLMs

2.4.1 Compositional data analysis and multinomial framework.

This section is devoted to the technical issues of likelihood estimation in a multinomial framework. We show that the CoDA theory (see Aitchison (1986)[2]) can be incorporated into the usual GLM theory developed by Mccullagh and Nelder (1989)[96] to obtain variants of the usual multinomial GLM. Associated with P-splines function as a predictor, the multinomial GLM allows us to extrapolate the contribution of causes of death to all-cause mortality at older ages in a very smooth way. Actually, due to its flexibility, P-spline allows to catch the non-linear evolution of cause of death proportion while extrapolating them linearly. The regression theory proposes to link response variables to associated explanatory variables. We denote by Y the I-dimensional response variable we wish to predict using the p-dimensional explanatory variables z, and consider N observed couples (Y_k, z_k). We denote by Z_k the matrix of dimension $I \times I(p+1)$, with each row containing the information of z_k such that :

$$Z_k = \begin{pmatrix} 1 & 0 & ... & 0 & z_k^{'} & 0_p & ... & 0_p \\ 0 & 1 & ... & 0 & 0_p & z_k^{'} & ... & 0_p \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0_p & 0_p & ... & z_k^{'} \end{pmatrix}.$$

where 0_p is a row vector whose components are null. The GLM theory allows to estimate the relationship between Y and z for a probability distribution belonging to the exponential family through a structural hypothesis on the linear predictor and the link function. A distribution belonging to the exponential family assumes the following probability function :

$$f(Y_k|\zeta_k, \phi, w_k) = exp\left(\frac{Y_k^{'}\zeta_k - b(\zeta_k)}{\phi} w_k - c(Y_k, \phi, w_k)\right) \text{ for all } k = 1, ..., N,$$

where ζ_k is the natural parameter associated to the observation k, w_k a specific constant depending on the observation k, ϕ the dispersion parameter common to

all observations, $c(.)$ a function independent of the natural parameter and $b(.)$ a differentiable function whose the gradient is bijective. One of the most important aspects with multivariate GLM is that the vector Y_k must not be constrained. Assuming the variables Y has a distribution belonging to the exponential family framework, we have :

$$u_k = \mathbb{E}[Y_k|Z_k] = \nabla b(\zeta_k),$$
$$Var[Y_k|Z_k] = \nabla^2 b(\zeta_k)\phi.$$

In order to estimate u_k, the conditional mean function, a structural hypothesis is made on $\zeta_k = Z_k\beta$, where β is the $I(p+1)$-dimensional vector parameter relating the response and explanatory variables. As a result, we have $g(u_k) = \zeta_k$, where $g(.)$ referred to the link function, and is equal to the inverse of the function $\nabla b(.)$. The estimation of the parameter β is performed using IWLS algorithm as detailed in Appendix C for the case where the observations Y are multivariate.
Considering the cause of death multinomial framework, the observation index k is replaced by the age observation x and the probability function writes

$$\begin{aligned} f(\mathbf{S}_x, \mathbf{D}_x) &= \binom{D_x}{D_x^1, \ldots, D_x^I} \left((S_x^1)^{D_x^1}, \ldots, (S_x^I)^{d_x^I}\right) \\ &= exp(\mathbf{D}_x^{'} log(\mathbf{S}_x) + c(\mathbf{D}_x)). \end{aligned}$$

As the vector $\mathbf{D}_x$ is constrained to sum up to D_x, this function cannot be considered as belonging to the exponential family. A way to get around this issue is to consider the probability function of a transformation of the vector $\mathbf{D}_x$ through an application $T(.)$ such that the constraint vanishes. The usual transformation consists in dropping one component of this vector. We denote by $T(\mathbf{D}_x) = \mathbf{D}_x^{-I}$, the vector obtained by setting aside the last component of the vector $\mathbf{D}_x$. As detailed in Appendix B.2, this transformation leads to in the well-known multi-logistic link function :

$$g(\mathbf{S}_x)_i = log\left(\frac{S_x^i}{1 - \sum_{j \neq I} S_x^j}\right).$$

We propose here other transformations which conducts to define particular link functions. In addition of the multi-logistic transformation which is named thereafter the additive log ratio transformation (alr), we introduce two other transformations from the CoDA theory which are the centered log-ratio (clr) and the isometric log-ratio transformations (ilr). Each of these functions is applied on a compositional data, which is a vector whose the components are positive and sum

up to 1. Let C be a I-dimensional compositional vector such that $\sum_{i=1}^{I} C^i = 1$. Formally, these transformations are defined as follows :

- $alr(C) = \left(\log\left(\frac{C^1}{C^I}\right), \log\left(\frac{C^2}{C^I}\right), \cdots, \log\left(\frac{C^{I-1}}{C^I}\right)\right)$;
- $clr(C) = \left(\log\left(\frac{C^1}{\left(\prod_i^I C^i\right)^{\frac{1}{I}}}\right), \log\left(\frac{C^2}{\left(\prod_i^I C^i\right)^{\frac{1}{I}}}\right), \ldots, \log\left(\frac{C^I}{\left(\prod_i^I C^i\right)^{\frac{1}{I}}}\right)\right)$;
- $ilr(C) = \log\left(W'C\right)$ where W is an orthonormal base of vectors of the I-dimensional simplex space (see Appendix A for more details).

Each of these transformations is the result of the application of a specific function $T(.)$ on the vector $\mathbf{D}_x$, leading to particular ζ_x and $b(.)$. We summarize the three distributions in Table 2.2.

	$T(\mathbf{D}_x)$	ζ_x	$b(\zeta_x)$	w_x
alr	$\mathbf{D}_x^{-I}/D_x$	$alr(\mathbf{S}_x)$	$log(1 + \sum_{i=1}^{I-1} exp(\zeta_x^i))$	D_x
clr	$(\mathbf{D}_x^{-I} - D_x^I)/D_x$	$clr(\mathbf{S}_x)^{-I}$	$log\left(\sum_{i=1}^{I-1} exp(\zeta_x^i) + exp(-\sum_{i=1}^{I-1} \zeta_x^i)\right)$	D_x
ilr	$(V'\mathbf{D}_x)/D_x$	$ilr(\mathbf{S}_x)$	$log\left(\sum_{i=1}^{I} exp\left(V'_{i,.}\zeta_x\right)\right)$	D_x

Table 2.2 – *CoDA Interpretation for GLMs*

Proofs of these relations are provided in Appendix B, while the definition of the matrix V for the ilr transformation is provided in Appendix A.

2.4.2 P-splines for Multinomial models

The P-splines technique is a statistical method of regression allowing for a non-linear relationship between the variable of interest and the explanatory variables. This method introduced in Eilers and Marx (1996) [50] has become notorious among the statisticians due to its great flexibility. The principle is to combine a large amount of B-spline functions with an L^2 penalization to obtain a good quality of fit without the downside of overfitting.
Initially developed for interpolation, this technique is also designed for extrapolation as shown by Currie et al. (2004)[33], who apply the P-splines to extrapolate mortality by ages and periods. In our approach, we use the P-splines as a predictor to extrapolate the contributions of each cause of death at older ages.

The B-splines method is a flexible non-parametric approach which allows good fitting without considering specific model. A B-spline is defined by its degree and associated knots. We denote by q of the degree of the B-splines, r the number of knots and $K_0 < K_1 < ... < K_r$ the knots.
The B-splines of degree q is an n sequence of B-spline functions, where $n = r - q$.

We denote by $B_{j,q}(x)$ the j-th component of the B-splines of degree q, considered at x. The B-spline functions are linearly combined to obtain the spline curve, which allows approximating univariate functions. De Boor (1978)[35] provided a recursive algorithm to build a B-spline function of any degree starting from an indicator function :

- $B_{j,0}(x) := \begin{cases} 1 & \text{if} \quad K_{j-1} \leqslant x < K_j \\ 0 & \text{else} \end{cases}$
- $B_{j,q}(x) := \frac{x - K_{j-1}}{K_{j+q-1} - K_{j-1}} B_{j,q-1}(x) + \frac{K_{j+q} - x}{K_{j+q} - K_j} B_{j+1,q-1}(x).$

In the following, we only consider B-splines of degree 2. We denote by $B(x) = (B_{1,2}(x), B_{2,2}(x), ..., B_{n,2}(x))'$ the vector containing the sequence of B-spline values considered at age x. For each age x, we denote by Z_x the $I^\star \times I^\star n$ matrix predictor associated to the $I^\star$-dimensional multivariate response Y_x which verifies :

$$Z_x = \begin{pmatrix} 1 & 0 & ... & 0 & B(x)' & 0_n & ... & 0_n \\ 0 & 1 & ... & 0 & 0_n & B(x)' & ... & 0_n \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0_n & 0_n & ... & B(x)' \end{pmatrix},$$

with 0_n the row vector of length n whose components are 0. This matrix is the input of the multinomial GLM for age x. The aggregated predictor Z is defined by regrouping all Z_x matrices :

$$Z = (Z'_{x_0}, ..., Z'_{x_{\omega_1}})'.$$

Using the notations introduced in Section 2.4.1, the likelihood maximization step writes as :

$$max_\beta \sum_{x=x_0}^{x_{\omega_1}} D_x \left[T(\mathbf{D}_x)'(Z_x\beta) - b(Z_x\beta) \right],$$

where β is the parameter vector of the model. If the number of B-spline functions n is too large, a straightforward optimization of the above likelihood can easily lead to overfitting issues. To avoid this pitfall, a penalization of the parameters can be added to the likelihood function for parameters relating to the B-spline functions. We denote by λ the penalty parameter and by $\Delta^d(.)$ the difference function of order d such that for a sequence v, $\Delta^d(v)_t = \Delta^{d-1}(\Delta(v))_t$ and $\Delta(v)_t = v_{t+1} - v_t$. Then, penalized likelihood writes (if not penalizing the intercept) :

$$max_\beta \sum_{x=x_0}^{x_{\omega_1}} D_x \left[T(\mathbf{D}_x)'(Z_x\beta) - b(Z_x\beta) \right] - \frac{\lambda}{2} \sum_{j \geq I^\star + 1} (\Delta^2(\beta)_j)^2.$$

The optimization of this likelihood is achieved using the IWLS algorithm as detailed in Appendix C.1.

2.5 Application

2.5.1 Data

The data come from the US Center of Disease Control (CDC)[56]. We use the Female mortality data corresponding to the calendar year 2017 and age range 51-90 years as reference for the fitting. As women experience longer life expectation than men, the question of extrapolation at oldest ages is particularly relevant for this segment of population. The US female exposure is taken from the HMD database [71]. For this application the Female causes of death, data are grouped in 7 categories presented in Table 2.3.

Cause of Death	1999 to 2017 (ICD 10 : UCOD 358)
Cancer	69 - 146
Metabolic	156-173
Mental & Nervous	174-194
Circulatory system	197-246
Respiratory system	247-278
External	381-456
Others	All others

Table 2.3 – *Mapping of the causes of death for ICD 10.*

Table 2.3 displays the International Classification of Diseases (ICD) codes used by a physician to record each death. The pathology recording system being very accurate, it leads to high level of granularity which might not be suitable to obtain robust statistics or to propose experts' judgments. Focusing on the most important causes of death, we regroup the causes into larger categories. We provide the mapping codes allowing to rebuild these classes starting from the original database. As indicated in the last row in Table 2.3, deaths falling outside each category are grouped into another category called "Others".

2.5.2 Extrapolation of the all-cause mortality

The first step consists in extrapolating the all-cause force of mortality at older ages through the life table closure methods.

Table 2.4 presents the likelihoods for the three completion methods. We note that the likelihoods are very close for all methods even if the numbers of parameters are not the same for each method. Nevertheless, regarding the fitting ages the rank

Method	Log Likelihood	Life Expectancy at 90
Coale-Kisker	768273.6	3.91
Gompertz-Makeham	768077.3	4.61
Kannisto	768465.2	5.14

TABLE 2.4 – *Information relative to the life table closure methods*

is the following : Kan, CK and GM. We observe that each model produces very different projected mortality as illustrated by the important variations among the induced life expectancies after 90.

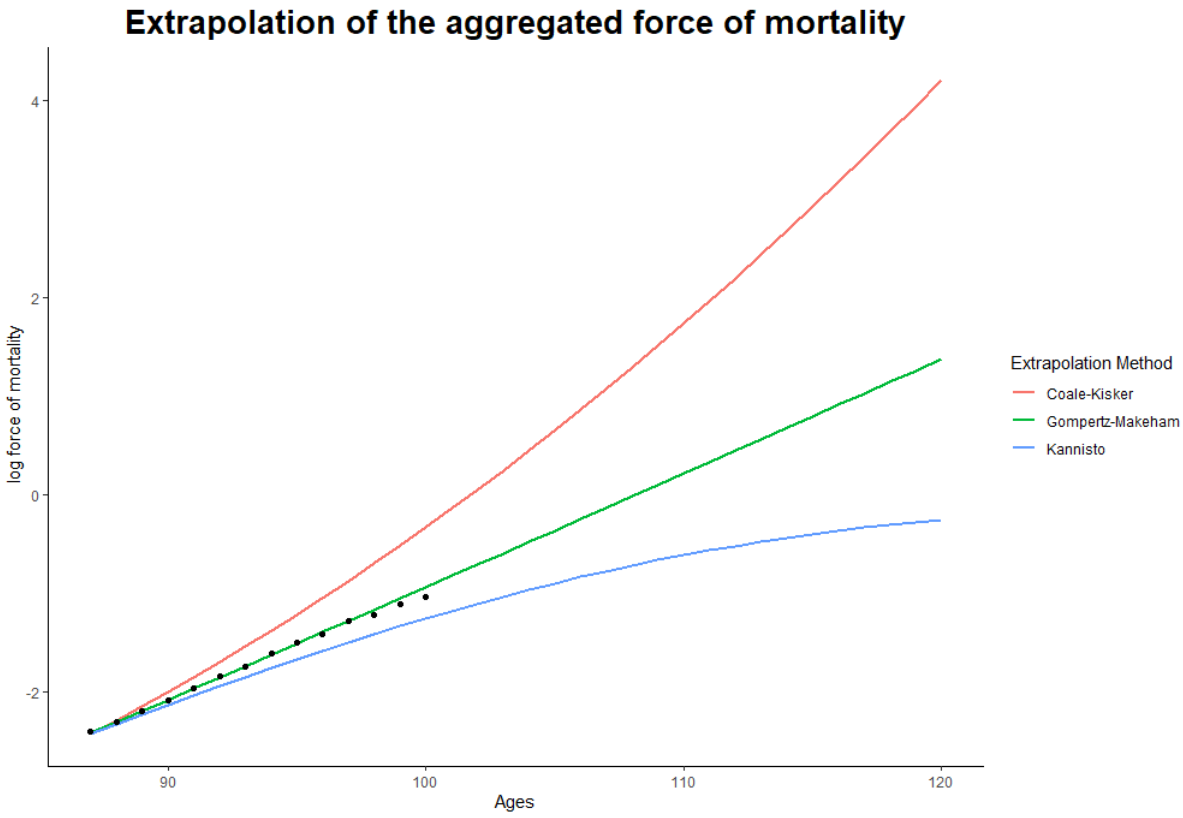


FIGURE 2.2 – *Extrapolation of the general mortality vs the observed realization of mortality*

Figure 2.2 displays the extrapolated forces of mortality until age 120. It can be seen on the graph above that forces of mortality at advanced ages are substantially different between CK, GM and Kan models. The slope of the CK mortality force is stronger than the one of the GM method. As for the Kan method, there is a significant difference compared to the others, due to the convergence to 1 at extreme ages, which constrains its growth.

The high mortality induced by CK implies a low life expectancy compared to

the others. The Kan and GM methods also present important differences. Indeed, the curves seem to diverge in the completely opposite direction leading to almost 1 year of life expectancy difference between CK and GM, and half this difference between GM and Kan methods. We also add the forces of mortality observations until 100. We can observe that for this age range the GM model seems to be the most suitable closure method.

2.5.3 Extrapolating the causes of death contributions

This section presents the results relating to the extrapolation of the proportions of deaths at older ages. The method was introduced in Sections 2.3 and 2.4. Table 2.5 provides the penalty parameters and corresponding fitting criteria of the three multinomial GLMs introduced in Section 2.4. Each P-splines methods has been calibrated in the same way, at the exception of the penalty parameter. We empirically tested the P-splines model with different parameters and selected 20 knots for the fitting age group and 16 for the extrapolation age range. The selection of the model parameter λ was done by minimizing the AIC, as presented in Eiler and Marx (1996)[50].

Method	Penalty Parameter	Log Likelihood	AIC
alr	0.07879205	1673811	3347735
clr	0.1502377	1673811	3347737
ilr	0.1176795	1673811	3347736

Table 2.5 – *Information regarding the parameters for the multinomial GLMs.*

The three methods provide similar results regarding the likelihood. It should be noted that the penalization must be stronger for the transformations *clr* and *ilr* than for the transformation *alr*.

We observe that the P-splines fitting applied to the three multinomial GLMs capture very well the individual cause of deaths characteristics. The small fitting differences observed in the age range 51-90 is due to the low volatility of the data and the large number of splines.

However, as outlined in Figure 2.6 the extrapolated forces of mortality by cause on the age range 91-120 differ for some causes despite having a similar fitting with the three GLMs. Although cancers are not subject to major differences regarding the method, extrapolations for circulatory, external and metabolic diseases seem to experience important variations according to the GLM link function chosen. We invite the reader to refer to Appendix A.2 for an explanation about the differences

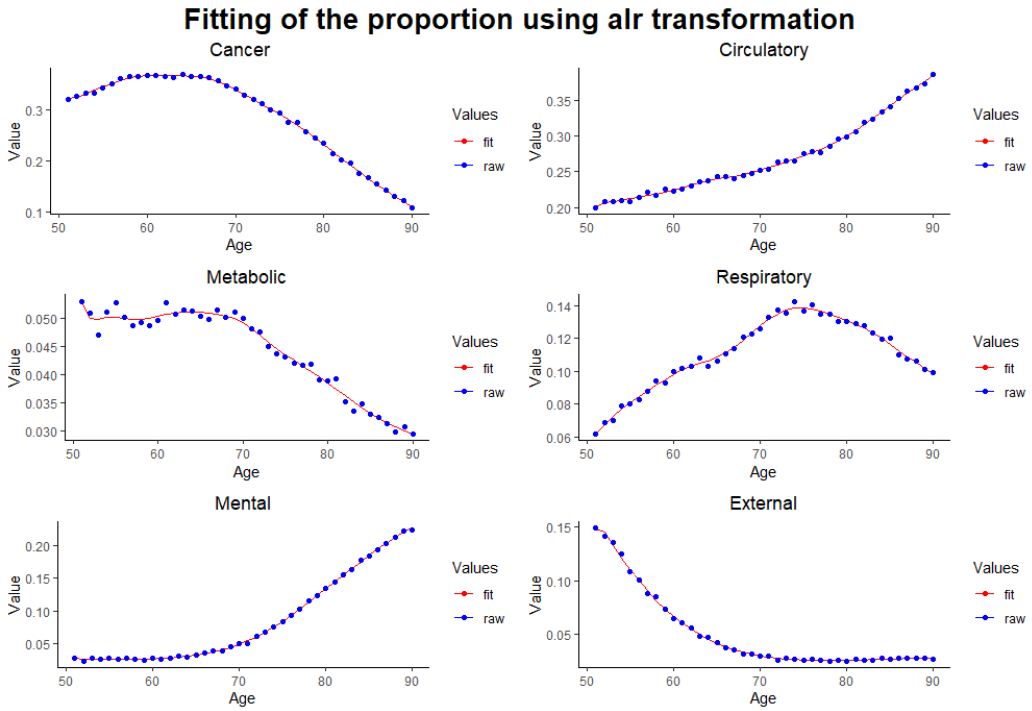


FIGURE 2.3 – *P-splines fitting using alr transformation*

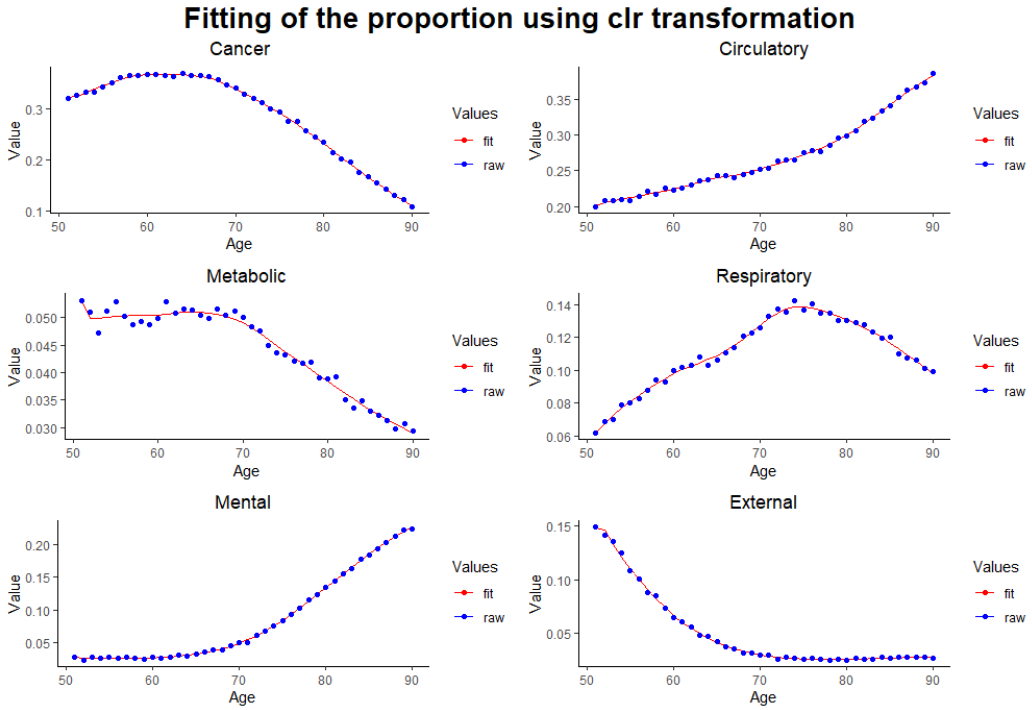


FIGURE 2.4 – *P-splines fitting using clr transformation*

in the evolution of the cause of death contributions according to each multinomial GLM.

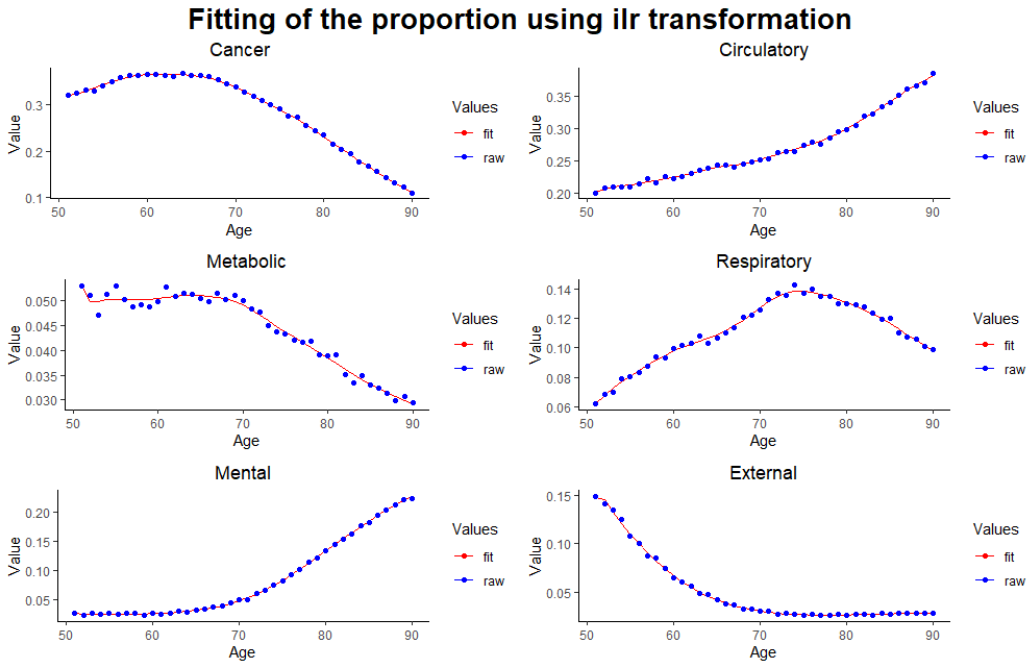


FIGURE 2.5 – *P-splines fitting using ilr transformation*

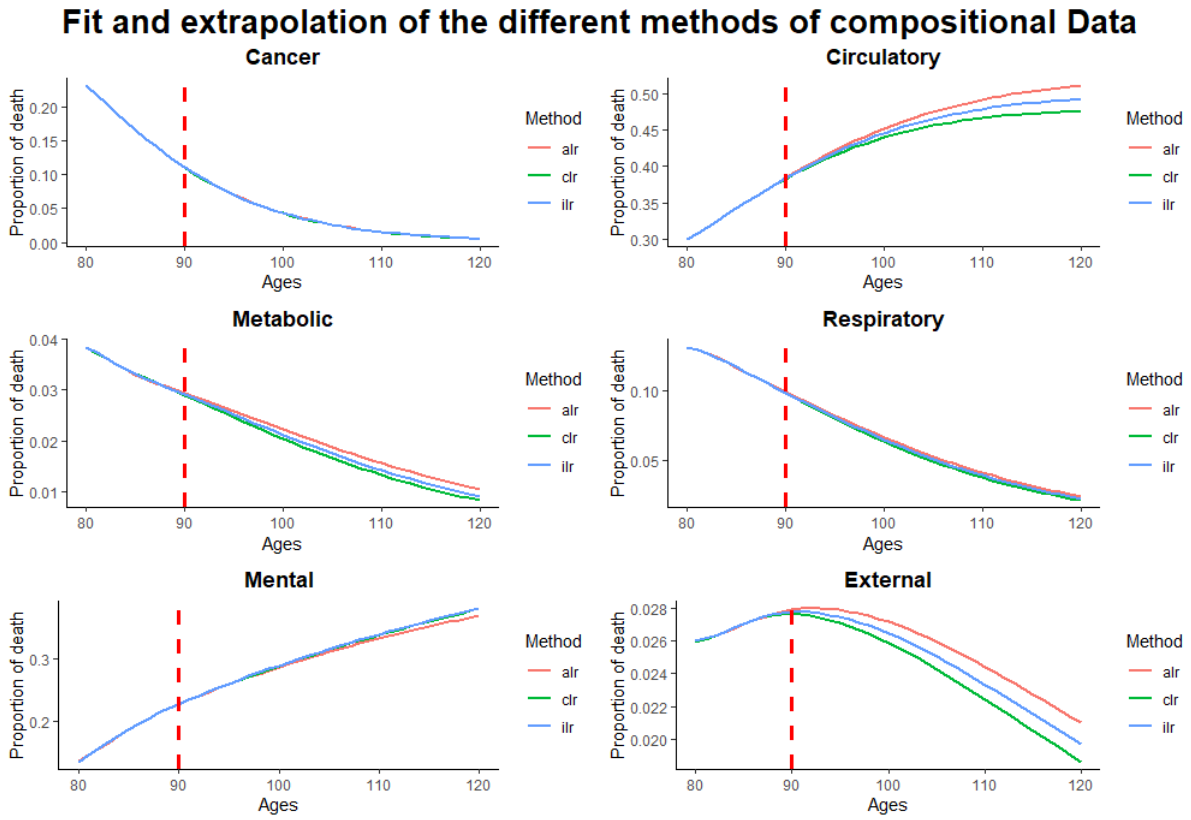


FIGURE 2.6 – *Extrapolation of the Cause of death contribution to the aggregated force of mortality*

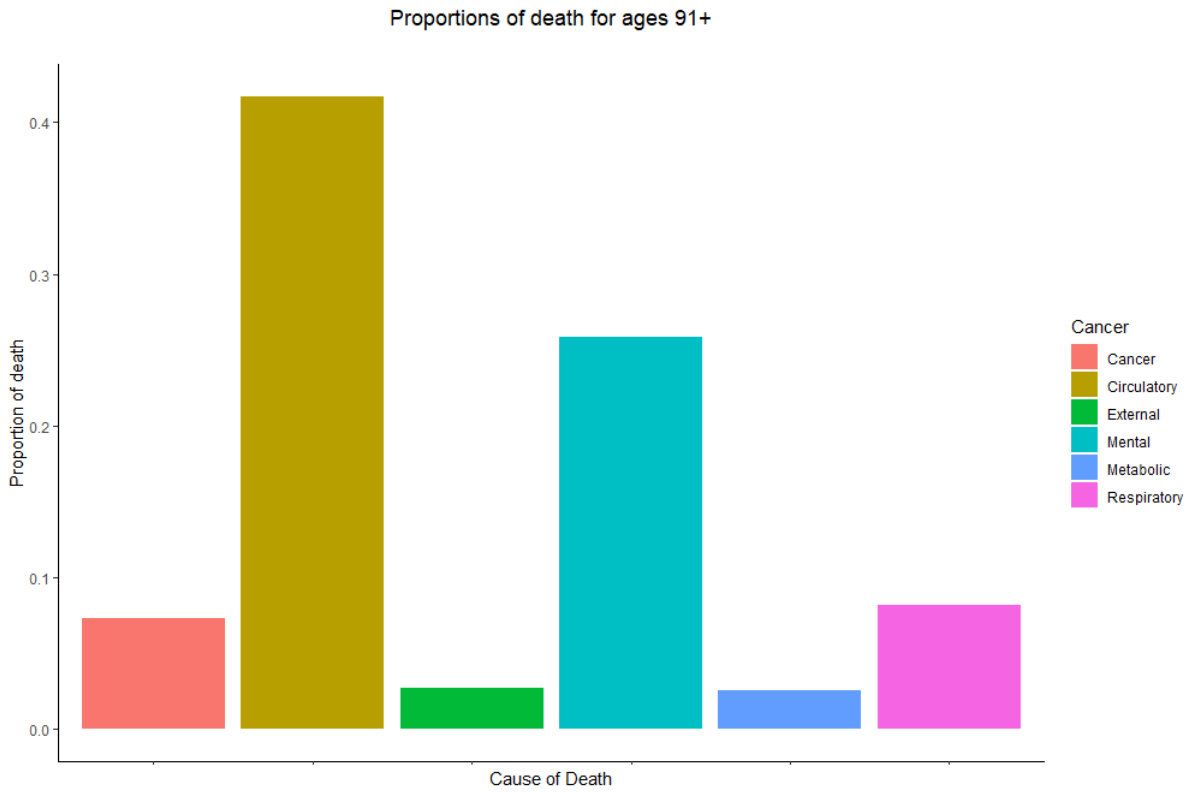
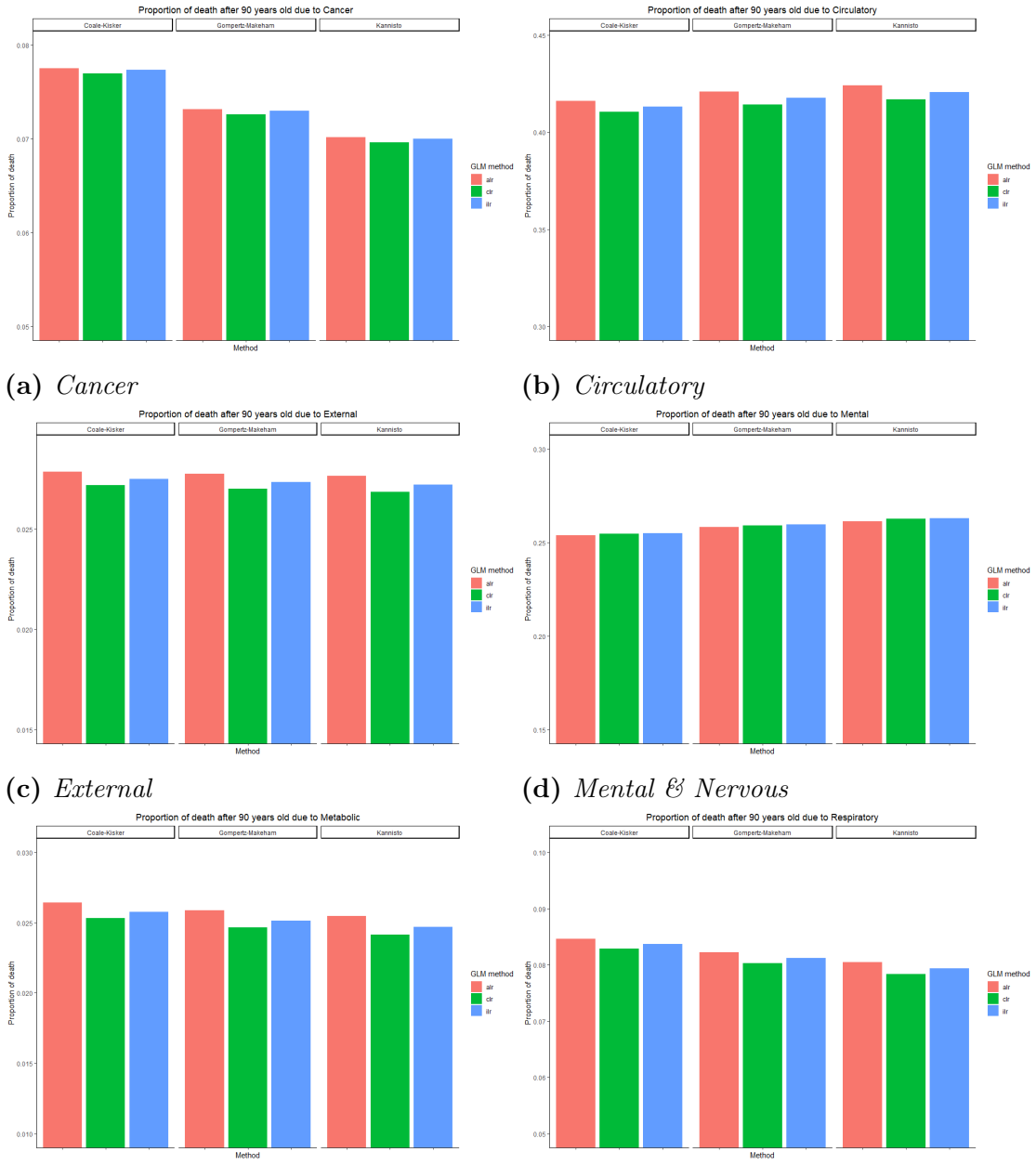


FIGURE 2.7 – *View on the proportion of death by cause after 90, by averaging all methods*

2.5.4 Global impact on death by cause at the advanced ages

Figure 2.7 presents the extrapolated proportions of deaths by cause after age 90. The proportions are obtained by averaging the proportions from all methods. As the proportion order is the same for all approaches, this calculus does not alter significantly the overall results. We observe that Circulatory and Mental correspond together to 67% of the deaths. Then follow Respiratory and Cancer with 8% and 7.5% respectively. Finally, External and Metabolic represent the least important causes after age 90 with both 2.5%. These results are consistent with the medical opinion according to which a significant proportion of deaths at advanced ages is due to either neurological and mental diseases or age-related vascular stiffening. These two causes "compete" with the other pathologies which gradually become less important as ageing, although the associated force of mortality may still increase.



(a) *Cancer* **(b)** *Circulatory*

(c) *External* **(d)** *Mental & Nervous*

(e) *Metabolic* **(f)** *Respiratory*

FIGURE 2.8 – *View on the proportion of death by cause after 84*

If we focus on the results obtained regarding the different methods of life table closure, it appears in Figure 2.8 that the method generating the lowest slope produces more important proportions of deaths from the causes whose contributions are rising. Actually, the methods which delay the death to extreme ages allow an increase in the number of people alive at high ages. The causes of death whose contribution in the overall mortality increases can then "catch" more people. The method which generates the youngest ages at death is the CK method for which we expect a larger proportion of deaths for causes whose contribution is decreasing.

On the other hand, the Kan extrapolation method, whose slope decelerates with ages, increases the proportions of causes for which the level increases at advanced ages.

An overall observation is that the choice of the GLM method for proportion extrapolation may have an important impact on the results even if it appears that the most important determinant is the choice of the life table closure method. We observe that the choice of link function is important for Circulatory, External and Metabolic as seen in 2.6.

All causes of death except for Mental & Nervous and Circulatory, display a larger proportion for life table closure methods having a sharper slope. Then for all these causes, the models inducing the largest number of deaths is, by decreasing order, CK, GM and Kan. At the opposite, the Mental and Circulatory proportions are more important for Kan model than for GM and more important for GM than for CK. The explanation is that these causes of death are the only ones to experience an increase in their contribution to the all-cause mortality.

Table 2.6 contains the extrapolation errors of forces of mortality by cause, closure method and link function using the sum of squared errors criterion, computed for age range 91-95. We can observe that the lowest errors derive from the GM extrapolation. The worst method among the three seems to be the parametric adaptation of the CK approach, due to the too important increase of mortality resulting from a quadratic log-mortality. Regarding the log-ratio transformations, the best one depends of the cause and the associated closure method. For the GM extrapolation, the best method is the *clr* for Cancer, *ilr* for Metabolic, *alr* for Mental, *ilr* for Circulatory, *alr* for respiratory and *alr* for External.

2.6 Conclusion

In this chapter, we proposed a method for extrapolating mortality by cause of death to advanced ages. We put forward a Top-Down method capable of maintaining the consistency of extrapolations made on aggregate mortality. We put forward the decomposition of the likelihood between the contribution of causes and aggregate mortality. Several statistical tools were then used to extrapolate contributions and total mortality. We presented three methods proven in the actuarial community for the completion of life tables : Kannisto, Coale-Kisker, Gompertz-Makeham. P-splines were used to extrapolate causal contributions to total mortality in multinomial models. We described two variants of the usual multi-logistic

	Coale-Kisker	Gompertz-Makeham	Kannisto
Cancer			
alr	0.00007133	0.00000034	0.00001231
clr	0.00006662	**0.00000028**	0.00001370
ilr	0.00006974	0.00000031	0.00001277
Metabolic			
alr	0.00000913	0.00000008	0.00000119
clr	0.00000613	0.00000009	0.00000212
ilr	0.00000728	**0.00000004**	0.00000171
Mental			
alr	0.00078626	**0.00000209**	0.00012580
clr	0.00080549	0.00000272	0.00012086
ilr	0.00081443	0.00000306	0.00011859
Circulatory			
alr	0.00201451	0.00000083	0.00037609
clr	0.00179802	0.00000217	0.00044299
ilr	0.00190525	**0.00000058**	0.00040869
Respiratory			
alr	0.00005336	**0.00000346**	0.00003417
clr	0.00004132	0.00000632	0.00004149
ilr	0.00004663	0.00000489	0.00003805
External			
alr	0.00000923	**0.00000004**	0.00000172
clr	0.00000738	0.00000008	0.00000233
ilr	0.00000817	0.00000005	0.00000205

Table 2.6 – *SSE of the extrapolation using the observed cause-specific forces of mortality from age 91 to 95*

model based on Compositional Data Analysis theory. We showed that the transformations usually used in this theoretical framework have equivalence in multinomial models. We applied the method of extrapolation of mortality by cause to the US Female population at year 2017. We found that in our example, the choice of the multinomial variant had a smaller impact on the results obtained in comparison with the closure methods. In contrast with the observed realizations on ages 91-95, we observe that the cause-specific mortality is better adjusted by the GM model than by the two others. Regarding the GLM approach, the results are less significant, the forecast quality depending on the cause considered. From a theoretical point of view, extrapolation can provide information on the assumed mortality structure of populations. This can be the subject of expert judgments in order to correct possible inconsistencies related to spline extrapolation. Finally, other avenues are being explored in order to combine information on mortality over se-

veral years. The resulting extrapolation would then be less subject to temporal volatility. From a theoretical point of view, other applications of the multinomial variants presented in this paper are to be experimented in order to better identify the differences between the different constraints.

Appendix

A Compositional Data Analysis

A.1 Compositional data framework

The vector $S = \left(S^1, S^2, ..., S^I\right)$ has some particular properties which we should have in mind. Before starting the extrapolation process. S is called a compositional data. It belongs to the simplex space denoted $\mathbb{S}^I$. A simplex space, $\mathbb{S}^I$, is a space in which the elements S satisfy the following constraints :

(i) S is a I-dimensional vector ;

(ii) for each $i \in \{1, \ldots, I\}$, $S^i > 0$;

(iii) $\sum_{i=1}^{I} S^i = 1$.

The elements of the simplex space $\mathbb{S}^I$ are named compositional data or compositional vectors. Within the simplex space, we define the following operator :

- Let S be a I dimensional vector of which each component is positive. We define the operator $\mathscr{C} : \mathbb{R}_+^I \to \mathbb{S}^I$ such that $\mathscr{C}(S)_i = \frac{S^i}{\sum_{j=1}^{I} S^j}$.
- $S_1 \oplus S_2 = \mathscr{C}\left(S_1^1\, S_2^1, S_1^2\, S_2^2, \ldots, S_1^I\, S_2^I\right)$, where S_1 and S_2 are two compositional vectors from $\mathbb{S}^I$.
- $S_1 \ominus S_2 = \mathscr{C}\left(\frac{S_1^1}{S_2^1}, \frac{S_1^2}{S_2^2}, \ldots, \frac{S_1^I}{S_2^I}\right)$, where S_1 and S_2 are two compositional vectors from $\mathbb{S}^I$.
- $\alpha \otimes S = \mathscr{C}\left((S^1)^\alpha, (S^2)^\alpha, \ldots, (S^I)^\alpha\right)$, where S be compositional vectors of $\mathbb{S}^I$ and α a scalar of $\mathbb{R}$.

A norm $||.||_a$, called the Aitchison norm, is also given to the simplex space :

$$||S||_a = \sqrt{\sum_{i<j} \frac{1}{I} \log\left(\frac{S^i}{S^j}\right)^2}.$$

A distance can be drawn naturally from this norm :

$$d_a(S_1, S_2) = ||S_1 - S_2||_a = \sqrt{\sum_{i<j} \frac{1}{I} \left(\log\left(\frac{S_1^i}{S_2^j}\right) - \log\left(\frac{S_1^i}{S_2^j}\right)\right)^2}.$$

Such data are structurally constrained. A straight application of statistical methods is not suitable. A previous step is required to allow us to study the properties of the compositional vectors. The principle is simple. The data are transformed into non-constrained spaces, on which a model is fitted. Then, statistical tools are used to extrapolate the transformed data. The final results are obtained after re-transforming the data by using the inverse function of the mentioned above transformation. This methodology finds its origin in Aitchison (1986)[2].

There are 3 well-known transforms used in the compositional data framework.

(i) $alr(S) = \left(\log\left(\frac{S^1}{S^i}\right), \log\left(\frac{S^2}{S^i}\right), \cdots, \log\left(\frac{S^{I-1}}{S^i}\right)\right)'$;

(ii) $clr(S) = \left(\log\left(\frac{S^1}{\left(\prod_i^I S^i\right)^{\frac{1}{I}}}\right), \log\left(\frac{S^2}{\left(\prod_i^I S^i\right)^{\frac{1}{I}}}\right), \ldots, \log\left(\frac{S^I}{\left(\prod_i^I S^i\right)^{\frac{1}{I}}}\right)\right)'$;

(iii) $ilr(S) = \log\left(W'S\right)$ where W is a matrix of which the columns are an orthonormal basis of vectors of $\mathbb{S}^I$.

The additive log ratio transformation (i) depends on the arbitrary choice of the i-th component of the compositional vector S_x to be used as the denominator. Its choice can alter the results if the method of forecasting is not linear. Nevertheless, the transformation is bijective, and alr(S) is not constrained.

The centered log-ratio transformation (ii) has the advantage of being easily interpretable. The more the component is important, the more is the associated clr component. The other advantage is that this transformation is isometric between $\left(\mathbb{S}^I, ||.||_a\right)$ and $\left(\mathbb{R}^I, ||.||_2\right)$. The downside is that the transformation is not injective because the vector $clr(S)$ is always constrained, i.e. $\sum_{i=1}^I clr(S)_i = 0$.

The isometric log ratio transformation (iii) was first introduced by Egozcue and al (2003)[49]. The function presents the advantage of being isometric for the $\mathbb{S}^I$ between the $\mathbb{R}^{I-1}$ and non-constrained. The main disadvantage is its weak interpretability.

Definition : Isometric transformation : let W_i, $i = 1, ..., I-1$ be an orthonormal basis in the simplex space $\mathbb{S}^I$. The isometric log ratio transform, denoted by $ilr : \mathbb{S}^I \to \mathbb{R}^{I-1}$, is defined as : $ilr(S) = \left(< S, W_1 >_a, < S, W_2 >_a, \ldots, < S, W_{I-1} >_a\right)$.

As said before, this transformation regroups the advantages of both alr and clr transformations. Unfortunately, it is rather difficult to give an interpretation of the statistics computed on this quantity. Let V be the matrix whose columns

are $V_{.,i} = clr(W_i)$ for $i = 1, .., I-1$. We have the following result :

$$Y = ilr(S) = V^{'} clr(S)$$

furthermore

$$VV^{'} = \mathbb{I}_I - \frac{1}{I}\mathbb{1}_I\mathbb{1}_I^{'},$$

and

$$V^{'}V = \mathbb{I}_{I-1},$$

where $\mathbb{1}_I$ is the I-dimensional column vector whose components are 1 and $\mathbb{I}_I$ is the identity matrix of dimension I.

A.2 Interpretation of the log ratio dynamics

The dynamic of the proportions regarding a quantitative variable may be difficult to visualize when using log-transformations. An interesting aspect of the trends with the Compositional Data is the fact they can be interpreted as sum of exchange over the variable (age or period for instance).

Let $\mathbf{S}_x$ be a compositional data which evolves regarding the quantitative variable x such that : $clr(\mathbf{S}_x) = \alpha + \Phi(x)$, where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_I)^{'}$ and $\Phi(x) = (\Phi(1,x), \Phi(2,x), ..., \Phi(I,x))^{'}$. $\Phi(x)$ is a multivariate function which relates the transformed proportion to the variable x. For instance $\Phi(x)$ may be the P-splines function linked to the age x, and the proportions may be the cause of death proportion as discussed in Section 2.2. We first start with the *clr* transformation, which implies the following constraints :

$$\sum_{j=1}^{I} \Phi(j,x) = 0,$$
$$\sum_{i=1}^{I} \alpha_i = 0.$$

We can prove that :

$$dS_x^i = \sum_{j=1}^{I} \left(d\Phi(i,x) - d\Phi(j,x)\right) S_x^j S_x^i dx \ \forall i \in \{1, ..., I\}.$$

Proof :

$$dS_x^i = \sum_{j=1}^{I} \left(\frac{d\Phi(i,x)}{dx} - \frac{d\Phi(j,x)}{dx} \right) S_x^j S_x^i dx$$

$$\frac{\frac{dS_x^i}{dx}}{S_x^i} = \sum_{j=1}^{I} \left(\frac{d\Phi(i,x)}{dx} - \frac{d\Phi(j,x)}{dx} \right) S_x^j$$

$$log(S_x^i) = \int_0^x \sum_{j=1}^{I} \left(\frac{d\Phi(i,s)}{ds} - \frac{d\Phi(j,s)}{ds} \right) S_s^j ds.$$

$$log\left(\frac{S_x^i}{(\prod_{j=1}^{I} S_x^j)^{\frac{1}{I}}} \right) = \int_0^x \sum_{j=1}^{I} \left(\frac{d\Phi(i,s)}{ds} - \frac{d\Phi(j,s)}{ds} \right) S_s^j ds$$
$$- \frac{1}{I} \sum_{j=1}^{I} \int_0^x \sum_{k=1}^{I} \left(\frac{d\Phi(j,s)}{ds} - \frac{d\Phi(k,s)}{ds} \right) S_s^k ds.$$

We note that :

(1) $\frac{1}{I} \sum_{j=1}^{I} \int_0^x \sum_{k=1}^{I} \left(\frac{d\Phi(j,s)}{ds} - \frac{d\Phi(k,s)}{ds} \right) S_s^k ds = \int_0^x \frac{1}{I} \sum_{j=1}^{I} \frac{d\Phi(j,s)}{ds} ds - \int_0^x \sum_{k=1}^{I} \frac{d\Phi(k,s)}{ds} S_s^k ds.$

(2) By assumption : $\frac{1}{I} \sum_{j=1}^{I} \frac{d\Phi(j,s)}{ds} = 0.$

Hence

$$clr(S_x)_i = \int_0^x \sum_{j=1}^{I} \frac{d\Phi(i,s)}{ds} S_s^j ds + Cons_i = \Phi(i,x) + Cons_i.$$

We set $Cons_i = \alpha_i$ to find the expected form.

This demonstration can be expended to *ilr* and *clr* transformations. Indeed, if we assume the following form for the S_x dynamic :

$$ilr(S_x) = \alpha + \Phi(x),$$

with α a vector of constants and $\Phi(x)$ a vector of differentiable functions. As $ilr(S_x) = V' clr(S_x)$, with V the matrix defined in the Section 2.2, then we obtain :

$$VV' clr(S_x) = clr(S_x) + 1_I \frac{1}{I} \sum_{i=1}^{I} clr(S_x)_i = clr(S_x) = V\alpha + V\Phi(x).$$

Hence :

$$dS_x^i = \sum_{j=1}^{I} ((V d\Phi(x))_i - (V d\Phi(x))_j) S_x^i S_x^j dx.$$

For the *alr* transformation, the proof is similar to the one for the *clr* transfor-

mation, except the constraint is not

$$\sum_{j=1}^{I} \Phi(j,x) = 0,$$
$$\sum_{i=1}^{I} \alpha_i = 0,$$

but

$$\alpha_I + \Phi(I,x) = 0.$$

These interpretations are particularly useful to better understand the dynamic of the proportions with these methods. We have shown that all the variation can be interpreted as exchanges between proportions.

In the P-splines framework with penalization of order 2, the extrapolation is linear. In consequence, each CoDA transformation has the following dynamic :

$$dS_x^i = \sum_{j=1}^{I} ((\rho_i - \rho_j) S_x^i S_x^j dx,$$

where ρ_i is the slope of $\Phi(i,x)$ for the *alr* and *clr* transformations, and the slope of $(Vd\Phi(x))_i$ for the *ilr* transformation.

B Relations between CoDA and GLM theories

B.1 Exponential family and the log ratio transformations

In this section, we show that the *alr*, *clr* and *ilr* transformation belong to the exponential family.

alr transformation

The *alr* transformation is the most commonly used transformation to deal with Compositional Data. This transformation has already been studied in the GLM framework. The *alr* leads to the standard multinomial transformation as expressed below. As $\mathbf{D}_x$ is a constrained vector, we drop one component (usually the last one), and we consider the vector $(S_x^1, ..., S_x^{I-1})$ compared to the component S_x^I.

The probability function is :

$$\begin{aligned}\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) &= \binom{d_x}{d_x^1, ..., d_x^I} \prod_{i=1}^{I} (S_i^x)^{d_x^i} \\ &= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\langle log(\mathbf{S}_x), \mathbf{d}_x \rangle\right) \\ &= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\sum_{i=1}^{I-1} d_x^i log(S_x^i) + (d_x - \sum_{i=1}^{I-1} d_x^i) log(S_x^I)\right) \\ &= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(d_x \left(\sum_{i=1}^{I-1} Y_x^i \zeta_x^i + log(1 + \sum_{i=1}^{I-1} exp(\zeta_x^i))\right)\right).\end{aligned}$$

It can be written as :

$$\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) = exp\left((\zeta_x^{'} Y_x - b(\zeta_x)) w_x + c(\mathbf{d}_x, d_x)\right),$$

where

$$\begin{aligned}\zeta_x &= alr(\mathbf{S}_x), \\ Y_x = T(\mathbf{d}_x) &= \left(\frac{d_x^1}{d_x}, \frac{d_x^2}{d_x}, ..., \frac{d_x^{I-1}}{d_x}\right), \\ b(\zeta_x) &= -log\left(1 + \sum_{i=1}^{I-1} exp(\zeta_x^i)\right), \\ w_x &= d_x, \\ c(\mathbf{d}_x, d_x) &= log\left(\binom{d_x}{d_x^1, ..., d_x^I}\right).\end{aligned}$$

clr transformation

clr is constrained by $\sum_i clr(S_i) = 0$, so we also need to drop one component in the vector. Here we compare the last one to the other component.

In this case, the probability function writes :

$$\begin{aligned}
\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) &= \binom{d_x}{d_x^1, ..., d_x^I} \prod_{i=1}^{I} (S_i^x)^{d_x^i} \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\langle log(\mathbf{S}_x), \mathbf{d}_x \rangle\right) \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\sum_{i=1}^{I-1} (d_x^i - d_x^I) clr(\mathbf{S}_x)^i - \frac{d_x}{I} \sum_{j=1}^{I} log(S_x^j)\right) \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(d_x \left(\sum_{i=1}^{I-1} T(\mathbf{d}_x)^i \zeta_x^i - log\left(\sum_{i=1}^{I-1} exp\left(\zeta_x^i\right) + exp\left(-\sum_{i=1}^{I-1} \zeta_x^i\right)\right)\right)\right).
\end{aligned}$$

$$\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) = exp\left(w_x \left(\zeta_x^{'} T(\mathbf{d}_x) - b(\zeta_x)\right) + c(\mathbf{d}_x, d_x)\right),$$

where

$$\begin{aligned}
\zeta_x = clr(\mathbf{S}_x)^{-I} &= \left(clr(\mathbf{S}_x)^1, clr(\mathbf{S}_x)^2, ..., clr(\mathbf{S}_x)^{I-1}\right), \\
Y_x = T((\mathbf{d}_x)) &= \left(\frac{d_x^1 - d_x^I}{d_x}, \frac{d_x^2 - d_x^I}{d_x}, ..., \frac{d_x^{I-1} - d_x^I}{d_x}\right), \\
b(\zeta_x) &= log\left(\sum_{i=1}^{I-1} exp(\zeta_x^i) + exp(-\sum_{i=1}^{I-1} \zeta_x^i)\right), \\
w_x &= d_x, \\
c(\mathbf{d}_x, d_x) &= log\left(\binom{d_x}{d_x^1, ..., d_x^I}\right).
\end{aligned}$$

ilr transformation

We consider the *ilr* transformation as described in Section 2.4. The probability function is

$$\begin{aligned}
\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) &= \binom{d_x}{d_x^1, ..., d_x^I} \prod_{i=1}^{I} (S_i^x)^{d_x^i} \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\langle log(\mathbf{S}_x), \mathbf{d}_x \rangle\right) \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(\mathbf{d}_x^{'} \left(VV^{'} + \frac{1}{I}\mathbb{I}_I\right) log(\mathbf{S}_x)\right) \\
&= \binom{d_x}{d_x^1, ..., d_x^I} exp\left(d_x \left(\langle \frac{V^{'}\mathbf{d}_x}{d_x}, ilr(\mathbf{S}_x)\rangle + \frac{1}{I}\langle log(\mathbf{S}_x), 1\rangle\right)\right).
\end{aligned}$$

which can be written as

$$\mathbb{P}(D_x^1 = d_x^1, ..., D_x^I = d_x^I) = exp\left(w_x\left(\zeta_x^{'} Y_x - b(\zeta_x)\right) + c(\mathbf{d}_x, d_x)\right),$$

where

$$\zeta_x = ilr(\mathbf{S_x}),$$

$$Y_x = T(\mathbf{d}_x) = V^{'}\mathbf{d}_x/d_x,$$

$$b(\zeta_x) = log\left(\sum_{i=1}^{I} exp\left(V_{i,.}^{'}\zeta_x\right)\right),$$

$$w_x = d_x,$$

$$c(\mathbf{d}_x, d_x) = log\left(\binom{d_x}{d_x^1, ..., d_x^I}\right).$$

B.2 GLM formulation of log ratio transformations

alr transformation

Regarding the *alr* transformation, the canonical link function has been widely studied as the standard transformation for the multinomial regression :

$$g(u_x)_i = log\left(\frac{u_x^i}{1 - \sum_j u_x^j}\right).$$

Using the equation $B(\zeta_x) = log(1 + \sum_{i=1}^{I-1} exp(\zeta_x^i))$, we have :

$$\nabla_{\zeta_x} b(\zeta_x) = \frac{exp(\zeta_x^i)}{1 + \sum_{j=1}^{I-1} exp(\zeta_x^j)}.$$

Regarding the Jacobian of $g^{-1}(\zeta_x)$, we obtain by differentiating $\nabla_{\zeta_x} b(\zeta_x)$:

$$J_{\zeta_x}(g^{-1}(\zeta_x)) = Cov(\mathbf{D}_x^{-I})_{i,j} = \begin{cases} S_x^i(1 - S_x^i) & \text{if } i = j, \\ S_x^i S_x^j & \text{else.} \end{cases}$$

clr transformation

Here, we consider the centered log ratio transformation and the functions included in the GLM when the natural parameter ζ_x is as follows :

$$\zeta_x = clr(\mathbf{S}_x) = clr_B(\mathbf{S}_x^{-I}),$$

where clr_B is the centered log ratio applied to a vector by omitting its last component, i.e. if $\mathbf{S}_x = (S_x^1, ..., S_x^I)$ is a compositional data, we denote $\mathbf{S}_x^{-I} = (S_x^1, ..., S_x^{I-1})$, and the application is :

$$clr_B(\mathbf{S}_x^{-I}) = clr\left(\left(S_x^1, S_x^2, ..., S_x^{I-1}, 1 - \sum_{i=1}^{I-1} S_x^i\right)\right).$$

Then we have

$$u_x = \frac{E[T(\mathbf{D}_x)]}{D_x} = \frac{E[D_x^i] - E[D_x^I]}{D_x} = S_x^i - S_x^I,$$

and

$$\zeta_x = clr_B(\mathbf{S}_x^{-I}) = clr_B\left(C^{-1}(u_x + 1)\right),$$

with,

$$C_{i,j} = \begin{cases} 2 & \text{if } i = j, \\ 1 & \text{else.} \end{cases}$$

The link function is :

$$g(u_x) = clr_B\left(C^{-1}(u_x + 1)\right).$$

In a GLM, using the *clr* function as a transformation of the probabilities implies that the variable of interest $T(\mathbf{D}_x) = \mathbf{D}_\mathbf{x}^{-\mathbf{I}} - D_x^I$. As a consequence, starting from the covariance structure of a multinomial vector, we obtain :

$$Cov(T(\mathbf{D}_x))_{i,j} = Cov(\mathbf{D}_x^{-I})_{i,j} + \sum_{i=1}^{I-1} S_x^i(1 - \sum_{i=1}^{I-1} S_x^i) + (S_x^i + S_x^j)(1 - \sum_{i=1}^{I-1} S_x^i),$$

with $Cov(\mathbf{D}_x^{-I})_{i,j}$ defined as in the *alr* case.

ilr transformation

Considering the isometric log transformation, the natural parameter is :

$$\zeta_x = ilr(\mathbf{S}_x) = ilr\left(Vu_x + \frac{1}{I}\right) = g(u_x).$$

Starting from the Covariance matrix of the multinomial vector, we can derive the Covariance structure of $T(\mathbf{D}_x)$.
Due to the form of the variable of interest, i.e. $T(\mathbf{D}_x) = V'\mathbf{D}_x$, the Jacobian becomes

$$J_{\zeta_x}(g^{-1}(\zeta_x)) = Cov(T(\mathbf{D}_x)) = V'Cov(\mathbf{D}_x)V.$$

C IWLS algorithm for random vector

We described in this section the algorithm used to obtain the parameters of interest for a GLM with canonical link function. We suppose the number of categories is I, then the number of natural parameters is $I^\star = I - 1$.

For each individual k, we denote by Y_k the observed vector of interest of length $I^\star$, by z_k the p length vector of observed explanatory variables. We introduce a synthetic matrix Z_k which allows to write the model in a matrix form for each individual.

$$Z_k = \begin{pmatrix} 1 & 0 & ... & 0 & z_k^{'} & 0_p & ... & 0_p \\ 0 & 1 & ... & 0 & 0_p & z_k^{'} & ... & 0_p \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 0_p & 0_p & ... & z_k^{'} \end{pmatrix}.$$

The matrix Z_k is pooling the information for each component of the multivariate variable. This matrix is of dimension $I^\star \times (I^\star(p+1))$. In order to link the response variables to the explanatory variables, we define the β, the $I^\star(p+1)$ dimensional vector parameter, such that with canonical link function we have :

$$\eta_k = \zeta_k = Z_k\beta.$$

For a component k of the vector Y_k the parameters that link the explanatory variable z_k to the i-th component are (β_s) for $s \in \{i\} \cup \{I^\star + (i-1)p + 1, ..., I^\star + ip\}$.

The log-likelihood has the form :

$$\sum_{k=1}^{N} l_k(\beta) = \sum_{k=1}^{N} (\zeta_k(\beta)^{'} Y_k - b(\zeta_k(\beta))).$$

For each observation k the gradient of the likelihood is :

$$\nabla l_k(\beta) = d_\beta(\zeta_k(\beta))(Y_k - \nabla b(\zeta_k(\beta))).$$

As $\zeta_k(\beta) = Z_k\beta$, we obtain :

$$d_\beta \zeta_k(\beta)_{s,t} = \frac{\partial(\zeta_k(\beta))_k}{\partial \beta_s} = (Z_k)_{t,s}.$$

Hence, we have :

$$d_\beta \zeta_k(\beta) = (Z_k)^{'}.$$

The gradient function is written as follows :

$$\nabla l(\beta) = \sum_{k=1}^{N} \nabla l_k(\beta) = \sum_{i=1}^{N} (Z_k)^{'}(y_k - \nabla b(\zeta_k(\beta))) = \overline{Z}^{'}(Y - g^{-1}(\zeta(\beta))),$$

where

$$\overline{Z} = (Z_1^{'}, Z_2^{'}, ..., Z_N^{'})^{'},$$
$$Y = (Y_1^{'}, Y_2^{'}, ..., Y_N^{'})^{'},$$

and

$$g^{-1}(\zeta(\beta)) = (g^{-1}(\zeta_1(\beta))^{'}, g^{-1}(\zeta_2(\beta))^{'}, ..., g^{-1}(\zeta_N(\beta))^{'})^{'}.$$

Optimizing the likelihood requires to find the solution of the following equation :

$$\nabla l(\beta) = 0.$$

C.1 Newton-Raphson algorithm

Since a closed form solution is not available, we have to use Newton-Raphson algorithm to estimate the parameters β. The principle of this algorithm is as follows : after initializing the parameter of interest $\beta^{(0)}$, we iterate :

$$\beta^{(t+1)} \leftarrow \beta^{(t)} - (H_\beta(l(\beta^{(t)}))^{-1} \nabla l(\beta^{(t)}),$$

where $\beta^{(t)}$ is the parameter obtained after t iterations.

This algorithm presents a specific form in GLM framework.

First, we notice that :

$$H_\beta(l_k(\beta)) = Z_k^{'} J(g^{-1}(\zeta_k(\beta)) Z_k,$$

where $J_\zeta(g^{-1}(\zeta_k(\beta))$ is the Jacobian of $g^{-1}(\zeta_k)$.

By summing all the Hessian matrices we obtain the following matrix form :

$$\overline{Z}^{'} W(\beta) \overline{Z},$$

where

$$W(\beta) = \begin{pmatrix} J(g^{-1}(\zeta_1)) & 0_{I^\star,I^\star} & ... & 0_{I^\star,I^\star} \\ 0_{I^\star,I^\star} & J(g^{-1}(\zeta_2)) & ... & 0_{I^\star,I^\star} \\ ... & ... & ... & ... \\ 0_{I^\star,I^\star} & 0_{I^\star,I^\star} & ... & J(g^{-1}(\zeta_{I^\star}0)) \end{pmatrix}.$$

and $0_{I^\star,I^\star}$ is the matrix of dimension $I^\star \times I^\star$ of each element is 0.

We detail the process allowing to obtain the Newton-Raphson algorithm which in the case of vector GLM leads to the IWLS (Iterative Weighted Least Square) algorithm. Using a Taylor expansion, and by denoting $\beta^\star$ the optimal parameter, we get :

$$\begin{aligned} \nabla l(\beta^\star) &\approx \nabla l(\beta) - H_\beta(l(\beta))(\beta^\star - \beta) \\ &= \overline{Z}^{'}(Y - g^{-1}(\zeta)) - \overline{Z}^{'} W(\beta)\overline{Z}(\beta^\star - \beta) \\ &= \overline{Z}^{'} W(\beta)(W^{-1}(\beta)(Y - g^{-1}(\zeta)) + \overline{Z}\beta - \overline{Z}\beta^\star) \\ &= \overline{Z}^{'} W(\beta)(M^\star(\beta) - \overline{Z}\beta^\star) \approx 0, \end{aligned}$$

where $M(\beta) = W^{-1}(\beta)(Y - g^{-1}(\zeta)) + \overline{Z}\beta$ is referred as the pseudo-response, W is the matrix of pseudo weights and $\beta^\star$ the parameter to estimate. We recognize in the last row, the solving equation of a Weighted Least Square Error algorithm. The IWLS algorithm consists in iteratively solving the equation :

$$\beta^{(t+1)} \leftarrow (\overline{Z}^{'} W(\beta^{(t)})\overline{Z})^{-1}\overline{Z}^{'} W(\beta^{(t)})M(\beta^{(t)}).$$

We describe the process as detailed in a note of Dutang (2017)[48].

1. Initialization
   (a) We use the response function shifted by a small amount : $u^{(0)} = Y + \epsilon_1$, which allows computing $\eta^0 = g(u^{(0)})$.

   (b) $M^0 = (\eta^0 + W^{-1}(u^0)(Y - u^0)$.

   (c) Calculate $\beta^{(0)} = (\overline{Z}^{'} W(u^0)\overline{Z})^{-1}\overline{Z}^{'} W(u^0)M^{(0)}$.

2. Iterative process :
   (a) $M^{(t)} = (\overline{Z}(\beta^{(t)}) + W^{-1}(\beta^{(t)})(Y - g^{-1}(\beta^{(t)}))$.

(b) $\beta^{(t+1)} = (\overline{Z}^{'} W(\beta^{(t)})\overline{Z})^{-1}\overline{Z}^{'} W(\beta^{(t)})M^{(t)}$.

(c) Using $\beta^{(t)}$, we update the pseudo weights W and pseudo responses M until convergence : i.e. if $||\beta^{(t+1)} - \beta^{(t)}|| < \epsilon_2$, then we stop the algorithm, else we reiterate from the step (b).

Dealing with the P-splines predictors $\overline{Z}$, we have the following likelihood function :

$$max_\beta \sum_{x=x_0}^{x_{\omega_1}} D_x \left[T(\mathbf{D}_x)^{'}(Z_x\beta) - b(Z_x\beta)\right] - \frac{\lambda}{2}\sum_{i=3}^{l}(\Delta^2\beta_i)^2,$$

with Z_x as described in Section 2.4.2. Considering the following $I^\star(p+1) - 2 \times (I^\star(p+1))$ matrix,

$$\mathbf{\Delta}_2 = \begin{pmatrix} 0 & ... & 0 & 1 & -2 & 1 & 0 & ... & 0 \\ 0 & ... & 0 & 0 & 1 & -2 & 1 & ... & 0 \\ 0 & ... & 0 & ... & ... & ... & ... & & \\ 0 & ... & 0 & 0 & 0 & ... & 1 & -2 & 1 \end{pmatrix},$$

we notice that the vector $\Delta^2(\beta) = \mathbf{\Delta}_2\beta$, hence $\sum_{j\geq I^\star+1}(\Delta(\beta)_j)^2 = (\mathbf{\Delta}_2\beta)^{'}\mathbf{\Delta}_2\beta = \beta^{'}\mathbf{\Delta}_2^{'}\mathbf{\Delta}_2\beta$. By following the optimization process detailed above and considering the penalty, we obtain the Jacobian :

$$Z^{'}(Y - g^{-1}(\zeta_x)) - \lambda\mathbf{\Delta}_2^{'}\mathbf{\Delta}_2\beta = 0,$$

where $Y = (T(\mathbf{D}_{x_0})^{'}, ..., T(\mathbf{D}_{x_{\omega_1}})^{'})$, which allows to adjust the IWLS algorithm as follows :

$$\beta^{(t+1)} \leftarrow (\overline{M}^{'} W(\beta^{(t)})Z + \lambda\mathbf{\Delta}_2^{'}\mathbf{\Delta}_2)^{-1}Z^{'}W(\beta^{(t)})M(\beta^{(t)}),$$

where the pseudo response $M(\beta)$ writes

$$M(\beta) = W^{-1}(\beta)(Y - g^{-1}(\theta)) + Z\beta.$$

C.2 P-splines extrapolation

As illustrated in Currie et al. (2004)[33] for the Poisson GLM, the P-splines can be an appropriate method for extrapolation. We consider that the starting age of extrapolation is x_{ω_1} and the final age of the extrapolation range is x_{ω_2}. Because of the extension of the ages considered, we have to include further B-splines, of

which the support is on intervals posterior to x_{ω_1} and which are covering the whole extrapolation age range $[x_{\omega_1}+1, ..., x_{\omega_2}]$. We previously had n B-spline functions to capture the calibration age range $[x_0, ..., x_{\omega_1}]$ and we assume that we now have $n^\star$ B-spline functions to cover the whole age range. Let Y_f be the unknown response variable for the extrapolation age range $[x_{\omega_1}+1, x_{\omega_1}+2, ..., x_{\omega_2}]$.

$$Y_f = \left(T(\mathbf{D}_{x_{\omega_1}+1})', ..., T(\mathbf{D}_{x_{\omega_2}})'\right)'.$$

We denote $B^\star(x) = (B_1(x), ..., B_n(x), B_{n+1}(x), ..., B_{n^\star}(x))'$ the extended vector of B-splines at age x, which includes the new B-spline functions. By splitting the previous vector in two part, one including the former B-splines and the other the new B-spline functions, we obtain $B^\star(x) = \left(B(x)', B_f(x)'\right)'$, where $B_f(x) = (B_{n+1}(x), ..., B_{n^\star}(x))'$

$$Zf_x = \begin{pmatrix} B_f(x)' & 0 & ... & 0 \\ 0 & B_f(x)' & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & B_f(x)' \end{pmatrix}.$$

We build the new regressor matrix :

$$Z^\star = \begin{pmatrix} Z^1 & Zf^1 \\ Z^2 & Zf^2 \end{pmatrix}.$$

where

$$Z^1 = (Z'_{x_0}, ..., Z'_{x_{\omega_1}})',$$

$$Z^2 = (Z'_{x_{\omega_1}+1}, ..., Z'_{x_{\omega_2}})',$$

$$Zf^1 = (Zf'_{x_0}, ..., Zf'_{x_{\omega_1}})',$$

$$Zf^2 = (Zf'_{x_{\omega_1}+1}, ..., Zf'_{x_{\omega_2}})'.$$

Because the new B-splines start being positive after the age x_{ω_1}, Zf^1 is a zero matrix. We now define a matrix U which will only select the data for which information is available, i.e. $U = (I\ 0)$, such that

$$UZ^{\star} = \begin{pmatrix} Z^1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then we modify the IWLS algorithm to obtain β for the new B-splines.

$$\beta^{(t+1)} \leftarrow (Z^{\star})' UW(\beta^{(t)})Z^{\star} + \lambda \left((\mathbf{\Delta}_2^{\star})' \mathbf{\Delta}_2^{\star} \right)^{-1} (Z^{\star})' UW(\beta^{(t)}) M^{\star}(\beta^{(t)}),$$

where $M^{\star}(\beta^{(t)}) = W^{-1}(\beta^{(t)})(Y^{\star} - g^{-1}(\theta)) + Z^{\star}\beta^{(t)}$ and $\mathbf{\Delta}_2^{\star}$ is the extended penalization matrix. As explained in Currie et al. (2004)[33], the second order penality will generate linear extrapolation.

Chapitre 3

Mortality clustering with the K-Lee-Carter model

Abstract

In this paper, we propose a method named the K-Lee-Carter model allowing to cluster different mortality time series. This method produces K different groups on which the applications of the Lee-Carter model leads to the minimization of the mean squared errors. While this algorithm may appear to be complex, we demonstrate that it is actually a particular case of the usual K-centroids approach, with a specific distance function. This equivalence makes the implementation easier and faster, using standard software and packages. Two variations of this method are proposed to overcome the potential erratic results and add expert judgments : the constrained and the penalized K-Lee-Carter model. We provide two applications : the first one the US population by age and sex and the second on the US female population clustered by cause of death and age. In each case, we observe that the K-Lee-Carter model generates, for the same number of groups, better results than the usual arbitrary clusters.

3.1 Introduction

Mortality studies are at a crossroads of several academic disciplines including demography, actuarial sciences and public health studies. All these fields of research have a particular interest in understanding and forecasting mortality. Demographers combine these forecasts with other indicators to derive population dynamics. The actuary estimates the technical risks of mortality and longevity portfolios for insurance companies and pension funds. Public health researchers focus on granular aspects of mortality such as the different causes of death and their relative evolution or the risk factors affecting them. The interest for this subject is not only theoretical but also economic, since large financial amounts are directly linked to the occurrence of deaths, as evidenced by the interest shown by pension funds in this topic. The modelling of mortality using probabilistic models originate from the pioneering work of Gompertz (1825)[64] and Makeham (1860)[95] who introduced methods to estimate death rates on a population for a specific period. With medical innovations and rising standards of living throughout the twentieth century, it soon became apparent that the probability of death at a given age could not remain the same over time. Various methods were then introduced to extrapolate observed mortality over time to obtain a reliable estimate of future mortality. Pollard (1987)[110] lists several methods to achieve this : extrapolation of transformed death rates, extrapolation of the parameters of periodically calibrated survival models (Kannisto, Gompertz, Makeham for example), projection based on a reference population assumed to be in advance in terms of medical innovations. It should also be noted that the type of model depends significantly on the quality of the data. If extensive mortality data are available with individual death realizations and explanatory variables, complex parametric models can be used to exploit all the available information. Unfortunately, most of the time, the practitioner has only access to aggregate data. These data often contain deaths and exposures for specific ages or age groups, as well as characteristics allowing distinction between different populations. The most commonly observed characteristics are gender, socio-professional classes, localization, income level or cause of death.

The approach traditionally adopted (see for example Villegas and Haberman (2014)[130] for forecasting by socio-professional group or Forman et al. (2018)[57] for forecasting by cause of death) is to extrapolate the mortality trends of different groups defined by their characteristics. The extrapolation models involved are diverse and can be broadly identified by their position on a continuum, the

two extremes of which being parsimony and accuracy. Thus, the approach of applying a sophisticated time-series model to each of the age-specific mortalities is a method that may appear attractive in terms of the accuracy. However, the number of parameters and the computation time involved often act as a disincentive, and the technique is considered to be inefficient. On the other hand, the idea of projecting mortality for all observed ages by means of a simple model containing few parameters is a practical and parsimonious approach, but can be deficient in accuracy. In their pioneer paper, Lee and Carter (1992)[114] proposed an elegant and computationally inexpensive approach to model mortality dynamics with a stochastic projection model containing two sets of parameters for ages and one set of parameters for time dynamics. The model summarizes the mortality dynamics for all ages in the data using a univariate time series. The singular value decomposition estimation method allows the model parameters to be estimated quickly and easily. Finally, in terms of accuracy, the variance explained by the model is very satisfactory for developed countries and both genders (see Girosi et al. (2007)[62]). Following Lee and Carter's paper, a number of articles have been published with the aim of improving the model and extending it to different mortality data : Lee (2000)[85] refers to the various applications and extensions of the model, Brouhns et al. (2002)[20] suggest calibrating the model using a Poisson's law rather than a Gaussian law, Renshaw and Haberman (2006)[116] propose to include a cohort effect in the Lee-Carter model. Bergeron et al. (2017)[14] proposed to apply the Lee-Carter model on transformed densities rather than on the log mortality and Curie et al. [32] integrated the Lee-Carter variants in a more general GLM framework. Lee-Carter state-space models have been proposed by Pedroza (2006)[108] and Kogure and Kurachi (2010)[83] for Bayesian estimation and by De Jong and Tickle (2006)[36] for a maximum likelihood estimation. Fung et al. (2017)[58] unify the framework of state-space mortality for the most common mortality stochastic models both with Bayesian and frequentist methods. More recently, Machine learning approaches have been experimented by Richman and Wuthrich (2018)[117] to extend the Lee-Carter model with neural network model.

Despite these remarkable qualities, the Lee-Carter model is not always able to best explain mortality trends. First, the model links all ages to a single mortality dynamic. This may seem inadequate in a stochastic framework when mortalities are independent or may lack of precision. For example, Li et al. (2013)[90] observe that infant mortality observed in previous years has a different trend compared to older ages. Indeed, the decrease in infant mortality tends to decelerate over time, contrary to adult mortality. This phenomenon, known as the rotation phenome-

non, justifies the use of adapted methods. Similarly, Bongaarts (2009)[16] suggests analyzing the mortality dynamics of a population by dividing the population into three age sub-groups : infant, adult or advanced age mortality. One of the major drawbacks of the Lee-Carter model is that mortality shocks are applied similarly at all ages. However, the emergence of certain pathologies, such as HIV or neurodegenerative diseases, leads to a shock to these causes of death implying an non-homogeneous change in mortality by age.

Secondly, when considering numerous characteristics other than age and sex, the question arises as on which data groups applying a Lee-Carter model. Lee (2000)[85] raises this issue when dealing with both genders or populations from different countries. The issue is to consider a single matrix of deaths including all ages and characteristics, or a split of these characteristics into several matrices. The question is not insignificant, since the answer determines the randomness structure of the model and the quality of the adjustment. On the other hand, if one wishes to create clusters from the data, we have to ask if it is necessary to strictly separate the characteristics (causes of death, for instance) or if they can be mixed into age groups in order to improve the fit and the explanatory character of the model. Girosi et al. (2007)[62] applies the Lee-Carter model to each cause separately. However, it is clear that some causes of death share risk factors that explain similar trends.

In this chapter, we propose to answer to the following question : how to define the clusters of characteristics that will lead to a better quality of mortality adjustment for each group and possibly to more accurate forecasts. To do so, we propose a method for grouping characteristics (age, sex, cause of death, localization, and more) with similar dynamics regarding the LC framework. The trends are identified by minimization of the fitting error obtained by applying the Lee-Carter model to a defined number of groups to which the different characteristics are allocated. This unsupervised classification replaces the arbitrary distinction that usually precedes the calibration of Lee-Carter models. This algorithm is inspired from the K-centroids clustering method. Actually, we show that this method is equivalent to a K-centroids method where the distance is defined as the weighted Pearson correlation. The resulting algorithm increases the explanatory character of the model and allows us to identify similar mortality trends, which can, for example, enlighten common risk factors. Clustering methods have been applied several times in mortality analysis. Thus, Vallin et al. (2005)[127], Carracedo et al. (2018)[24] use spatio-temporal measures to find significant cluster countries. In addition, Débon

et al. (2017)[37] and Dolinar et al. (2019)[46] used hierarchical clustering methods on mortality indicators in order to distinguish different geographical areas according to the mortality behavior. But to our knowledge, no attempt has been made to create groups of mortality series according to the death characteristics of the observed population. Clustering methods allowing to group time series with similar patterns exist (see Paparrizos and Gravano (2015)[107] for example) but are not necessarily appropriate in explaining mortality by variance minimization, given that the series are standardized. In our analysis, we keep the series unchanged and propose to find the groups offering the best explanation of mortality by application of the Lee-Carter model. In this way, we combine the flexibility and simplicity of the Lee-Carter model with the search for groups of characteristics that can best discriminate mortality dynamics. Tsai and Cheng (2021)[125] proposed a similar approach for age grouping, using several clustering methods such as Ward's hierarchical clustering, divisive hierarchical clustering, K-means and mixture Gaussian to create clusters in order to achieve better forecasting results. The two main differences between this approach and the one developed in this paper, is that, for our method, the clustering and fitting steps are processed simultaneously and the method is designed exclusively for the Lee-Carter model.

In Section 3.2, we briefly review the Lee-Carter model, and its parameterization on a group combining several characteristics such as age and gender. This consideration leads to the question of dividing these characteristics into subgroups, each of which can be fitted by a Lee-Carter model. After a reminder of the K-centroids algorithm, Section 3.3 introduces the K-LC algorithm allowing to obtain K groups of characteristics offering the best fit by the Lee-Carter model. The link between the K-centroids algorithm and the K-LC is then established, which allows us to improve the computation speed of the clustering algorithm for mortality. Two variants of the K-LC are also proposed to overcome volatility issues : the K-constrained LC and the K-penalized LC. Section 3.4 proposes an application of the K-Lee-Carter and its variants with two examples. The first one is on US mortality data by age and by sex. We want to test if the division usually made between the two genders is the most relevant. The second application is on US females mortality data by age and by cause of death. We shall see that it may be interesting to follow the joint evolution of specific causes of death for specific age groups. Finally, Section 3.5 concludes this paper and proposes a discussion on further developments.

3.2 Methodology

3.2.1 Lee-Carter Model

For the purpose of presenting the Lee-Carter model, we assume mortality data whose only features are ages. We will then extend the model to include other characteristics. Within a given population, we observe the number of deaths $D_{x,t}$ and exposures to death $E_{x,t}$. The variable $D_{x,t}$ refers to the count of individual deceased in the population at risk at age x during the period t. The exposure $E_{x,t}$ represents the time spent by the observed population at age x during the period t. We assume these two variables are available for ages $x = x_1, ..., x_{n_1}$ and periods $t = t_1, ..., t_{n_2}$ with n_1 the number of ages considered and n_2 the number of observation periods. Usual considerations lead us to define the force of mortality $\mu_{x,t}$ as follows :

$$\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}.$$

The Lee-Carter model developed in 1992 fits the logarithm of observed mortality by age and period using three sets of parameters $(\alpha_x)_x$, $(\beta_x)_x$ and $(\kappa_t)_t$, assuming the following relationship :

$$log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t},$$

where $\epsilon_{x,t} \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed error terms. The hypothesis of independence between error terms across ages and periods is an important assumption that allows us to perform least squares estimation. Although it greatly facilitates the calculations, this assumption is unlikely to be valid since mortality shocks are likely to propagate over adjacent ages and years.

The parameters of the Lee-Carter model can be interpreted as follows :

- α_x is the time-independent age level parameter, whose interpretation depends on the constraint imposed on the other parameters.
- κ_t is the time varying parameter that determines the trend of the whole mortality.
- β_x is the sensitivity of mortality at age x to the dynamics κ_t.

The model is estimated by minimizing the sum of the quadratic errors. Other GLM estimation approaches are possible (see for instance Debón et al. (2010)[38] or Currie (2016)[32]), but in this framework, the easiest way to include clustering is to use the euclidian distance. If we define the loss function :

$$L((\alpha_x)_x,(\beta_x)_x,(\kappa_t)_t)) = \sum_x \sum_t (log(\mu_{x,t}) - \alpha_x - \beta_x\kappa_t)^2,$$

hence,

$$\Big((\hat{\alpha}_x)_x,(\hat{\beta}_x)_x,(\hat{\kappa}_t)_t)\Big) = \arg \min_{(\alpha_x)_x,(\beta_x)_x,(\kappa_t)_t} L((\alpha_x)_x,(\beta_x)_x,(\kappa_t)_t)).$$

Without constraints on the parameters, the Lee-Carter model is not identifiable. The usual constraints are :

$$\sum_t \kappa_t = 0,$$
$$\sum_x \beta_x = 1.$$

The first constraint allows us to estimate easily α_x by differentiating the error function.

$$\frac{\partial L((\alpha_x)_x,(\beta_x)_x,(\kappa_t)_t))}{\partial \alpha_x} = \sum_t 2(log(\mu_{x,t}) - \alpha_x - \beta_x\kappa_t) = 2\sum_t log(\mu_{x,t}) - 2n_2\alpha_x = 0.$$

It leads to $\hat{\alpha}_x = \frac{1}{n_2}\sum_t log(\mu_{x,t})$, the temporal mean of the log-mortality.

The solution of the minimization program for the other parameters of the model can be obtained by using the singular value decomposition (SVD) of the matrix of logarithms of the forces of mortality from which are subtracted α_x. Let us denote by Y the centred log mortality matrix, i.e. $Y_{x,t} = log(\mu_{x,t}) - \hat{\alpha_x}$. The development of the quadratic error terms with respect to the vectors κ and β leads to consider as optimal parameters the first eigenvectors of the matrices $Y'Y$ and YY'.
The largest eigenvalues in absolute value of these two matrices are identical. The eigenvectors associated with this eigenvalue are respectively standardized versions of $(\hat{\kappa}_t)_t$ and $(\hat{\beta}_x)_x$. Multiplying the standardized $(\hat{\kappa}_t)_t$ by the first eigenvalue leads to the desired decomposition. We notice that this optimization approach is part of a broader decomposition method. Indeed, it is possible to rewrite the Y matrix using the eigenvectors and eigenvalues of $Y'Y$ and YY'. Let λ_i be the i-th the largest eigenvalue in absolute value of $Y'Y$ or YY', u_i the eigenvector of $Y'Y$ associated with λ_i and v_i the eigenvector of YY' associated with σ_i^Y, then we have the following decomposition :

$$Y = \sum_i \lambda_i u_i v_i'$$

This decomposition is called singular value decomposition or SVD. We used it to compute the parameters of the Lee-Carter model using the eigenvectors associated with the largest eigenvalue. It will be used in Section 3.3 for the K-LC method. We now extend the model by including other features than ages.

3.2.2 The different characteristics

In addition to age, the series of the logarithm of the forces of mortality are likely to differ according to several characteristics such as for instance the sex of the individuals, the localization, socio-professional category, income quantile or cause of death. For each combination of characteristics, it is assumed that there is a single associated mortality series. We then define mortality indicators that depend on these observed characteristics. We use the following notations :

- $\mathbb{F}$ the set of characteristics, for instance $\mathbb{F} = \text{Age} \times \text{Sex} \times \text{Cause of death}$.
- $a \in \mathbb{F}$, the characteristic of the observed mortality.
- $\mu_{a,t}$ the force of mortality associated with characteristic a and observed in period t.
- $log(\mu_a) = (log(\mu_{a,t_1}), log(\mu_{a,t_2}), ..., log(\mu_{a,t_{n_2}}))$ is the log-mortality time series associated with characteristic a.

Usually, the practitioner applies the Lee-Carter model to several groups, these groups being defined arbitrarily beforehand. By "arbitrarily", we mean that the groups were defined by a distinction that seemed natural although not based on a statistical analysis of group dynamics. In order to identify common mortality trends and to reduce the proportion of unexplained variance, we wish to define groups of characteristics such that each group contains mortality series of similar dynamics, suitably fitted by a Lee-Carter model. Formally, the problem consists in distributing the characteristics in homogeneous groups from the point of view of Lee-Carter dynamics. We now denote by J the wished number of groups. We denote the groups we allocate the various characteristics by $C = (C_1, C_2, ..., C_J)$. C is a partition of $\mathbb{F}$, i.e.

- $\forall a \in \mathbb{F}, \exists j,$ such that $a \in C_j$,
- $\cup_{j=1}^{J} C_j = \mathbb{F}$,
- Let i and j such that $i \neq j$, hence $C_i \cap C_j = \emptyset$.

With these notations, the problem is to find the groups $(C_1, C_2, ..., C_J)$ and the

parameters $(\beta_a)_a, (\alpha_a)_a$ and $(\kappa_t^j)_{t,j}$ which minimize the following cost function :

$$L\left(C, (\beta_a)_a, (\alpha_a)_a, (\kappa_t^j)_{t,j}\right) = \sum_j \sum_{a \in C_j} \sum_t (log(\mu_{a,t}) - \alpha_a - \beta_a \kappa_t^j)^2. \tag{2.1}$$

To simplify the clustering algorithm, we define constraints exclusively on the parameters $(\kappa_t^j)_t^j$ of the model : $\forall j$, $\sum_t \kappa_t^j = 0$ and $\sum_t (\kappa_t^j)^2 = 1$. The modification of the constraint does not modify the original fit. Indeed, in the classical Lee-Carter framework, we just have to divide the κ_t by $\sqrt{\sum_t \kappa_t^2}$ and multiply each of the β_x by the same quantity. The result in terms of explained variance will then be identical to the initial fit, and we can rewrite the algorithm of the K-LC in a simpler form to implement.

3.3 The K-Lee-Carter

3.3.1 The K-centroids method

In this section, we present the clustering algorithms for J distinct classes (see Macqueen (1967)[94] for more details). The aim of a clustering algorithm is to identify similarities between individuals and to group them into homogeneous classes. Individuals are vectors whose components are called variables. In our framework, individuals are log-mortality time series, each identified by a unique characteristic. The variables are the realizations of the mortality series over the different periods. The similarity of an individual to its class is defined by means of a distance function and an element called a centroid. The lower is the distance between the individual and the centroid, the closer is the individual to its group. This element summarizes the characteristics of its class. An important distance between two individuals makes the likelihood lower to observe these individuals in a same group. In the opposite, a low distance between two individuals suggests that the individuals are almost identical and should be grouped together.

Let us denote by y_a the individuals with respect to the characteristic a, by $C_1, ..., C_J$ the groups previously defined and by $(c_1, ..., c_J)$ centroids. $d(.,.)$ defines the positive distance function. Even if the K-centroids algorithm is often assimilated to the Euclidean distance, its fundamental principle is valid with any type of distance. The K-centroid algorithm is the following :

- Step 0 : Allocate individuals within J groups $(C_1, ..., C_J)$ according to a defined rule or randomly.
- Step 1 : for each group C_j, one calculates the centroid of the class as follows :

$c_j = \arg\min_c \sum_{a \in C_j} d(y_a, c)$.

- Step 2 : we allocate the individuals within the different classes by comparing them to the centroids. For an individual y_a, $j = \arg\min_i d(y_a, c_j)$ and y_a is assigned to C_j.
- Step 3 : If no individual has experienced a class change then the algorithm is stopped. Otherwise, we start again from step 1.

At each step, the sum of the distances to the centroids $\sum_j \sum_{a \in C_j} d(y_a, c_j)$ decreases or remains stable. Given the limited number of combinations, the algorithm is converging. However, there is no guarantee that the resulting minimum is a global minimum. It is therefore necessary to run the algorithm using several initial allocations in order to obtain the wished solution.

3.3.2 The K-Lee-Carter and its relationship with the K-centroids

This section presents the clustering algorithm, named the K-Lee-Carter which minimize the loss function L defined in equation 2.1.

- Step 0 : Allocate the individuals within the J groups $(C_1, ..., C_J)$ according to a defined rule or randomly.
- Step 1 : for each group C_j, we calculate the parameters : $((\widehat{\alpha}_a)_{a \in C_j}, (\widehat{\beta}_a)_{a \in C_j}, (\widehat{\kappa}_t^j)_t) = \arg\min \sum_{a \in C_j} \sum_t (log(\mu_{a,t}) - \alpha_a - \beta_a \kappa_t^j)^2$.
- Step 2 : for each $log(\mu_{a,t})$ and each κ^j, we calculate the coefficients $\widetilde{\alpha}_a^j$, $\widetilde{\beta}_a^j$ minimizing $L_a^j(\alpha_a^j, \beta_a^j) = \sum_t (log(\mu_{a,t}) - \alpha_a^j - \beta_a^j \widehat{\kappa}_t^j)^2$.
- Step 3 : The time series associated with the characteristic a switches to the group j that best matches its dynamics, i.e. $C(a) = C_j$ with $j = \arg\min_i L_a^i(\alpha_a^{i,\prime}, \beta_a^{i,\prime})$.
- Step 4 : if no individual has experienced a group change then the algorithm is stopped. Otherwise, the algorithm is reiterated from step 1.

We now demonstrate that this algorithm can be written as a K-centroids algorithm, for which we have to define the distance. This rewriting allows us to merge steps 2 and 3 into a single step that is faster in terms of computation time. Moreover, since the algorithm is nothing more than a K-centroid method, we are then assured of the convergence of the K-LC algorithm to a local minimum. Thanks to the constraint $\sum_t \kappa_t^j = 0$, we can avoid the computation of the parameter α_a, as shown in Section 3.2.1. We now denote :

$$y_{a,t} = (log(\mu_{a,t}) - \hat{\alpha}_a)$$
$$y_a = (y_{a,t_1}, y_{a,t_2}, ..., y_{a,t_{n_2}})$$

Let's assume we are at stage 1, then :

$$\sum_{a\in C_j}\sum_t (y_{a,t}-\beta_a\kappa_t^j)^2 = \sum_{a\in C_j}\sum_t y_{a,t}^2 - 2\sum_{a\in C_j}\beta_a\sum_t y_{a,t}\kappa_t^j + \sum_{a\in C_j}\beta_a^2\sum_t(\kappa_t^j)^2$$

For a given κ^j by applying simple linear regression, the parameter β_a which optimizes the above equation is :

$$\beta_a = \frac{\sum_t(\kappa_t^j-\overline{\kappa}^j)(y_{a,t}-\overline{y}_a)}{\sum_t(\kappa_t^j-\overline{\kappa}^j)^2}.$$

Due to the fact that $\overline{y}_a = n_2(\alpha_a-\alpha_a) = 0$ and because of the constraints $\sum_t \kappa_t^j = 0$ and $\sum_t(\kappa_t^j)^2 = 1$, we obtain :

$$\beta_a = \sum_t \kappa_t^j y_{a,t}.$$

By exploiting the constraints and denoting $Var(y_a) = \frac{1}{n_2}\sum_t \left(y_{a,t}-\overline{y}_{a,t}\right)^2$ and $Cov\left(\kappa^j, y_a\right) = \frac{1}{n_2}\sum_t \left(y_{a,t}-\overline{y}_{a,t}\right)\left(\kappa_t^j-\overline{\kappa}^j\right)$, respectively the empirical variance and covariance functions, we obtain :

$$\frac{1}{n_2}\sum_{a\in C_j}\sum_t(y_{a,t}-\beta_a\kappa_t^j)^2 = \sum_{a\in C_j}Var(y_a) - \sum_{a\in C_j}Cov(\kappa^j,y_a)^2 = \sum_{a\in C_j}Var(y_a)(1-cor(\kappa^j,y_a)),$$

with $Cor(\kappa^j, y_a)) = \frac{Cov(\kappa^j,y_a)^2}{Var(\kappa_j)Var(y_a)}$.

The distance is therefore defined as

$$d(x,y) = Var(x)Var(y)(1-Cor(x,y)). \tag{3.1}$$

It can be shown that minimizing this distance from κ^j, that is, maximizing $\sum_{a\in C_j} cov(\kappa^j, y_a))$ does indeed correspond to a problem of Rayleigh's Quotient (see Golub and Van Loan (2013)[63]) which can be solved by decomposition into singular values. If we denote by Y^j the matrix of the mortality series of the group j, then the solution of $\sum_{a\in C_j} cov(\kappa^j, y_a))$ is :

$$\kappa^j = \arg\min_x x'(Y^j)'Y^j x,$$

under the constraint $x'x = 1$.

The solution of such a problem is as expected the standardized eigenvector of the matrix $(Y^j)'Y^j$ associated to the largest eigenvalue in absolute value. This

solution is therefore the one resulting from the singular value decomposition of Y^j. Thus, the algorithm of the K-Lee-Carter is equivalent to the algorithm of the K-centroids with a measure of distance $d(x,y) = Var(x)Var(y)(1-cor(x,y))$ and a constraint on the centroids $(\kappa^j)_{j=1,...,J}$.

3.3.3 Constraint and penalized clustering

In some applications, we may include prior information to constrain the classification. For example, qualitative studies can show that certain time series are closely linked in their dynamics. In addition, sampling fluctuations can lead to incoherent classification. For this reason, we may wish to artificially constrain some series to belong to the same group. We present two approaches using either a strong constraint and a weak one also known as penalization. As soon as we incorporate prior information on groups, we talk about a penalized clustering problem.

Approach using strong constraint

The strong constraint consists in associating pairs of series that we want to group together or oppositely to separate. This approach has been proposed by Wagstaff et al. (2001)[132] and widely used in practice. A constraint, called the must-link constraint, imposes that both individuals belong to the same group. In contrast, a cannot-link constraint impose that two individuals to belong to two distinct groups. In order to apply such constraints, we follow the procedure proposed by Leisch (2020)[86].

We define $G_N = \{g_1, ..., g_N\}$ where $g_n \in \{1, ..., M\}$ a preliminary classification of the various time series within M groups, and we denote by $|G_N|$ its cardinal. We define thereafter specific constraints for each of these groups. It is also assumed that defined groups do not share common elements. We also add a last group G_{m+1} including all the series not being subjected to any constraint.

The constraint O_m is associated with a group and implies either that the members of this group must be placed in different classes, or oppositely that they must belong to the same class. The algorithm is then as follows :

- Step 1 : Groups are initialized either deterministically or randomly. Individuals are then assigned to the centroids to which they are the closest, regardless of the previously defined constraints.
- Step 2 : Centroids are updated by minimizing their distance from cluster members.

- Step 3 : For each group of constraint G_m, perform the following procedures according to the nature of the constraint :
  O_m = must-link : assign all points of G_m to the group with the smallest distance from the centroid,
  O_m = cannot-link : allocate the points of G_m in the $|G_m|$ groups which minimizes the distance to the centroids.
- Step 4 : If there is convergence, then the algorithm stops, otherwise we start again from step 2.

It is desirable to have such constraints when dealing with small populations for which the age dynamics are too volatile for consistent clustering. For example, mortality series can be grouped into 5-ages group to avoid erratic clustering for adjacent ages. Moreover, this approach allows including expert opinions regarding the defined clusters.

A penalization approach

We have seen previously an approach introducing a strong prior. Another approach described as weak by contrast, may also be interesting. It is based on an incentive to group together series that are considered to belong to the same group. Our approach consists in penalizing the heterogeneity of the classes defined by the prior. Although there is a large literature on penalized mixing models and unsupervised classification (see for example Xie et al. (2008) [136]), we have not found such approaches in the theoretical framework of K-centroids. The approach consists in introducing prior information to the individual vectors (here the time series) and to modify the objective function to include this new component without disturbing the estimation of the parameter κ^j. We provide a prior group for each time series. Let us denote by H^a the vector modelling the group belonging to the individual i. $H^a_j = 1$ if the individual a belongs to the group j and 0 otherwise. We define a distance $d^P(\cdot,\cdot)$ measuring the heterogeneity of the class with respect to the predefined groups, and a penalty parameter λ. The principle of our approach is to minimize the following objective function :

$$L = \sum_j \sum_{a \in C_j} \sum_t (log(\mu_{a,t}) - \alpha_a - \beta_a \kappa^j_t)^2 + \lambda \sum_j \sum_{a \in C_j} d^P(H^a, H^j).$$

H^j is the vector that minimizes the term $\sum_{a \in C_j} d^P(H^a, H^j)$. If $d^P(\cdot,\cdot)$ is the Euclidean distance, then H^j is the vector whose k-th coordinate represents the proportion of individuals initially placed in the group k that end up together in group j. The more heterogeneous the group is, the more the penalty increases. The

penalization allows series to move into a group with fewer associated series only if the improvement induced by this change exceeds the penalty increase. So the new algorithm becomes :

- Step 0 : Allocate individuals within J groups $(C_1, ..., C_J)$ using a defined rule or randomly.
- Step 1 : for each group C_j, we calculate the parameters : $((\widehat{\alpha}_a)_{a\in C_j}, (\widehat{\beta}_a)_{a\in C_j}, (\widehat{\kappa}_t^j)_t, H^j) = \arg\min \sum_{a\in C_j} \sum_t (log(\mu_{a,t}) - \alpha_a - \beta_a\kappa_t^j)^2 + \lambda \sum_{a\in C_j} d^P(H^a, H^j)$.
- Step 2 : for an individual $log(\mu_{a,t})$, we calculate for each of the κ^j, the coefficients $\widetilde{\alpha}_a^j$, $\widetilde{\beta}_a^j$ minimizing $L_a^j(\alpha_a^j, \beta_a^j) = \sum_t (log(\mu_{a,t}) - \alpha_a^j - \beta_a^j\widehat{\kappa}_t^j)^2$.
- Step 3 : the individual a changes to the group j such that $j = \arg\min_i L_a^i\left(\widetilde{\alpha}_a^j, \widetilde{\beta}_a^j\right) + \lambda d^P(H^a, H^j)$.
- Step 4 : if no individual has experienced a group change then the algorithm is stopped. Otherwise, the algorithm is reiterated from step 1.

3.4 Application

3.4.1 Data presentation and validation criteria

We present two applications based on data related to the US population mortality. The first illustration applies the K-LC algorithm to the all-causes mortality split by gender. We investigate whether the distinction usually made between the two genders for the application of the Lee-Carter model is the most appropriate. The second application concerns mortality data by cause focusing on the female population. For these two examples, we study the period from 1980 to 2012. We then make forecasts from the resulting Lee-Carter models to see if the method leads to improvements in forecasting for the years 2013 to 2017. The parameters κ^j are extrapolated using an ARIMA model.

In order to evaluate the quality of the calibration and the forecasts, we use three measurement criteria : Mean Square Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) whose formulas are given below.

$$MSE(y^1, y^2) = \frac{1}{n_2}\sum_{t=1}^{n_2}\left(y_t^1 - y_t^2\right)^2,$$

$$MAE(y^1, y^2) = \frac{1}{n_2}\sum_{t=1}^{n_2}\left|y_t^1 - y_t^2\right|,$$

$$MAPE(y^1, y^2) = \frac{1}{n_2} \sum_{t=1}^{n_2} \left| \frac{y_t^1 - y_t^2}{y_t^1} \right|.$$

For each application, we compare Lee-Carter models fitted on clusters defined by the K-LC with an application of Lee-Carter on arbitrary clusters, frequently met in practice. The validation criteria are computed by considering the logarithmic transformation of the forces of mortality. We also apply the variants with strong "must-link" and penalization constraints for the application relative to the causes of death.

The death counts are taken from the CDC (Centers for Disease Control) database, while the exposures used to calculate the forces of mortality are taken from the HMD (Human Mortality Database). The forces of mortality are available by cause, sex and age attained for each calendar year. Table 3.4.1 below reconstructs the data used in this study from the CDC database. Each death is recorded by a physician using the International Classification of Diseases (ICD) nomenclature to identify the cause of death. As the pathology recording system is very detailed, it leads to a high level of granularity which is not well suited to the construction of robust statistics. Since our attention is focused on the most important causes of death, we group the different causes of death into fewer classes. We provide the transcription codes to reconstruct these classes from the original database. These codes are different for ICD 9 and 10, which reflect changes in nomenclature as new medical considerations are incorporated.

Cause of Death	1979 to 1998 (ICD 9 : UCOD 282)	1999 to 2017 (ICD 10 : UCOD 352)
Cancer	04600 – 11800	69 - 146
Metabolic	11900 – 13200	156-173
Nervous and Mental	14100 – 15900	174-194
Circulatory system	16000 – 20300	197-246
Respiratory system	20400 – 22700	247-278
External	29900 – 35800	381-456
Others	All others	All others

Table 3.4.1 – *Cause of death classification according ICD 9 and 10.*

3.4.2 Clustering by age and gender

In the first application, we observe all-causes mortality for the US population according to two characteristics : age and sex. The observations include ages from 21 to 85. The reference cluster is obtained by applying the Lee-Carter model to

both genders separately. Figure 3.4.1 below shows the parameters κ_t^i for each sex i, as well as the centered logarithmic series of each group.

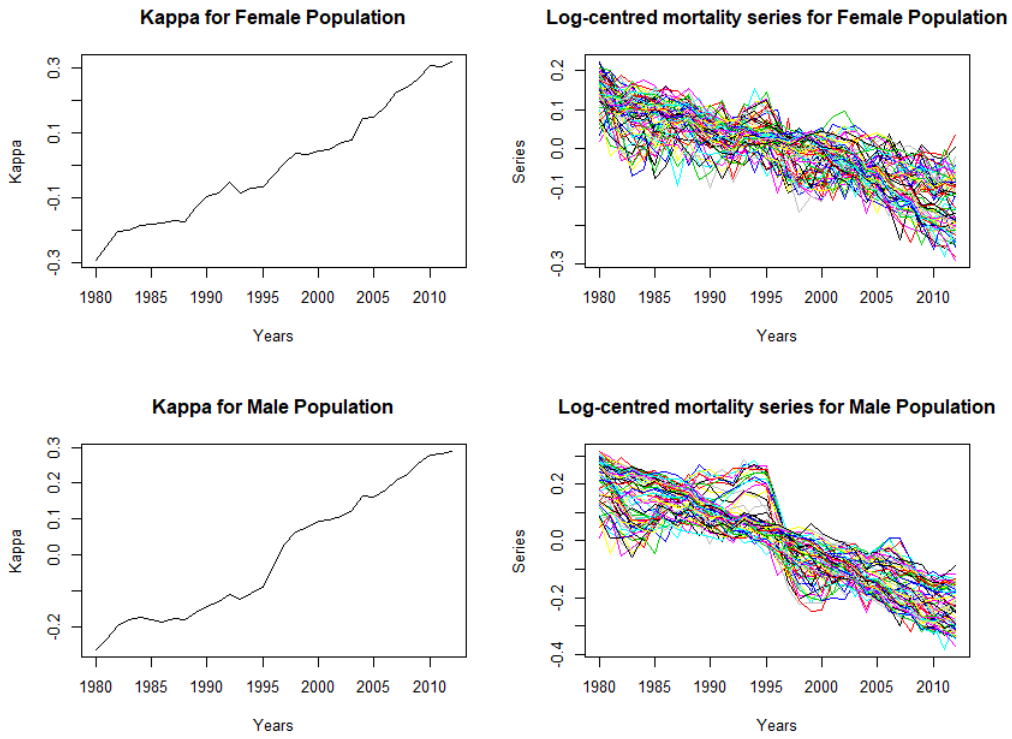


FIGURE 3.4.1 – *Mortality trends by gender*

We find that the female and male series show on average a common linear trend, as suggested by the similarity between the respective time parameters, see figure 3.4.1, upper and bottom left panels. Nevertheless, some male series seem to deviate from the linear trend, which is not as evident for the female population.

We now apply the K-LC algorithm on a database including both genders. Figure 3.4.2 shows the distribution of individuals according to groups. The abscissa shows the genders, the ordinate shows the ages and the groups are indicated by a different color. Two groups are identifiable. The first group highlighted in blue contains male mortality between 26 and 45 years of age, and female mortality between 25 and 40 years of age. It seems to have undergone a particular evolution during the period from 1980 to 2012. The rest of the mortality series are included in the second group. In order to better understand what distinguishes these groups, we propose to examine the resulting κ_t^i parameters, as well as the series associated to these dynamics.

We can see in Figure 3.4.3 that the population is divided into two groups with very distinct trends. The first group consists of the series that do not follow a

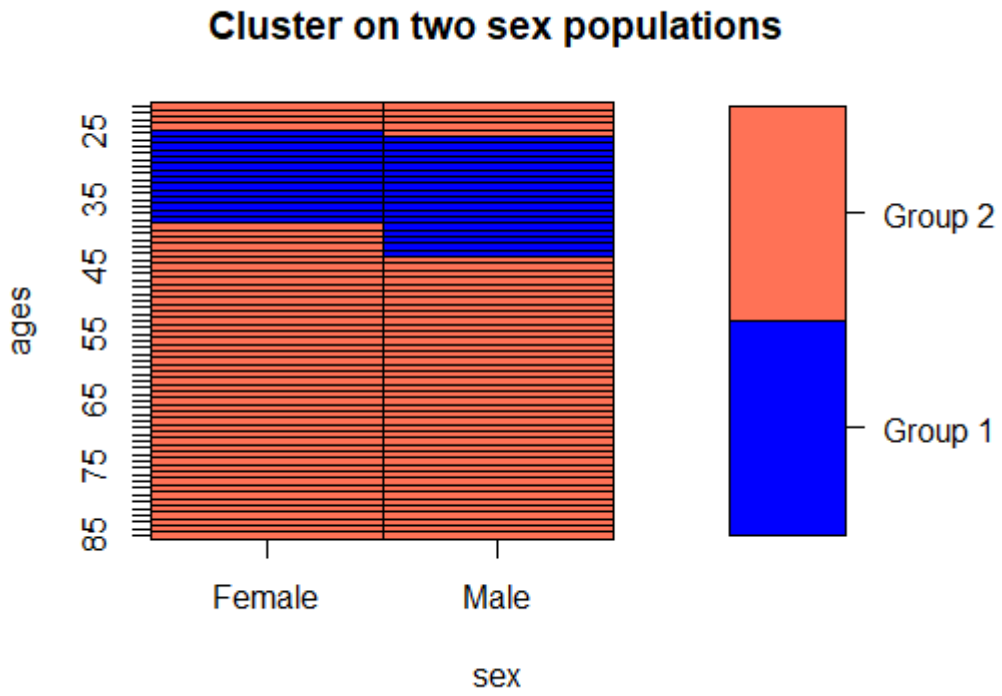


FIGURE 3.4.2 – *Clustering in two groups for Male and Female populations*

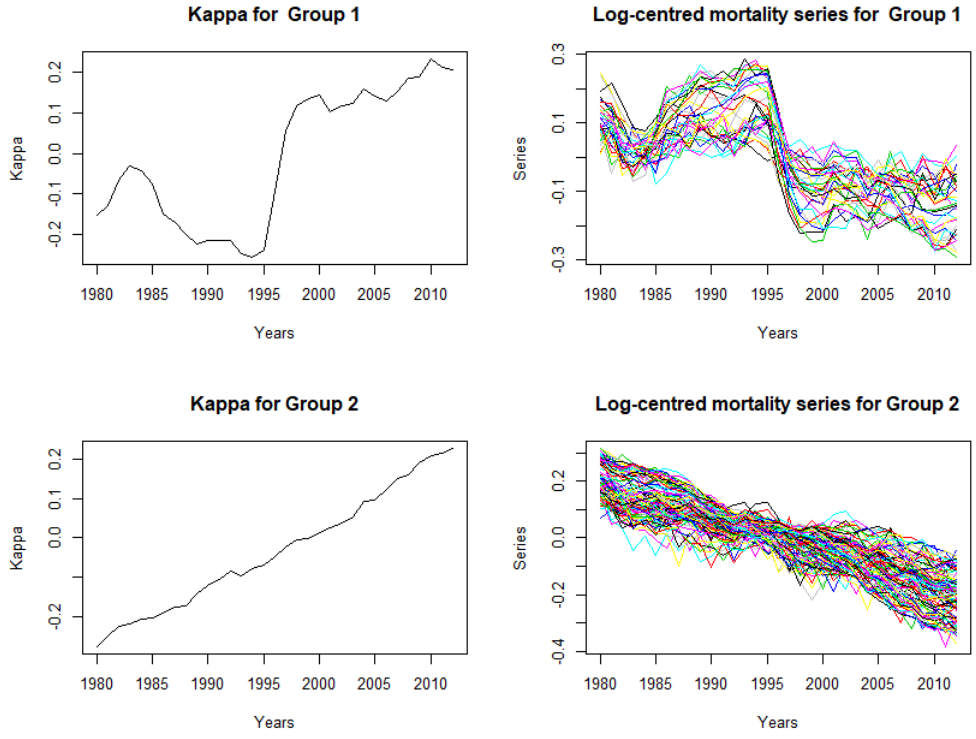


FIGURE 3.4.3 – *Mortality trends by groups obtained by K-LC*

linear pattern and show an increase in the years 1985 to 1995 before decreasing and then stagnating or slightly increasing. The model detect the young population which suffers from the HIV epidemic and is mainly composed of young people. This mortality rate has stopped decreasing by the end 1990, due to the diffusion of effective treatments. The second group includes mortality series that appear to follow a linear path. We then project the time varying parameters obtained on the reference clustering and on the clustering obtained by K-LC using an ARIMA model. Table 3.4.2 compares the quality of the calibration of the LC model for the reference groups (Male vs Female) and the groups resulting from the clustering by K-LC.

Cluster	MSE	MAE	MAPE
Reference cluster	0.002425	0.03715	0.007663
K-LC cluster	**0.001537**	**0.03022**	**0.006266**

Table 3.4.2 – *In-sample errors*

Cluster	MSE	MAE	MAPE
Reference cluster	0.01373	0.08237	0.01510
K-LC cluster	**0.01060**	**0.07706**	**0.01464**

Table 3.4.3 – *Out-sample errors*

Adjustment errors are lower for groups obtained from the K-LC. This result was expected because the MSE criterion is the one used to calibrate the parameters and constitute the groups. When observing the forecasts, we can also see a gain in accuracy for all error measures. The method is therefore not only more efficient for the adjustment but also improves the predictions. The results are sufficiently regular (in the sense that there is no alternation of groups for adjacent ages) so that we do not have to consider methods of constraint by penalization or "must-link". Applications of these methods are illustrated in the next example.

3.4.3 Application to the US female population by cause of death

In the second application, we observe the female mortality for the six causes of death selected in Table 3.4.1 (we exclude the cause "Others" from our analysis in order to lighten the graphs). The ages studied for each cause range from 41 to 80 years. This choice is due to the lack of mortality data for some causes at young ages. To carry out this application, we first observe the trend parameters obtained for a Lee-Carter model applied separately to each cause. Each of the parameters κ^j thus obtained is fitted by means of an ARIMA model and extrapolated. We present three successive applications of the K-LC model on the female cause of death mortality using an approach without constraints, with strong constraints and including a penalization.

Application of Lee-Carter to specific causes of death for the US female population

Applying the Lee-Carter model to each of the causes separately results in the time varying parameters presented in Figure 3.4.4.

Firstly, we can see that the time varying parameters obtained are very different from each other. While some appear to follow a linear pattern similar the observed overall mortality, other trends seem to have a totally different shape. This is the case for external and respiratory causes of death, for which the associated parameters seem to show clear changes in trend.

Application of the K-Lee-Carter on the US females cause of death mortality

We apply the unconstrained K-LC method to the US female mortality by cause of death. Figure 3.4.5, presents the resulting clustering. In order to better visualize the distribution of groups within causes and ages, we propose the following. The x-axis represents the cause of death while the y-axis corresponds to the attained age.

From Figure 3.4.5, several major groupings seem to stand out. The first group includes cancer mortality between the ages of 41 and 60, mortality from cardiovascular disease across most ages, and mental illness in advanced ages. Another group includes cancers between the ages of 60-75 and external diseases between the ages of 41-55. A third group consists of respiratory and external diseases at ages 61-80. The other groups have either scattered series or a single cause for a specific age

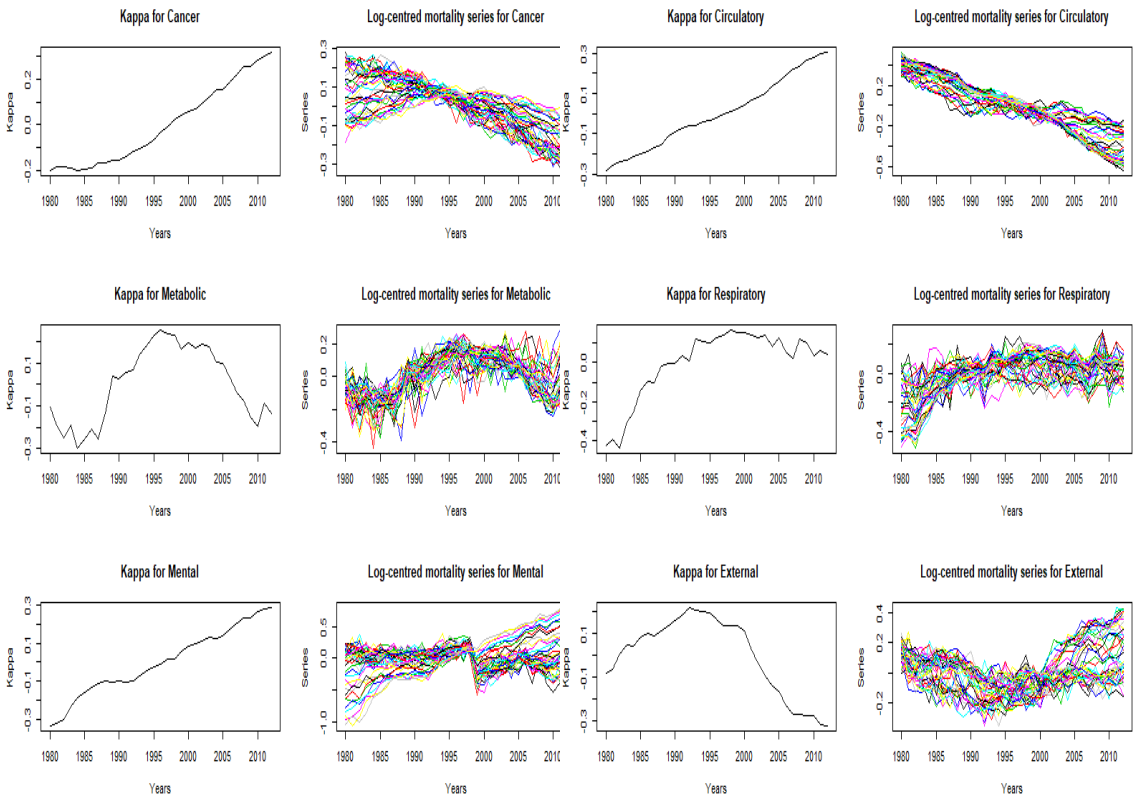


FIGURE 3.4.4 – *Trends for each cause and associated centred series*

group and are therefore of less interpretative interest. Figure 3.4.6 presents the trend of the time varying parameters of the resulting clusters as well as the series associated with each of the parameters.

We observe that Group 1 corresponds to mental and nervous illnesses between 50 and 67 years of age, who experienced a jump between 1998 and 1999. This year corresponding to change from ICD 9 to ICD 10, it is possible Group 2 includes trends whose dynamics seem linear, notably mental illness after age 68, almost all circulatory diseases and cancers up to age 60. Groups 3 and 5 concern mortality series that experienced a change in trend, namely cancers after age 60, a large proportion of mortality due to metabolic diseases, as well as mortality by external causes up to about age 61. The fourth group contains series whose dynamics seem to stagnate after 1995. It contains a large proportion of deaths from respiratory diseases (especially after age 60), and deaths by external causes after age 60. Finally, Group 6 contains some relatively erratic series, most of which are mental illnesses at early ages (about 41 to 50 years). Because of the erratic nature of the mortality series, some groupings are difficult to read. This is particularly the case

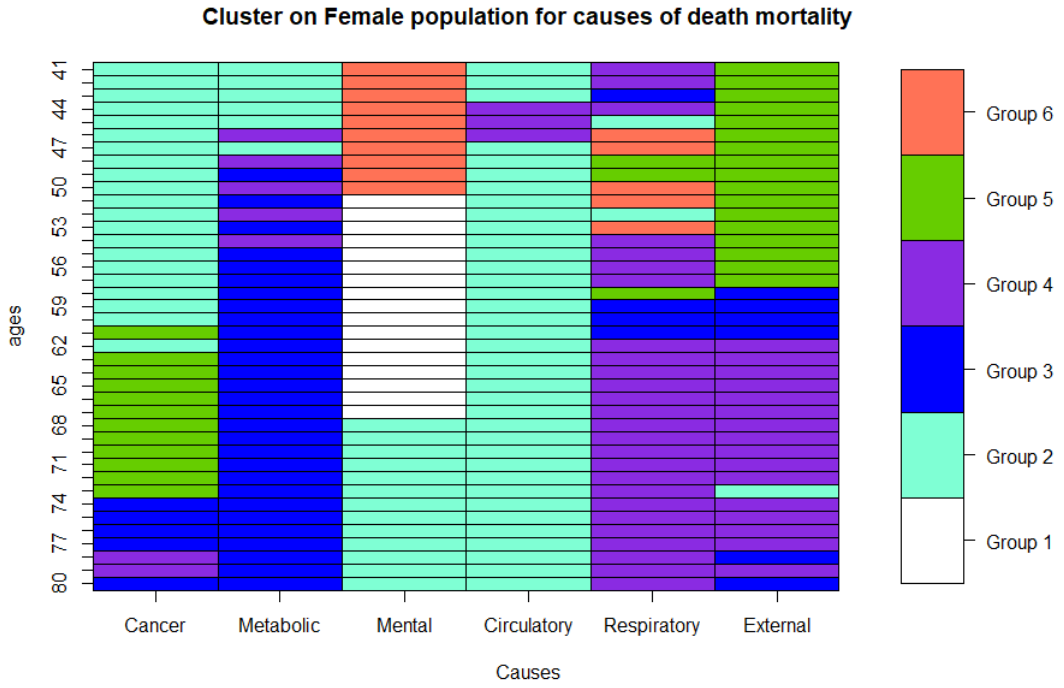


Figure 3.4.5 – *Clustering for Female population by causes of death*

for respiratory diseases before the age of 60, for which 5 different groups are present. To overcome this issue, we apply the K-LC with constraints and penalization.

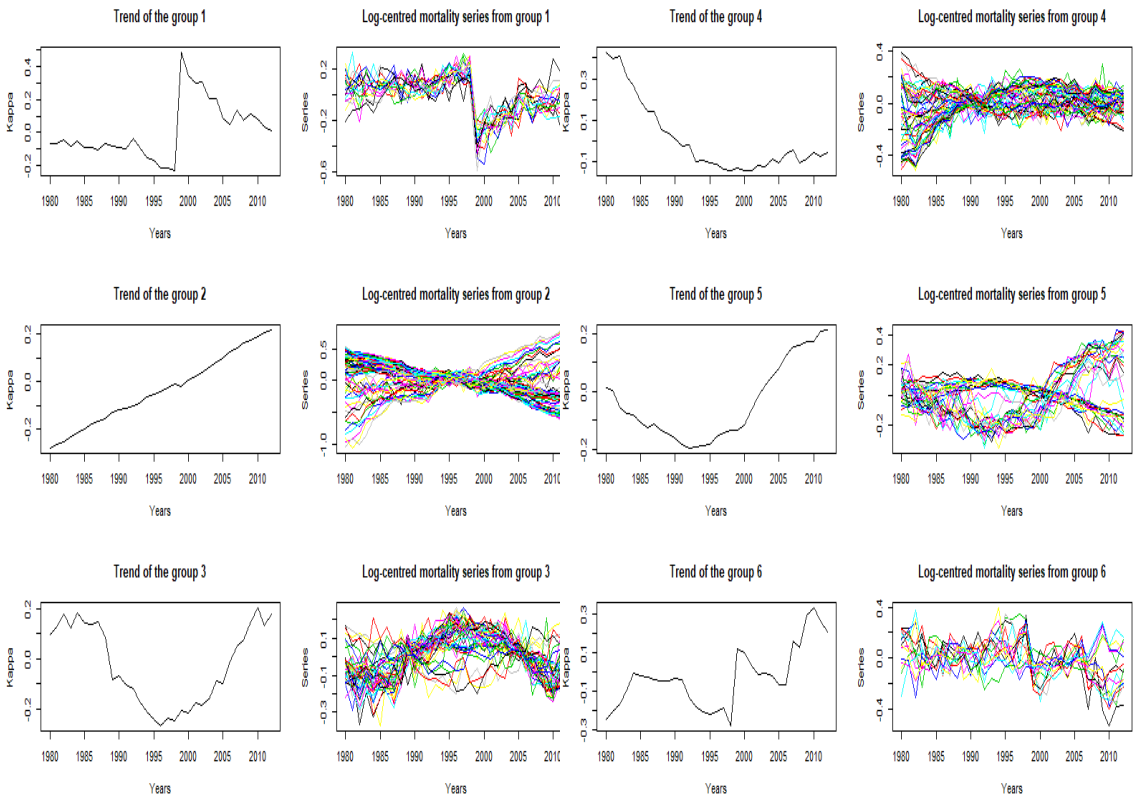


FIGURE 3.4.6 – *Trends for each group and associated centred series*

Here we have imposed constraints on the female population in 5 year age groups. For each of the causes, mortality time series are attached by groups of 5 using the "must-link" constraint technique detailed in Section 3.3.3.

Figure 3.4.7 presents the resulting clusters obtained with constraints. The groups resulting are similar to those obtained with the unconstrained method, but the groups seem less erratic regarding ages and causes of death. It should be noted, however, that the constraint method revealed similar mortality between mental illness at young ages and respiratory illness between 45 and 55 years of age. We can see from Figure 3.4.8 that the dynamics of the time varying parameters are not significantly modified by the constraint.

We now apply the K-LC method with penalization. For this purpose we define the distance $d^P(.,.)$ as the Euclidean distance. In this way, the vector H^j obtained at each step summarizes the distribution of the predefined groups within the group C^j. The parameter λ can be chosen arbitrarily or by other means such as cross-validation. The number of group J is here defined ex-ante. Again, other methods can be used to define this number (cross validation, BIC and AIC criteria), but

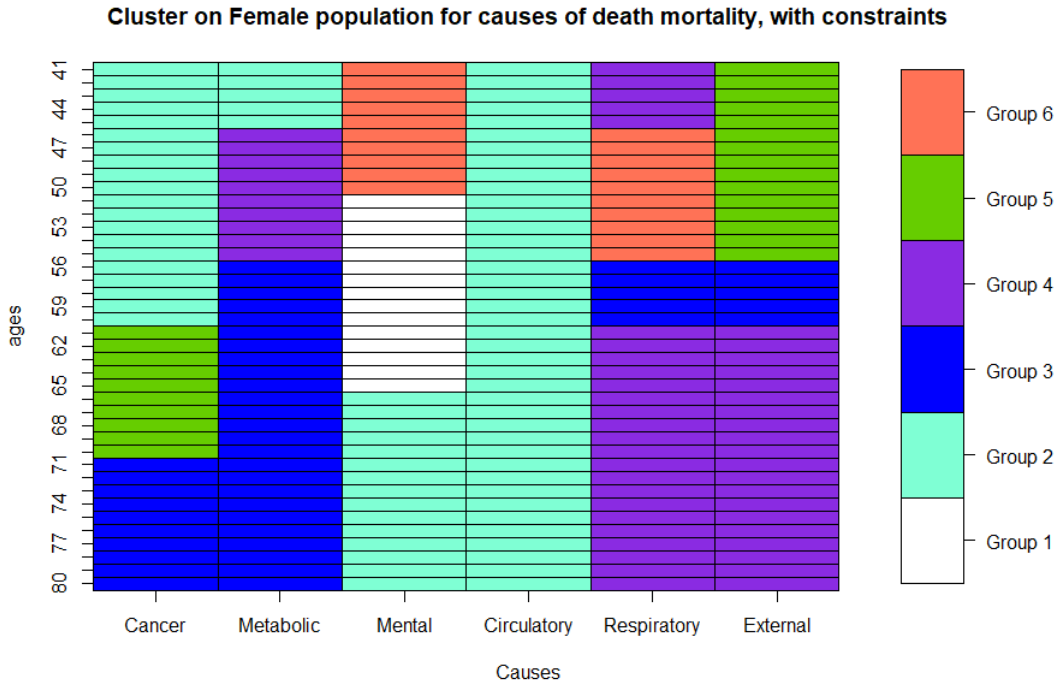


FIGURE 3.4.7 – *Constrained clustering for Female population by causes of death*

this is not the purpose of this chapter.

Several tests have been performed empirically in order to select a satisfactory penalty parameter λ. We propose for the female population the parameter $\lambda = 0.008$. Figure 3.4.9 illustrates the groups thus obtained.

This third application provides intermediate clusters compared to the method constrained by five-year age groups and the unconstrained method. As Figure 3.4.10 illustrates, the changes resulting from the penalization do not lead to a significant difference in the parameters κ_t^i of the model.

For each of the clustering methods, an ARIMA model is associated to the temporal parameters κ_t^i before projecting them. By combining this projection to the parameters α_x^i and β_x^i, we obtain an estimate of the future mortality for each group. Finally, we compare the results obtained on each clustering model to the fits resulting from the arbitrary grouping by cause of death. Table 3.4.4 lists the calibration errors according to the method used.

The K-Lee-Carter provides a better calibration results than cause separate clustering. The approach not only leads to the smallest MSE criterion, as expected due to the nature of the objective function to be optimized, but smaller MAE and MAPE criteria as well. Comparison of the three variants of the K-LC shows relatively similar performance for the three criteria. Naturally, the unconstrained

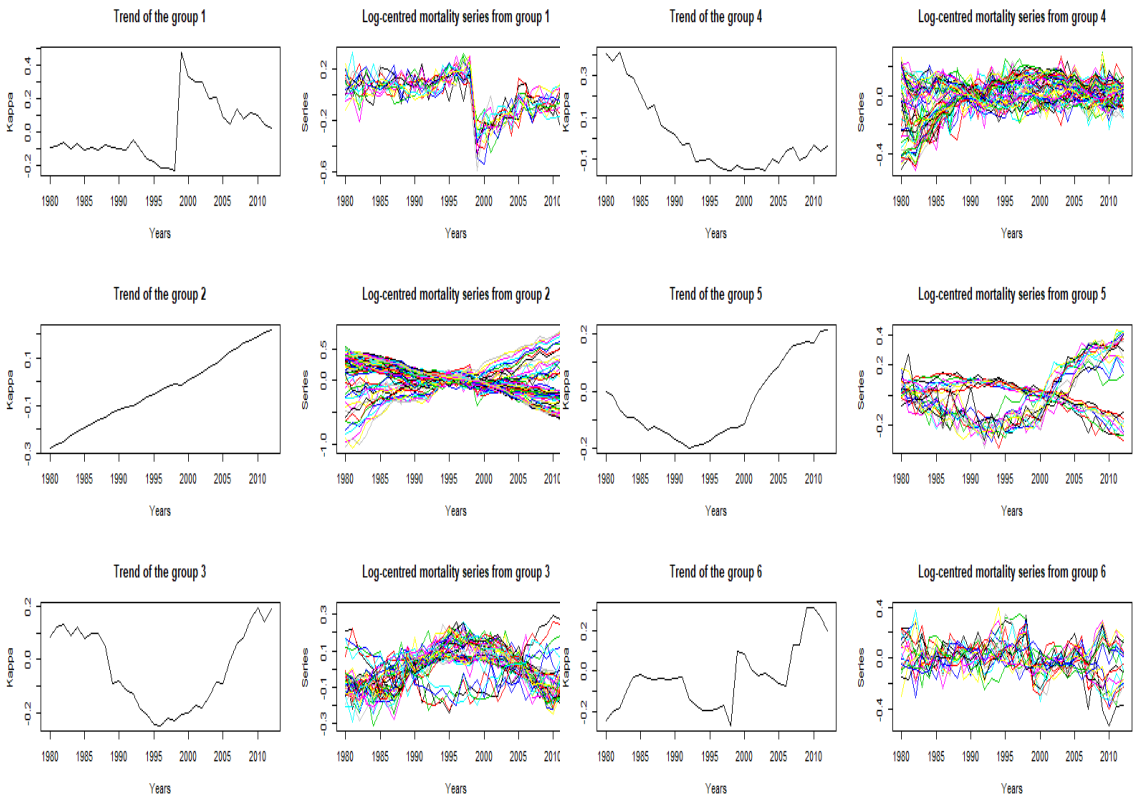


FIGURE 3.4.8 – *Trends for each group and associated cantered series*

Cluster	MSE	MAE	MAPE
Reference cluster	0.005571466	0.05283739	0.007000738
K-LC	**0.003887775**	**0.0465221**	**0.006256402**
K-LC with constraints	0.00401991	0.04733094	0.006379771
K-LC penalized	0.00398385	0.04723434	0.006371326

TABLE 3.4.4 – *Errors in the calibration sample*

model is slightly better performing in that its calibration only takes into account the quadratic errors and not ancillary criteria such as grouping or penalization. Table 3.4.5 presents the errors on the test sample to see whether mortality by cause can be better projected by grouping similar dynamics than by considering the causes of death independently.

The results in Table 3.4.5 show the best performance of the clustering methods on this sample. Indeed, we observe that for all the criteria, the groups constituted by K-LC show lower values than those resulting from the reference method. For all criteria, the error is slightly lower when adding the constraint or the penalty. Finally, between the constrained and penalized methods the differences are small,

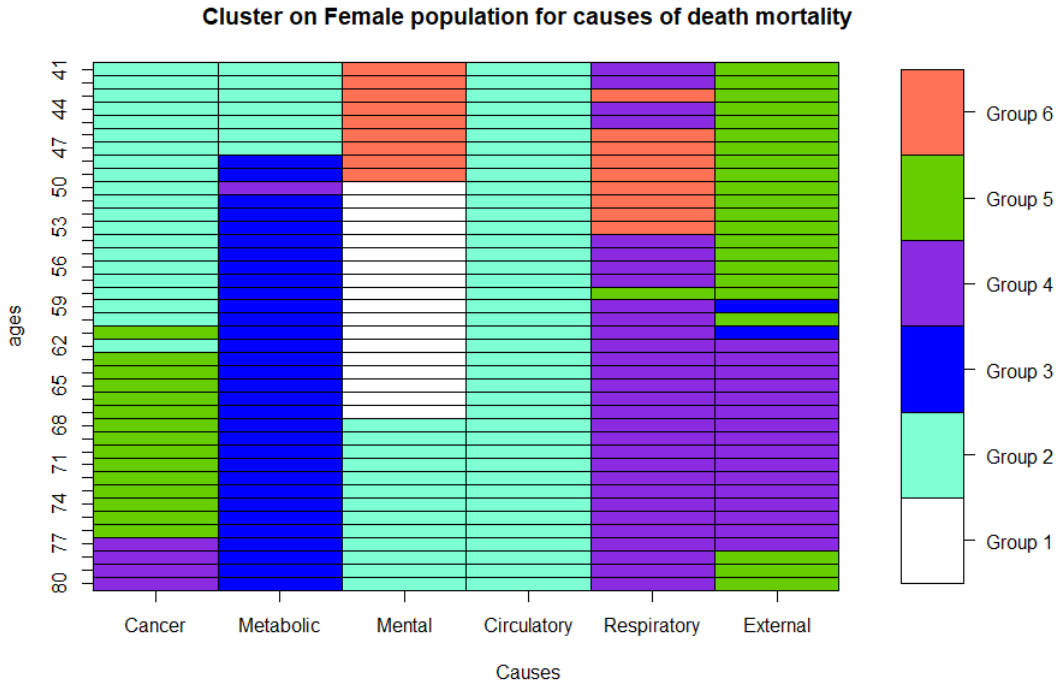


FIGURE 3.4.9 – *Constrained clustering for Male population by causes of death*

Cluster	MSE	MAE	MAPE
Reference cluster	0.0169789	0.1029326	0.01419595
K-LC cluster	0.01310745	0.08807986	0.01228526
K-LC with constraints	0.01287824	**0.08643274**	**0.01208211**
K-LC penalized	**0.01248059**	0.08665662	0.01217052

TABLE 3.4.5 – *Out-sample errors*

with a slight advantage for the penalization according to the MSE criterion and a slight superiority of the constrained method for the MAE and MAPE criteria.

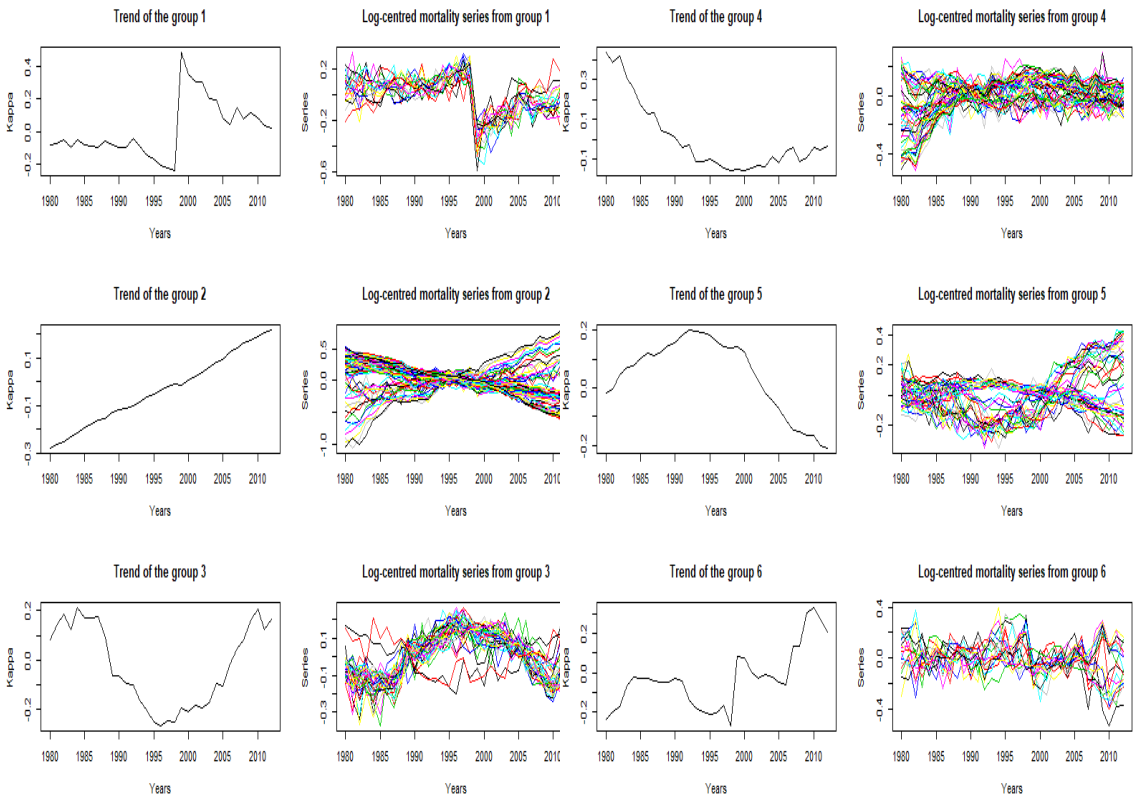


FIGURE 3.4.10 – *Trends for each group and associated centered series*

3.5 Conclusion

The clustering of mortality time series has received little attention in the literature. When the choice of groups has an explanatory purpose, it is indeed pointless to address the question of optimal clustering. However, in the absence of qualitative information in grouping data, it may be beneficial to process a complete statistical analysis rather than defining the groups arbitrarily. A rigorous approach of clustering can lead to better fitting, new explanations of mortality dynamics and possibly better forecasts. In this paper, we proposed a clustering approach based on the Lee-Carter model. We showed that the algorithm of the K-LC is the K-centroids method for a specific distance. In order to compensate for the volatility of certain series, two developments of the K-LC method have been proposed : a method with a strong grouping constraint and a penalty method with prior on groups. Two applications have been proposed based on the US population morality. In these two applications the fitting and out-sample errors have been reduced using the K-LC method rather than arbitrary clusters. If the results obtained are encouraging, several items remain to be explored. The question of the optimal number of groups

has not been addressed, nor the question of the choice of the penalty parameter. A rigorous process to select these two parameters (number of K groups and λ parameter) would likely shed light on mortality trends and perhaps improve the quality of the forecasts. Finally, application to richer databases would certainly be instructive in identifying mortality patterns. Such an application could include, in addition to mortality data by cause, sex and age, information on geographical location or standard of living.

Chapitre 4

A multivariate Poisson state-space model to forecast cause of death mortality

Abstract

This chapter deals with the cause of death mortality forecasting by applying a Poisson state-space model on mortality data. The choice of the hidden variables dynamic is such that both cause of death dependency and trend shift are taken into account. The parameter estimation using a likelihood maximization is achieved through the Expectation Maximization algorithm which articulates particle smoothing, numerical methods and closed formulae. The algorithm is applied on US female mortality data. We examine the parameters obtained and their implications in terms of cause of death dynamics. Finally, a comparison is made with the standard Lee-Carter approach, in which we observed that, the model developed in the paper outperforms the Lee-Carter model thanks to its greater flexibility.

4.1 Introduction

Mortality study is a key topic at the crossroads of several disciplines : demography, actuarial science, gerontology and public health. One of the most challenging aspects in mortality modelling is the projection of mortality series, which is a necessary step to capture properly the longevity risk, i.e. the risk that people live longer than expected and outlive their retirement savings. A better understanding of mortality dynamic may be achieved by analyzing the causes of death which compose the all-cause mortality. Such an approach have been first experienced by Pollard (1990)[111], Crimmins (1981)[31] and Girosi et al. (2007)[62]. In practice, the cause-specific time series are forecasted and then summed up to retrieve the projected all-cause mortality. However, Wilmoth (1995)[135] demonstrates that for the forecasting models used in practice, the predictions resulting from the cause of death aggregation were always more pessimistic in the long term than the all-cause forecast. However, the cause of death projections remain attractive as they allow explaining several pieces of the mortality puzzle, such as the mortality trend changes and the basis risk between different populations. Furthermore, several works have emerged over the last decade to overcome the divergence issue between causes of death and all-cause dynamics such as the reconciliation approaches developed by Hyndman et al. (2011)[73], which has been applied successfully on causes of death by Li et al. (2019)[87]. In the insurance industry, deaths by cause are projected using the same models as for projecting all-cause mortality, despite the fact that cause-specific mortality exhibits some particularities. First, the series may display dependency structure between causes that we need to properly apprehend in order to estimate the longevity risk. One of the most important topic linked to this aspect is the question of impact of a cause of death elimination on the others causes. As the different causes of death compete with each other, the elimination of one specific cause of death may leave the floor to other causes, thereby reducing the expected mortality improvement. This case has been investigated by Honore and Lleras-Muney (2006)[72] to understand the dependency structure between cancers and other causes and explain why the cancer may have experienced a stagnation over the last decades in spite of the huge investments made in US to tackle the disease. In order to understand the dependency between causes of death, two approaches have been suggested. The first one consists in catching the dependency structure of the mortality rates across the time using parametric models such as the multinomial model proposed by Alai et al. (2015)[3], co-integrated series like Arnold and Sherris (2013)[10] or adaptation of multi-population mortality models like Lyu et al. (2020)[93]. Although these approaches are good at capturing the

time-dependency structure of the mortality series, they are not tailored to capture the within-cohort dependency. For this aspect, specific models have been proposed based on the copula theory (see Carriere (1994)[25], Zheng and Klein (1995)[139] and Dimitrova et al. (2013)[44] for cause of death application of the copulas).

Li and Lu (2019)[88] propose an approach to fill the gap between these two approaches using hierarchical Archimedean copulas to transform the forecasted crude cause-specific mortality intensity into the forecasted net mortality intensity. This technique, while being promising, requires in a first instance to have reliable projections for the crude mortality by cause. However, it appears that the current forecasting methods, derived from the standard Lee-Carter model (see Lee and Carter (1992)[114]) do not provide sufficiently accurate results on forecasts because of the unsuitability of the usual dynamic. In a first hand the dynamic have to include the cause of death dependency to properly catch the longevity risk. As discussed early, important works have emerged in this field, but these models are either focused on long-run relationships between causes of death or on other indicators than the crude mortality. In another hand, the cause-specific mortality experience trend shift more frequently than the all-cause mortality, which we need to integrate in the dynamic to adjust the trend on the most recent observations. In consequences, we propose a multivariate dynamic for the Lee-Carter model including both the dependency structure between causes of death and potential trend change. In the Lee-Carter (LC) model, the structural mortality parameter are estimated in a first step, then the dynamic is estimated on the time-dependent parameters. As explained in Fung et al. (2017)[58], the usual two-step estimation of the LC model leads to a statistical lack of efficiency in comparison with models of which the structural parameters and the dynamic are conjointly estimated. Another issue with the LC model is that it forecasts the logarithm of the crude mortality intensity, and not the crude mortality intensity itself, which introduces a bias in the estimation of the last one. This aspect is particularly important for some methods of reconciliation such as the one used in Athanasopoulos et al. (2009)[12], Hyndman et al. (2011)[73] or Wickramasuriya et al. (2018)[134]). For these reasons, we propose a Poisson state-space model to estimate the cause-specific dynamic. The Poisson framework for LC model was introduced by Brouhns et al. (2002)[20] to overcome the LC unrealistic feature of homogeneity. In association with the state-space framework, it also ensures to obtain non-biased estimation of the crude mortality intensity as illustrated in Andersson and Lindholm (2020)[8], in which a more general Markov structure is proposed to forecast the age-period mortality. State-space models refer to a class of probabilistic models describing

relationship between the observed and latent variables (see Durbin and Koopman (2001)[47]). It allows estimating simultaneously the dynamic and the structural parameters of the model. LC state-space models have been proposed by Pedroza (2006)[108] and Kogure and Kurachi (2010)[83] for Bayesian estimation and by De Jong and Tickle (2006)[36] for a maximum likelihood estimation. Fung et al. (2017)[58] unify the framework of state-space mortality for the most common mortality stochastic models both with Bayesian and frequentist methods. Liu and Siu (2016a)[91] modify the state-space Cairns-Blake-Dowd (CBD)[21] model to include multi-population all-causes mortality. Liu and Siu (2016b)[92] propose a two-layers latent variable structure for the state-space CBD model to take into account the potential trend change in mortality dynamic. We follow this approach in great adaptability, especially in comparison with the CBD model which is more adapted to capture all-cause mortality with its linear age sensitivity regarding the trend. In our model, the dependency structure is integrated by considering multivariate Gaussian dynamic for the two layers of the state variable. We describe the estimation process for Likelihood maximization, adapting the approach developed by Schön et al. (2011)[123] to the multivariate state-space Poisson LC. The model is applied on the US female population for 7 causes of death. We compare the precision of the approach with the LC model applied on each cause individually.
This paper is organized as follows : Section 4.2 presents the state-space model ; the dependency relationships between observed and the latent variables are specified to identify our model as a state-space model. Section 4.3 deals with filtering and smoothing methods specific to the type of state-space model used. The Sequential Importance Sampling (SIS) algorithm, also known as Particle Filtering, which is described in this section, is useful regarding two aspects : it is the first step in likelihood smoothing, and it allows us to obtain unbiased forecasts. The Smoothing algorithm, based on the filtering results, is a necessary preliminary step for the application of the Expectation Maximization (EM) algorithm. Section 4.4 deals with application of the EM algorithm to the cause-specific mortality model. We use the algorithm developed by Schön et al. (2011)[123] for the parameter calibration. Section 4.5 presents an application of the approach using US Female mortality data. We compare the Lee-Carter model applied to each cause of death separately to the multivariate Poisson state-space approach. Finally, Section 4.6 concludes this paper.

4.2 Model

4.2.1 Notations and variables

Data are available by age $x = x_1, ..., x_{n_a}$, calendar year $t = t_1, ..., t_{n_y}$ and causes of death $i = 1, ..., I$, where n_a, n_y and I refer respectively to the number of ages observed, the number of years considered and the causes of death studied. For simplicity, the ages and calendar years observed are equally spaced such that $x_{k+1} - x_k = 1$ and $t_{k+1} - t_k = 1$, $\forall k$. We observe the number of deaths $D^i_{x,t}$ occurring for age x, at year t and due to cause of death i and exposures to death $E_{x,t}$. This exposure represents the time spent by the observed population at age x during year t. We assume that lifespan of an individual is represented by a random variable X. In presence of causes of death, the competing risk framework models the lifespan as $X = min(X^1, ..., X^I)$, where X^i represents the age at death from cause i. There are two distinct forces of mortality to analyze : the crude cause-specific force defined as

$$\mu^i_x = \lim_{s \to 0_+} \frac{\mathbb{P}\left(x < X^i \leq x + s \mid x < X\right)}{s}, \tag{2.1}$$

and the net force of mortality, defined as

$$\mu^{i,net}_x = \lim_{s \to 0_+} \frac{\mathbb{P}\left(x < X^i \leq x + s \mid x < X^i\right)}{s}. \tag{2.2}$$

Equation (2.1) implies that the all-cause force of mortality writes :

$$\mu_x = \lim_{s \to 0_+} \frac{\mathbb{P}(x < X \leq x + s \mid x < X)}{s} = \sum_{i=1}^{I} \mu^i_x, \tag{2.3}$$

while in the absence of extra identifying assumptions, it has been demonstrated (see Tsiatis (1975)[126]) that the net force of mortality is not identifiable and may not be linked to the all-cause force of mortality. Under the hypothesis of independence between the lifespan variables X_i, these two coefficients are equal, which may be wrong under other assumptions. As mentioned in the introduction, several methods have emerged to overcome this issue, notably the Li and Lu (2019)[88] which allows obtaining, given some hypotheses, the net mortality force from the crude. By the following, we consider the crude mortality forces and suppose that the following assumptions are satisfied :

- the forces of mortality depend on the calendar year during which the deaths occur, and are denoted by $\mu_{x,t}$,
- the mortality coefficient is piecewise constant regarding year and age, i.e. for

any integers x and t, for $u \in [0,1[$ and $v \in [0,1[$, we have :

$$\mu^i_{x+u,t+v} = \mu^i_{x,t}. \tag{2.4}$$

Under these hypotheses the likelihood writes (see Andersen et al. (1993)[7] pp. 402–406) :

$$L\left(\left(D^i_{x,t}\right)_{x,t,i}, \left(E_{x,t}\right)_{x,t}; \left(\mu^i_{x,t}\right)_{x,t,i}\right) = \prod_{t=t_1}^{t_{n_y}} \prod_{x=x_1}^{x_{n_a}} \prod_{i=1}^{I} \exp\left(-E_{x,t}\mu^i_{x,t}\right) \left(\mu^i_{x,t}\right)^{D^i_{x,t}}. \tag{2.5}$$

Regarding $\mu^i_{x,t}$ we recognize that Equation 2.5 is proportional to a product of independent Poisson distributions of parameter $E_{x,t}\mu^i_{x,t}$. Then we consider that

$$D^i_{x,t} \sim P\left(E_{x,t}\mu^i_{x,t}\right). \tag{2.6}$$

4.2.2 Lee-Carter Model

The Lee-Carter model (1992)[114] is a popular model allowing the forecast of age-period mortality measures in a very parsimonious and efficient way. It has been adapted to the Poisson mortality framework by Brouhns et al. (2002)[20] to consider heterogeneous volatility by age and period. The model is designed as follows for the all-cause mortality :

$$\left(D_{x,t}\right)_{x,t} \sim P\left(\lambda_{x,t}\right), \tag{2.7}$$

with $\lambda_{x,t} = E_{x,t} \exp\left(\alpha_x + \beta_x \kappa_t\right)$. The model parameters are interpreted in the same way as for the original Lee-Carter model :

- α_x is the time-independent level parameter, whose interpretation depends on the constraint imposed on the other parameters.
- κ_t is the time-varying parameter that determines the forces of mortality trend whose the dynamic follows an ARIMA model.
- β_x is the sensitivity parameter of mortality at age x with respect to the dynamics of the time parameter κ_t.

We notice that the model is not identifiable without additional constraints. The constraints usually applied are $\sum_t \kappa_t = 0$ and $\sum_x \beta_x = 1$. However, in the state-space framework, we will use constraints facilitating the estimation : $\kappa_{t_1} = 0$ and $\beta_{x_1} = 1$. In Brouhns et al. (2002)[20] approach, the estimation is decomposed in two steps : the first one consists in estimating the parameters $\left(\alpha_x, \beta_x, \kappa_t\right)_{x,t}$ using the Newton-Raphson algorithm, while in the second is devoted to the estimation

of the κ_t dynamic with an ARIMA model. As explained in Fung et al. (2017)[58] this choice may lead to a statistical lack of efficiency in comparison with models of which the structural parameters and the dynamic are conjointly estimated. For this reason, we propose to estimate the Poisson LC model by cause with a state-space approach.

4.2.3 State-space Poisson Model

As a preamble to the mortality model, we present in Section 4.2.3 the state-space model theory. The reader can refer to the rich monograph of Durbin and Koopman (2001)[47] for more details. The state-space method aims to model random events called the space variables using latent variables referred as state variables. We denote by $p_\theta(.)$ the probability density of the different variables conditionally to a time-independent vector parameter θ, Y_t the space variable and X_t the state variable. Both the state and space variables can be multidimensional. Then a state-space model assumes the relationships :

$$\begin{aligned} Y_t &\sim p_\theta\left(Y_t \mid X_t\right), \\ X_t &\sim p_\theta\left(X_t \mid X_{t-1}\right), \\ X_{t_1} &\sim p_{\theta,t_1}, \end{aligned} \tag{2.8}$$

with the following assumptions :

- independence between the space variable given the state variable taken at period t, and previous state variables, $\forall t_n > t_1, p_\theta\left(Y_{t_1:t_n} \mid X_{t_1:t_n}\right) = \prod_{t=t_1}^{t_n} p_\theta\left(Y_t \mid X_t\right)$;
- the state variable is a Markov chain, i.e. $p_\theta\left(X_{t_1:t_n}\right) = p_\theta\left(X_{t_1}\right) \prod_{t=t_1}^{t_n-1} p_\theta\left(X_{t+1} \mid X_t\right)$.

To adapt the state-space model to the causes of death mortality, we assume that $D_{x,t}^i \sim P\left(\lambda_{x,t}^i\right)$, with $\lambda_{x,t}^i = E_{x,t} \exp\left(\alpha_x^i + \beta_x^i \kappa_t^i\right)$ as in the LC model. We provide a dynamic to parameter κ_t^i flexible enough to take into account the potential drift changes and richer enough to integrate cause dependency. Let $\kappa_t = \left(\kappa_t^1, \kappa_t^2, \ldots, \kappa_t^I\right)'$ be the vector combining the time dependent parameters by cause of death and $\rho_t = \left(\rho_t^1, \rho_t^2, \ldots, \rho_t^I\right)'$ drift process of the same dimension as κ_t.

$$\begin{aligned} \kappa_t &= \kappa_{t-1} + \rho_{t-1} + \eta_t, \\ \rho_t &= \rho_{t-1} + \zeta_t, \end{aligned} \tag{2.9}$$

where $\eta_t \sim N\left(0, \Sigma_\eta\right), \zeta_t \sim N\left(0, \Sigma_\zeta\right)$ and η_t, ζ_s are independent $\forall t, s$. If we assume that η_{t_1} and ζ_{t_1} are multivariate Gaussian variables and denote $\epsilon_t = (\eta_t', \zeta_t')$ then the state variable $X_t = (\kappa_t', \rho_t')'$ has the following dynamic :

$$\begin{aligned} X_t &= AX_{t-1} + \epsilon_t, \\ X_{t_1} &\sim N\left(\bar{x}_{t_1}, \Sigma_{\rho_{t_1}}\right), \end{aligned} \tag{2.10}$$

where $\bar{x}_{t_1}$ and $\Sigma_{\rho_{t_1}}$ are respectively the initial state mean and variance parameters of X_{t_1}. Let us assume for now that these parameters are given (Section 4.4 deals with the estimation of the parameters). The matrix A is defined in such a way that the dynamics described in 2.9 for κ_t and ρ_t are respected, i.e.

$$A = \begin{pmatrix} A_{sup} \\ A_{inf} \end{pmatrix} = \begin{pmatrix} I_d & I_d \\ I_d & 0 \end{pmatrix} \tag{2.11}$$

,with I_d the identity matrix of the same dimension than κ_t and ρ_t. By denoting $Y_t = \left(\left(D^i_{x_1,t}\right)'_i, \left(D^i_{x_2,t}\right)'_i, \ldots, \left(D^i_{x_{n_a},t}\right)'_i\right)'$, we have the following state-space model :

$$\begin{aligned} Y_t &\sim p_\theta\left(Y_t \mid X_t\right) = \textstyle\prod_{i,x} \mathscr{P}\left(D^i_{x,t}; E_{x,t}\exp\left(\alpha^i_x + \beta^i_x\kappa^i_t\right)\right), \\ X_t &\sim p_\theta\left(X_t \mid X_{t-1}\right) = \mathscr{N}\left(X_t; \left(AX_{t-1}, \text{blockdiag}\left(\Sigma_\eta, \Sigma_\zeta\right)\right)\right), \\ X_{t_1} &\sim p_{\theta,t_1}, \end{aligned} \tag{2.12}$$

with $\mathscr{P}(n;\lambda)$ the probability mass distribution of a Poisson realization n and a parameter λ, $\mathscr{N}(x;(\mu,\Sigma))$ the density of a Gaussian distribution with mean μ and variance Σ taken at point x and blockdiag(.,.) the operator defined such that for two matrices A and B, we have :

$$\text{blockdiag}(A,B) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}. \tag{2.13}$$

In the state-space model (2.12), the observed variable is the vector of deaths number $D^i_{x,t}$, the state variable is (κ_t, ρ_t), the inputs are $E_{x,t}$, and the model parameters are grouped in the vector :

$$\theta = \left(\left(\alpha^i_x\right)'_{x,i}, \left(\beta^i_x\right)'_{x,i}, \text{vec}\left(\Sigma_\eta\right)', \text{vec}\left(\Sigma_\zeta\right)', \text{vec}\left(\Sigma_{\rho_{t_1}}\right)', \bar{x}'_{t_1}\right)' \tag{2.14}$$

Two observations can be made from model (2.12). The first one refers to the correlation of the dynamics of causes of death induced by the covariance matrix of the random vectors η_t and ζ_t. In this model, shocks on some causes can have an impact on other cause mortalities. This dependency is particularly important when considering cause of death elimination scenarios. The second observation is that, contrary to the usual modelling, we have chosen to introduce randomness in the step parameter ρ_t. This randomness reflects potential trend change by cause of death. A dependency structure is also proposed for the dynamics of the drifts

through the matrix Σ_ζ, allowing to integrate complex random structure in cause mortality trend shift.

To conclude this section, we propose a brief discussion on the constraint choice of the model. As said before, without constraints on the structural parameters, the LC model is not identifiable. Hence, an infinity of optimal solutions exist, which may be detrimental for the optimization process. In this paper, we assume that that for every i, $\kappa^i_{t_1} = 0$ and $\beta^i_{x_1} = 1$. These constraints, in addition to make the model identifiable, facilitate the simulation process detailed thereafter, as the first element of the time series $(\kappa_t)_t$ is fixed. If the choice of the constraint is not crucial for the estimation of the parameters, it may have an important impact on the forecasting. In this paper, we do not dwell on this question, but the reader interested may refer to Nielsen and Nielsen (2010)[104].

4.3 Sequential Monte Carlo filtering and smoothing

In order to forecast mortality, we first have to estimate the latent state variables $(X_t)_{t=t_1,\dots,t_{n_y}}$ as well as the parameter vector θ. The estimation of θ will be delayed until Section 4.4 while this section covers the question of estimating $(X_t)_{t=t_1,\dots,t_{n_y}}$. The estimation of the latent parameter X_t is achieved through the computation of the conditional expectancy $\bar{X}_t = \mathbb{E}\left[X_t \mid Y_{t_1}, \dots, Y_{t_{n_y}}\right]$. In the general frequentist state-space framework, this quantity is obtained through the successive applications of two algorithms : the filtering and the smoothing steps.

4.3.1 Filtering

In this section we explain the filtering method for non-linear, non-Gaussian state-space. The principle of filtration is to obtain the conditional expectation of a function of the state variables X_t at time t given the observations available up to that time, allowing for instance the estimation of the state variable X_t conditionally to the observation $Y_{t_1:t} = (Y_{t_1}, \dots, Y_t)$. When the state and space variables are both Gaussian and the relationships between Y_t and X_t and between X_t and X_{t-1} are linear, the filtering is achieved by using the Kalman Filter developed in Kalman (1960)[76]. In the non-Gaussian, non-linear framework, the absence of a closed formula for the likelihood prevents using such an algorithm, which makes it necessary to employ simulation methods. Let $h(.)$ be a function of $X_{t_1:t}$ and assume for now that the parameter vector θ is given. The function $\bar{h}_t = \mathbb{E}\left[h\left(X_{t_1:t}\right) \mid Y_{t_1:t}\right]$ is estimated as follows :

$$\bar{h}_t = \int h\left(x_{t_1:t}\right) p_\theta\left(x_{t_1:t} \mid Y_{t_1:t}\right) dx_{t_1:t}. \tag{3.1}$$

Assuming that the function $q_\theta(.|Y_{t_1:t})$ has the same support than $p_\theta(.|Y_{t_1:t})$, the following relationship holds :

$$\overline{h}_t = \int h(x_{t_1:t}) \frac{p_\theta(x_{t_1:t}|Y_{t_1:t})}{q_\theta(x_{t_1:t}|Y_{t_1:t})} q_\theta(x_{t_1:t}|Y_{t_1:t}) dx_{t_1:t} = E_q\left[h(X_{t_1:t}) \frac{p_\theta(X_{t_1:t}|Y_{t_1:t})}{q_\theta(X_{t_1:t}|Y_{t_1:t})}\right]. \tag{3.2}$$

In addition, by writing $p_\theta(x_{t_1:t}|Y_{t_1:t}) = \frac{p_\theta(x_{t_1:t}, Y_{t_1:t})}{p_\theta(Y_{t_1:t})}$ and noting that

$$p_\theta(Y_{t_1:t}) = \int \frac{p_\theta(x_{t_1:t}, Y_{t_1:t})}{q_\theta(x_{t_1:t}|Y_{t_1:t})} q_\theta(x_{t_1:t}|Y_{1:t}) dx_{t_1:t}, \tag{3.3}$$

we obtain :

$$\overline{h}_t = \frac{E_q\left[h(X_{t_1:t}) \frac{p_\theta(X_{t_1:t}, Y_{t_1:t})}{q_\theta(X_{t_1:t}|Y_{t_1:t})}\right]}{E_q\left[\frac{p_\theta(X_{t_1:t}, Y_{t_1:t})}{q_\theta(X_{t_1:t}|Y_{t_1:t})}\right]}. \tag{3.4}$$

We denote $\widetilde{w}_t = \frac{p_\theta(x_{t_1:t}, Y_{t_1:t})}{q_\theta(x_{t_1:t}|Y_{t_1:t})}$. To lighten the notation, we omit writing the dependency between $\widetilde{w}_t$ and the variables $x_{t_1:t}$ and $Y_{t_1:t}$. We draw M variables $x_{t_1:t}^{(j)}$ according to $q_\theta(x_{t_1:t}|Y_{t_1:t})$ and then we propose the following estimator :

$$\overline{h}_t \approx \frac{\frac{1}{M}\sum_{j=1}^M h(x_{t_1:t}^{(j)}) \widetilde{w}_t^{(j)}}{\frac{1}{M}\sum_{j=1}^M \widetilde{w}_t^{(j)}} = \sum_{j=1}^M h(x_{t_1:t}^{(j)}) w_t^{(j)}, \tag{3.5}$$

where $w_t^{(j)} = \frac{\widetilde{w}_t^{(j)}}{\sum_k \widetilde{w}_t^{(k)}}$.

Estimator 3.5 is called Importance Sampling (IS) estimator. In state-space models, some assumptions about the $q_\theta(x_{t_1:t}|Y_{t_1:t})$ function save valuable time in simulations. Rather than simulating a new sample $x_{t_1:t}$ every t period, it would be advantageous if the simulation $X_{t_1:t}^{(j)}$ could be done starting from $x_{t_1:t-1}^{(j)}$, i.e. by drawing $x_t^{(j)}$ from $x_{t_1:t-1}^{(j)}$. We then obtain :

$$\begin{aligned} q_\theta(x_{t_1:t}|Y_{t_1:t}) &= \frac{q_\theta(x_{t_1:t}, Y_{t_1:t})}{q_\theta(Y_{t_1:t})} = \frac{q_\theta(x_t|x_{t_1:t-1}, Y_{t_1:t}) q_\theta(x_{t_1:t-1}, Y_{t_1:t})}{q_\theta(Y_{t_1:t})} \\ &= q_\theta(x_t|x_{t_1:t-1}, Y_{t_1:t}) q_\theta(x_{t_1:t-1}|Y_{t_1:t}). \end{aligned} \tag{3.6}$$

By assuming that $q_\theta(x_{t_1:t-1}|Y_{t_1:t}) = q_\theta(x_{t_1:t-1}|Y_{t_1:t-1})$, we switch from the Importance Sampling algorithm to the Sequential Importance Sampling algorithm

(SIS) known also as the Particle Filtering algorithm. As a result, we have :

$$\begin{aligned}\widetilde{w}_t &= \frac{p_\theta(x_{t_1:t}, Y_{t_1:t})}{q_\theta(x_{t_1:t}|Y_{t_1:t})} = \frac{p_\theta(x_{t_1:t-1}, Y_{t_1:t-1})p_\theta(x_t, Y_t|x_{t_1:t-1}, Y_{t_1:t-1})}{q_\theta(x_{t_1:t-1}|Y_{t_1:t-1})q_\theta(x_t|x_{t_1:t-1}, Y_{t_1:t})} \\ &= \widetilde{w}_{t-1}\frac{p_\theta(x_t|x_{t-1})p_\theta(Y_t|x_t)}{q_\theta(x_t|x_{t_1:t-1}, Y_{t_1:t})}.\end{aligned} \tag{3.7}$$

In consequence, the SIS algorithm can be written as follows :

(I) for $t = t_1$

(1) draw of $x_{t_1}^j$ with probability $q_\theta(x_{t_1}|Y_{t_1})$ for $j = 1, ..., M$

(2) compute $\widetilde{w}_{t_1}^j = \frac{p_\theta(Y_{t_1}|x_{t_1}^j)p_{\theta,t_1}(x_{t_1}^j)}{q_\theta(x_{t_1}|Y_{t_1})}$, and normalize $w_{t_1}^j = \frac{\widetilde{w}_{t_1}^j}{\sum_k \widetilde{w}_{t_1}^k}$ for $j = 1, ..., M$.

(II) for $t = t_2, ..., t_{n_y}$

(1) draw of x_t^j with probability $q_\theta(x_t|x_{t_1:t-1}, Y_{t_1:t-1})$ for $j = 1, ..., M$

(2) compute $\widetilde{w}_t^j = w_{t-1}^j \frac{p_\theta(Y_t|x_t^j)p_\theta(x_t^j|x_{t-1}^j)}{q_\theta(x_t^j|x_{t_1:t-1}^j, Y_{t_1:t})}$, and normalize $w_t^j = \frac{\widetilde{w}_t^j}{\sum_k \widetilde{w}_t^k}$.

This is the basic algorithm from which several variations have emerged in the literature based on the choice of the q function (see chapters 11 and 12 in Durbin and Koopman (2001) [47]). By the following, we opt for the Bootstrap particle filter approach developed by Gordon et al. (1993)[66] for which $q_\theta\left(x_t \mid x_{t_1:t-1}, Y_{t_1:t-1}\right) = p_\theta\left(x_t \mid x_{t-1}\right)$. In this approach, the draws are re-sampled after each iteration to keep the most likely trajectories. It also results that the weights associated to these new sample are uniform. Then the algorithm becomes :

(I) Initialize $\tilde{x}_{t_1}^j$ with probability $p_\theta\left(x_{t_1}\right)$ for $j = 1, \ldots, M$.

(II) For each t, draw $\tilde{x}_t^j$ with probability $p_\theta\left(x_t \mid x_{t-1}^j\right)$ for $j = 1, \ldots, M$.

(III) Compute $w_t^j = \frac{p_\theta\left(Y_t|\tilde{x}_t^j\right)}{\sum_k p_\theta\left(Y_t|\tilde{x}_t^k\right)}$.

(IV) Draw a new sample $\left(x_t^j\right)_{j=1,\ldots,M}$ by resampling the variables $\left(\tilde{x}_t^j\right)_{j=1,\ldots,M}$ according to probabilities $\left(w_t^j\right)_j$, and set $t \leftarrow t + 1$.

With this approach, the filtered expectancy is provided by $\mathrm{E}[h\left(X_t\right) \mid Y_{t_1:t}] = \sum_{i=1}^M w_t^i h\left(\tilde{x}_t^i\right)$

4.3.2 Smoothing

Once the simulation using filtering algorithm is finalized, the simulations of the state variables from the smoothed probability $p_\theta\left(x_t|Y_{t_1:t_{n_y}}\right)$ can be drawn using

the weights and variables derived in the previous step :

$$\begin{aligned}
p_\theta\left(x_t|x_{t+1}, Y_{t_1:t_{n_y}}\right) &= p_\theta\left(x_t|x_{t+1}, Y_{t+1:t_{n_y}}, Y_{t_1:t}\right) \\
&= \frac{p_\theta\left(x_t, x_{t+1}, Y_{t+1:t_{n_y}}, Y_{t_1:t}\right)}{p_\theta\left(x_{t+1}, Y_{t+1:t_{n_y}}, Y_{t_1:t}\right)} \\
&= \frac{p_\theta\left(Y_{t+1:t_{n_y}}|x_t, x_{t+1}, Y_{t_1:t}\right) p_\theta\left(x_t, x_{t+1}, Y_{t_1:t}\right)}{p_\theta\left(x_{t+1}, Y_{t+1:t_{n_y}}, Y_{t_1:t}\right)} \\
&= \frac{p_\theta\left(Y_{t+1:t_{n_y}}|x_t, x_{t+1}, Y_{t_1:t}\right) p_\theta\left(x_t|x_{t+1}, Y_{t_1:t}\right) p_\theta\left(x_{t+1}, Y_{t_1:t}\right)}{p_\theta\left(x_{t+1}, Y_{t+1:t_{n_y}}, Y_{t_1:t}\right)} \\
&= \frac{p_\theta\left(Y_{t+1:t_{n_y}}|x_t, x_{t+1}, Y_{t_1:t}\right) p_\theta\left(x_t|x_{t+1}, Y_{t_1:t}\right)}{p_\theta\left(Y_{t+1:t_{n_y}}|x_{t+1}, Y_{t_1:t}\right)} \\
&= p_\theta\left(Y_{t+1:t_{n_y}}|x_{t+1}, Y_{t_1:t}\right)
\end{aligned} \tag{3.8}$$

with the last equality obtained using the Markov property of the model (2.12). Using the intermediary result (3.8), it follows that :

$$\begin{aligned}
p_\theta\left(x_t|Y_{t_1:t_{n_y}}\right) &= \int p_\theta\left(x_t|x_{t+1}, Y_{t_1:t}\right) p_\theta\left(x_{t+1}|Y_{t_1:t_{n_y}}\right) \mathrm{d}x_{t+1} \\
&= \int \frac{p_\theta\left(x_{t+1}|x_t\right) p_\theta\left(x_t|Y_{t_1:t}\right)}{p_\theta\left(x_{t+1}|Y_{t_1:t}\right)} p_\theta\left(x_{t+1}|Y_{t_1:t_{n_y}}\right) \mathrm{d}x_{t+1} \\
&= p_\theta\left(x_t|Y_{t_1:t}\right) \int \frac{p_\theta\left(x_{t+1}|x_t\right) p_\theta\left(x_{t+1}|Y_{t_1:t_{n_y}}\right)}{p_\theta\left(x_{t+1}|Y_{t_1:t}\right)} \mathrm{d}x_{t+1},
\end{aligned} \tag{3.9}$$

We can develop the denominator in the integrand :

$$p_\theta\left(x_{t+1}|Y_t\right) = \int p_\theta\left(x_{t+1}|x_t\right) p_\theta\left(x_t|Y_t\right) \mathrm{d}x_t \approx \sum_{i=1}^{M} w_t^i p_\theta\left(\tilde{x}_{t+1}^k|\tilde{x}_t^i\right), \tag{3.10}$$

based on the results presented in 4.3.1. If we assume there exist weights $w_{t|t_{n_y}}^i$ such that

$$\mathbb{E}\left[h(X_t)|Y_{t_1:t_{n_y}}\right] = \sum_{i=1}^{M} w_{t|t_{n_y}}^i h(x_t^i), \tag{3.11}$$

for x_t^i based on the Filtering algorithm, then we have :

$$\int \frac{p_\theta\left(x_{t+1}|x_t\right) p_\theta\left(x_{t+1}|Y_{t_1:t_{n_y}}\right)}{p_\theta\left(x_{t+1}|Y_t\right)} \mathrm{d}x_{t+1} \approx \sum_{k=1}^{M} \frac{w_{t+1|t_{n_y}}^k p_\theta\left(\tilde{x}_{t+1}^k|x_t\right)}{\sum_{i=1}^{M} w_t^i p_\theta\left(\tilde{x}_{t+1}^k|\tilde{x}_t^i\right)}. \tag{3.12}$$

By another application of the Filtering algorithm on equation 3.9, we obtain :

$$\mathbb{E}\left[h(X_t)|Y_{t_1:t_{n_y}}\right] = \sum_{j=1}^{M} h(x_t^j) w_t^i \sum_{k=1}^{M} \frac{w_{t+1|t_{n_y}}^k p_\theta\left(\tilde{x}_{t+1}^k|x_t^j\right)}{\sum_{i=1}^{M} w_t^i p_\theta\left(\tilde{x}_{t+1}^k|\tilde{x}_t^i\right)} = \sum_{j=1}^{M} h(x_t^j) w_{t|t_{n_y}}^i, \quad (3.13)$$

with

$$w_{t|t_{n_y}}^i = w_t^i \sum_{k=1}^{M} w_{t+1|t_{n_y}}^k \frac{p_\theta\left(\tilde{x}_{t+1}^k|\tilde{x}_t^i\right)}{v_t^k}, \quad (3.14)$$

where

$$v_t^k \triangleq \sum_{i=1}^{M} w_t^i p_\theta\left(\tilde{x}_{t+1}^k|\tilde{x}_t^i\right). \quad (3.15)$$

It can be observed that the weights $w_{t|t_{n_y}}^i$ are defined backwardly. Hence, because

$$\mathbb{E}\left[h(X_t)|Y_{t_1:t_{n_y}}\right] = \sum_{i=1}^{M} w_{t_{n_y}}^i h(x_t^i), \quad (3.16)$$

we can set $w_{t_{n_y}|t_{n_y}}^i = w_{t_{n_y}}^i$. We are then sure that the whole series $(w_{t_{n_y}|t_{n_y}}^i)_t$ exists for $t = t_1, ..., t_{n_y}$.

4.4 EM Algorithm

The Expectation Maximization algorithm is a popular method developed in Dempster et al (1977)[41], which aims to tackle maximum likelihood estimation for statistical models with unobserved latent variables. It allows, through two successive phases, to find local maximum estimates of parameters when the likelihood function is not fully observed. These two phases are the Expectation step during which an estimation of a pseudo-likelihood is provided and the Maximization step which aims to maximize the pseudo-likelihood. Although the pseudo-likelihood is different from the real likelihood, it has been demonstrated that its maximization is equivalent to the maximization of the real likelihood. For more details, we invite the reader to refer to Moon (1996)[101] or Do and Batzoglou (2008)[45].

The joint search for state and structural parameters θ requires the use of the likelihood :

$$L(\theta) = p_\theta\left(Y_{t_1}, \cdots, Y_{t_{n_y}}\right) = p_\theta\left(Y_{t_1}\right) \prod_{t=t_2}^{t_{n_y}} p_\theta\left(Y_t|Y_{t_1:t-1}\right). \quad (4.1)$$

The logarithm of the likelihood writes

$$l_\theta\left(Y_{t_1:t_{n_y}}\right) = \log p_\theta\left(Y_{t_1}\right) + \sum_{t=t_2}^{t_{n_y}} \log p_\theta\left(Y_t | Y_{t_1:t-1}\right), \tag{4.2}$$

and the maximum likelihood estimator is obtained by maximizing this expression :

$$\widehat{\theta} = \underset{\theta\in\Theta}{\arg\max}\ l_\theta\left(Y_{t_1:t_{n_y}}\right). \tag{4.3}$$

A closed form formula cannot be derived apart from the Gaussian and linear case. Several methods have been used to solve this problem both in a Bayesian and frequentist theoretical framework (see for example Kantas (2015) [78]). In this paper, we adopt a frequentist approach and propose to use the Expectation Maximization (EM) method. The principle is the following : as the likelihood defined in equation 4.2 is not exploitable, we introduce the full log likelihood function :

$$l_\theta\left(X_{t_1:t_{n_y}}, Y_{t_1:t_{n_y}}\right) = \log p_\theta\left(X_{t_1:t_{n_y}}, Y_{t_1:t_{n_y}}\right), \tag{4.4}$$

on which an iterative algorithm is applied. Supposing that we have an approximation of the optimal parameter θ_k, we compute the expectation of the complete likelihood with respect to the measure $p_{\theta_k}\left(X_{t_1:t_{n_y}}\right)$,

$$Q\left(\theta, \theta_k\right) = \mathbb{E}_{\theta_k}\left[l_\theta\left(X_{t_1:t_{n_y}}, Y_{t_1:t_{n_y}}\right) | Y_{t_1:t_{n_y}}\right]. \tag{4.5}$$

The second step named the maximization step consists in updating the parameter θ as follows :

$$\theta_{k+1} = \underset{\theta\in\Theta}{\arg\max}\ Q\left(\theta, \theta_k\right). \tag{4.6}$$

The procedure is iterated until convergence. The EM algorithm is designed in such manner that each step results in an increase in the log-likelihood $l_\theta\left(Y_{t_1:t_{n_y}}\right)$. The full likelihood writes :

$$\begin{aligned} l_\theta\left(X_{t_1:t_{n_y}}, Y_{t_1:t_{n_y}}\right) &= \log p_\theta\left(Y_{t_1:t_{n_y}} | X_{t_1:t_{n_y}}\right) + \log p_\theta\left(X_{t_1:t_{n_y}}\right) \\ &= \log p_\theta\left(X_{t_1}\right) + \sum_{t=t_1}^{t_{n_y}-1} \log p_\theta\left(X_{t+1} | X_t\right) + \sum_{t=t_1}^{t_{n_y}} \log p_\theta\left(Y_t | X_t\right). \end{aligned} \tag{4.7}$$

Applying the conditional expectation to $Y_{t_1:t_{n_y}}$ conditionally to the estimation $\theta = \theta_k$, we get the expectation as the sum of three terms :

$$\mathbb{E}\left[L_\theta\left(X_{t_1:t_{n_y}}, Y_{t_1:t_{n_y}}\right)\right] = Q_1 + Q_2 + Q_3, \tag{4.8}$$

where

$$\begin{aligned} Q_1 &= \int \log p_\theta\left(x_{t_1}\right) p_{\theta_k}\left(x_{t_1}|Y_{t_1:t_{n_y}}\right) \mathrm{d}x_{t_1}, \\ Q_2 &= \sum_{t=t_1}^{t_{n_y}-1} \iint \log p_\theta\left(x_{t+1}|x_t\right) p_{\theta_k}\left(x_{t+1}, x_t|Y_{t_1:t_{n_y}}\right) \mathrm{d}x_t \mathrm{d}x_{t+1}, \\ Q_3 &= \sum_{t=t_1}^{t_{n_y}} \int \log p_\theta\left(Y_t|x_t\right) p_{\theta_k}\left(x_t|Y_{t_1:t_{n_y}}\right) \mathrm{d}x_t. \end{aligned} \tag{4.9}$$

In the approach proposed by Schön et al. (2011)[123], each of these quantities is estimated by simulation. In the state-space model framework, the Expectation step of the algorithm is achieved using the weights $w^i_{t|t_{n_y}}$ defined in equation 3.14. These weights allow us to compute the conditional expectations relative to the death observations in our sample. These quantities are identified as the smoothed expectations. The quantities Q_1 and Q_3 are approximated using the smoothed weights :

$$Q_1 \approx \widehat{Q}_1 = \sum_{i=1}^{M} w^i_{1|t_{n_y}} \log p_\theta\left(\tilde{x}^i_1\right) \tag{4.10}$$

$$Q_3 \approx \widehat{Q}_3 = \sum_{t=t_1}^{t_{n_y}} \sum_{i=1}^{M} w^i_{t|t_{n_y}} \log p_\theta\left(Y_t|\tilde{x}^i_t\right). \tag{4.11}$$

We refer to Schön's lemma 6.1 (2011)[123] for the calculation of Q_2 :

$$Q_2 \approx \widehat{Q}_2 = \sum_{t=t_1}^{t_{n_y}-1} \sum_{i=1}^{M} \sum_{j=1}^{M} w^{ij}_{t|t_{n_y}} \log p_\theta\left(\tilde{x}^j_{t+1}|\tilde{x}^i_t\right), \tag{4.12}$$

with

$$w^{ij}_{t|t_{n_y}} = \frac{w^i_t w^j_{t+1|t_{n_y}} p_{\theta_k}\left(\tilde{x}^j_{t+1}|\tilde{x}^i_t\right)}{\sum_{l=1}^{M} w^l_t p_{\theta_k}\left(\tilde{x}^j_{t+1}|\tilde{x}^l_t\right)}. \tag{4.13}$$

The quantity to optimize is therefore

$$\widehat{Q}_M\left(\theta, \theta_k\right) = \widehat{Q}_1 + \widehat{Q}_2 + \widehat{Q}_3. \tag{4.14}$$

Each of the terms $\widehat{Q}_1$, $\widehat{Q}_2$ and $\widehat{Q}_3$ depend on specific subsets which constitute a partition of the parameter vector θ. This implies that the optimization can

be conducted separately for each objective function. While the maximization of $\widehat{Q}_1$ and $\widehat{Q}_3$ can be achieved through a closed formula, by equaling the gradient functions to zero, the maximization of $\widehat{Q}_2$ is more complex and requires numerical optimization methods such as the Newton-Raphson algorithm. In both cases the gradient function is necessary and we provide it in the following for the three objective functions displayed in 4.10, 4.11 and 4.12 in appendix A.

4.5 Application

We present an application of the state-space multivariate Poisson model on the US Female mortality. We consider the 6 following main causes of death : Cancers, Circulatory, Metabolic, Nervous and Mental (further referred as Mental), Respiratory and External, to which we add an "Other causes" category to account for the rest of deaths observed over the period. The ages studied for each cause range from 51 to 90 years of age, i.e., an adult population in a position to acquire or benefit from life annuities. The observation period covers the years 1979 to 2012. The observed periods start from the begining of ICD 9, to avoid another source of potential errors, we prefer to not include ICD 8 classification. The end of observation allows us to backtest the model on the period 2013 to 2017.

The death counts are taken from the CDC (Centers for Disease Control)[56] database, while the exposures used to calculate the forces of mortality are taken from the HMD (Human Mortality Database). The forces of mortality are available by cause, sex and age attained for each calendar year. Table 4.5.1 below reconstructs the data used in this study from the CDC database. Each death is recorded by a physician using the International Classification of Diseases (ICD) nomenclature to identify the cause of death. As the pathology recording system is very detailed, it leads to a high level of granularity which is not well suited to construct robust statistics. Since our attention is focused on the most important causes of death, we group the different causes of death into fewer classes. We provide the transcription codes to reconstruct these classes from the original database. These codes are different for ICD 9 and 10, which reflect changes in nomenclature as new medical considerations are incorporated.

Cause of Death	1979 to 1998 (ICD 9 : UCOD 282)	1999 to 2017 (ICD 10 : UCOD 352)
Cancer	04600 – 11800	69 - 146
Metabolic	11900 – 13200	156-173
Mental and Nervous	14100 – 15900	174-194
Circulatory system	16000 – 20300	197-246
Respiratory system	20400 – 22700	247-278
External	29900 – 35800	381-456
Other causes	All others	All others

Table 4.5.1 – *Cause of death classification according ICD 9 and 10.*

This application is composed as follows : Section 4.5.1 reviews the structural parameter resulting from this approach, Section 4.5.2 considers the model-induced

cause of death dependency relationship, Section 4.5.3 proposes to compare the forecasted forces of mortality for the Poisson state-space model and the standard LC.

4.5.1 Illustration of the model structural parameters

As described in Section 4.2, several sets of structural parameters are used to adjust historical mortality. We examine, cause by cause, the structural parameters obtained from the EM algorithm.

Circulatory

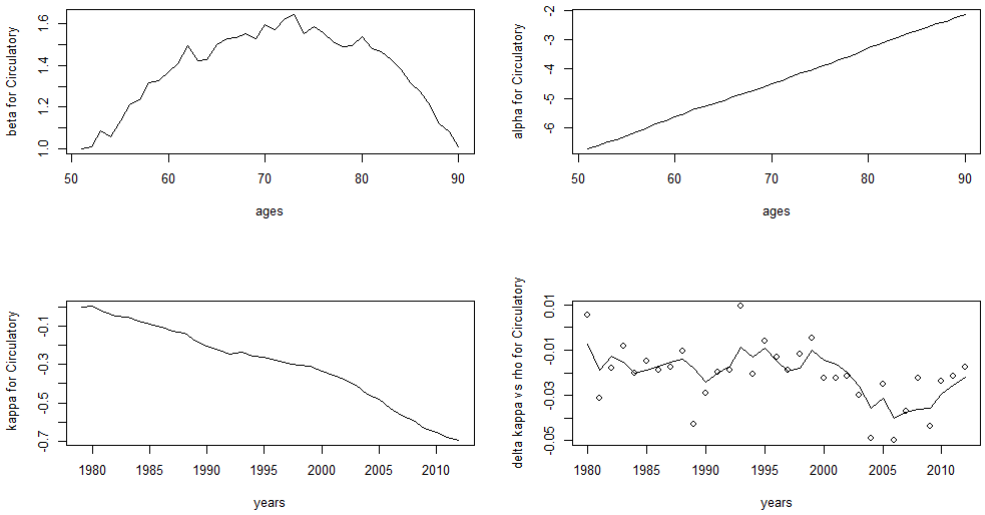


FIGURE 4.5.1 – *Structural parameters related to the Circulatory diseases.*

Figure 4.5.1 shows the structural parameters of the Circulatory diseases.

- The parameter α, in Figure 4.5.1 top right panel, is increasing linearly with respect to age, illustrating the increasing probability of dying due to Circulatory disease while aging.
- The parameter β in Figure 4.5.1 top left panel is positive, its structure is arc-shaped. We first observe an increase in the β up to age 75 and then a decrease.
- The parameter κ in Figure 4.5.1 bottom left panel is decreasing over the entire period studied. This decrease associated with positive β implies a mortality decrease over time of the Circulatory diseases. The arc shape of the parameter

β implies the decrease in mortality is lower at youngest and oldest ages and higher for ages around 75 year old.

- The parameter ρ in Figure 4.5.1 bottom right panel seems to oscillate slightly, which explains the deceleration phases of the parameter κ observed in certain years in Figure 4.5.1 bottom left panel.

Cancer

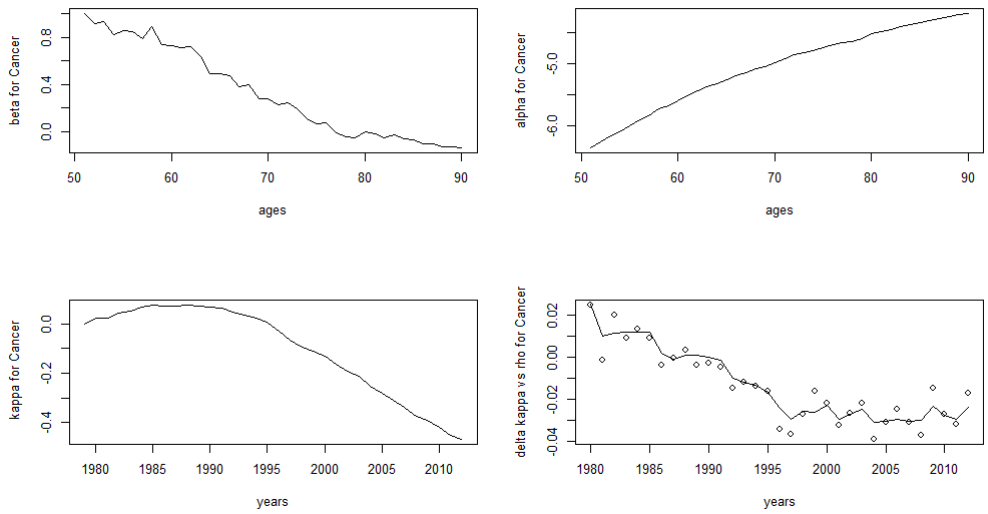


FIGURE 4.5.2 – *Structural parameters related to the cause of death Cancer.*

Figure 4.5.2 presents the structural parameters of the cause of death Cancer.

- The parameter α displayed in Figure 4.5.2 top right panel is increasing, illustrating the growing probability of dying from Cancer with age.
- The parameter β displayed in Figure 4.5.2 top left panel, is positive until about age 80, then it becomes slightly negative. It decreases monotonously with age.
- Figure 4.5.2 bottom left panel shows the time varying parameter κ which first increases until 1989, and then decreases linearly until 2012.
- Finally, Figure 4.5.2 bottom right panel illustrates the drift parameter ρ. It decreases during the 1990s before stagnating, which explains the linear trend in the κ observed over the recent years in Figure 4.5.2 bottom left panel.

Mental and Nervous

Figure 4.5.3 presents the structural parameters of the Mental diseases.

- The parameter α in Figure 4.5.3 top right panel is increasing linearly with respect to age, illustrating the increasing probability of dying from Mental diseases while aging.
- The parameter β in Figure 4.5.3 top left panel is positive until age 68 and negative after. The associated curve is initially stagnating and then decreases in a quasi-linear manner.

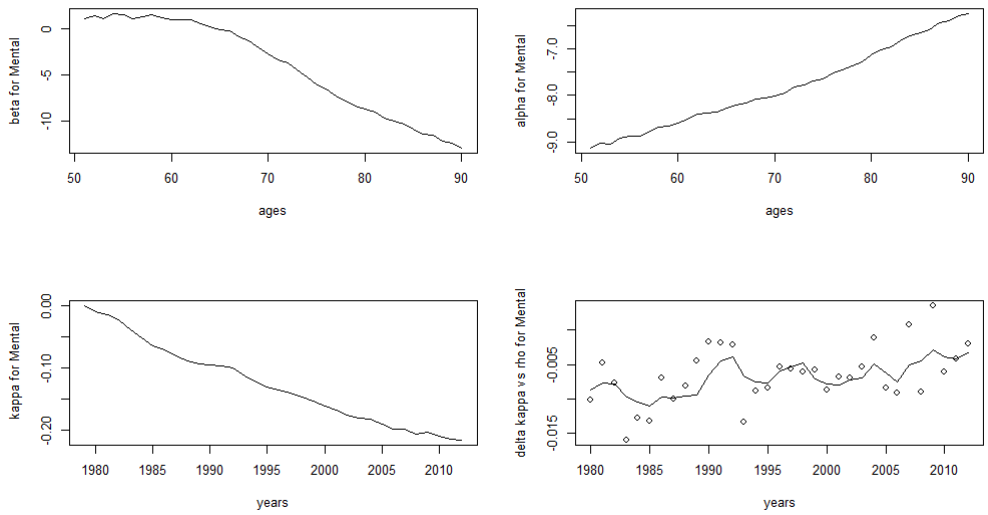


FIGURE 4.5.3 – *Structural parameters related to the Mental and Nervous diseases.*

- The parameter κ in Figure 4.5.3 bottom left panel is decreasing over the whole period, but seems to start a stagnation in the recent years.
- The parameter ρ in Figure 4.5.3 bottom right panel seems to increase through the period of observation. It is approaching zero in the recent years leading a stagnation of the κ parameter in Figure 4.5.3 bottom left.

Respiratory diseases

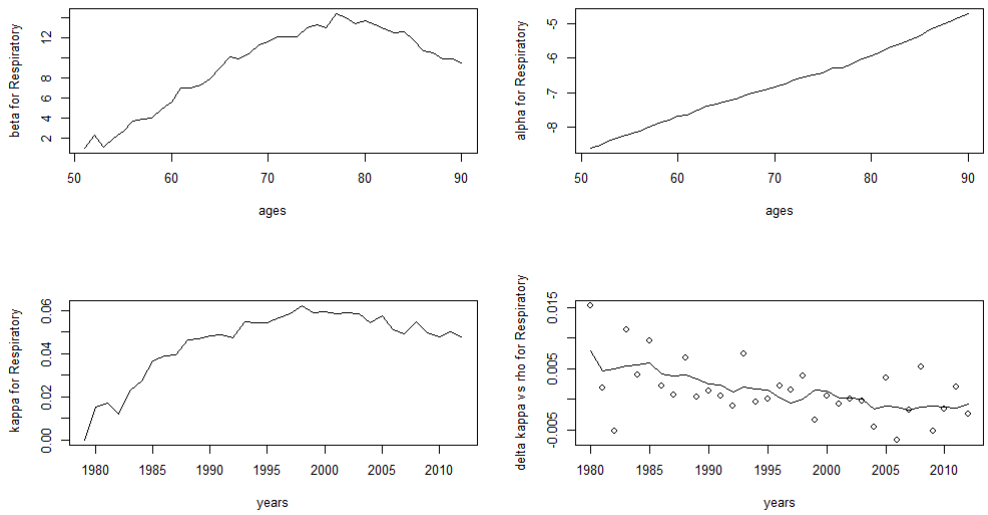


FIGURE 4.5.4 – *Structural parameters related to the respiratory diseases.*

Figure 4.5.4 presents the structural parameters of the respiratory diseases.

- The parameter α in Figure 4.5.4 top right panel is increasing linearly with age illustrating the increasing probability of dying from respiratory diseases while aging.
- The parameter β in Figure 4.5.4 top left panel is positive for the whole age range. The parameter increases until age 77 and then decreases.
- The parameter κ in Figure 4.5.4 bottom left panel is increasing over the period 1979 to 1998, and decreasing over the period 1999 to 2012. This implies for the first period 1979-1998 an increase in the number of deaths due to respiratory diseases for all ages. The opposite phenomenon occurs for the second period 1999-2012 for all the ages considered.
- The parameter ρ in Figure 4.5.4 bottom right panel decreases over the period of observation. It becomes negative after 2000 explaining the trend observed in the parameter κ in Figure 4.5.4 bottom left panel.

Metabolic

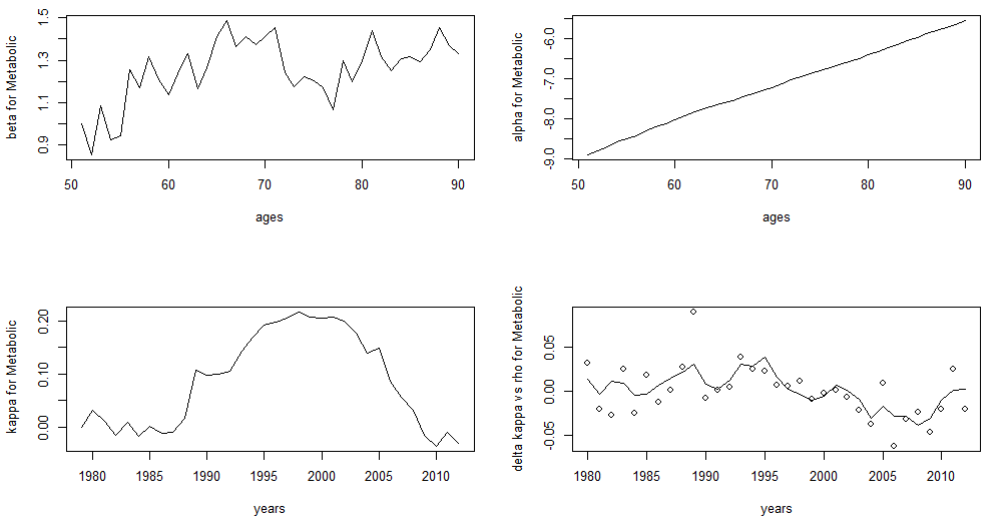


FIGURE 4.5.5 – *Structural parameters related to the metabolic diseases.*

Figure 4.5.5 illustrates the structural parameters of the metabolic diseases.

- The parameter α in Figure 4.5.5 top right panel is increasing linearly illustrating the increase of the probability of dying from metabolic diseases while aging.
- The parameter β in Figure 4.5.5 top left panel is positive. We observe first a growth of the β and then a stagnation around age 70.
- The parameter κ in Figure 4.5.5 bottom left panel is first increasing, then decreasing in the early 2000s. The β being positive, it implies that mortality

related to the metabolic diseases first increased at all ages and started to decrease after the early 2000s.

- The parameter ρ in Figure 4.5.5 bottom right panel is first positive and then decreases gradually to become negative after 2000, explaining the trend reversion of the κ parameter in Figure 4.5.5 bottom left panel. For the last years, we observe that the parameter ρ becomes closer to zero which illustrates a stagnation of the parameter κ.

External

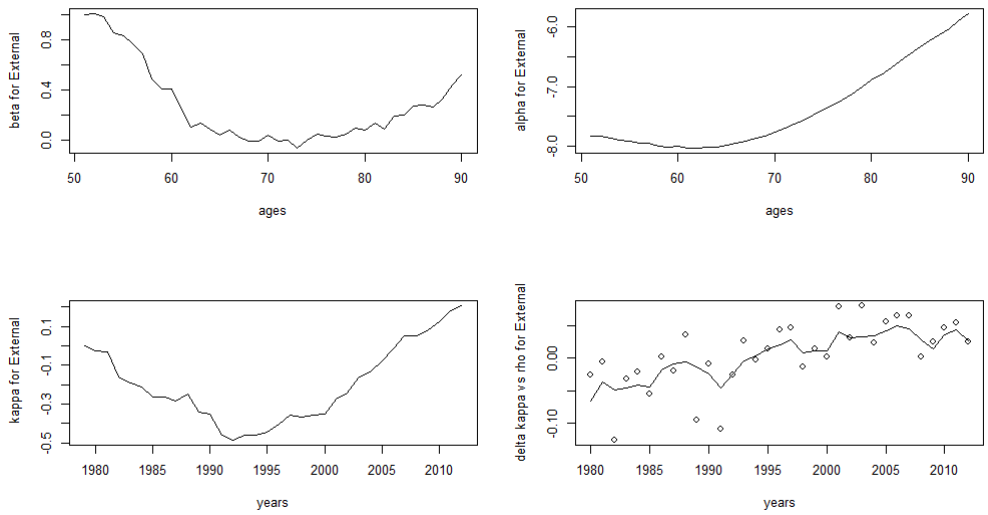


FIGURE 4.5.6 – *Structural parameters related to the external causes of death.*

Figure 4.5.6 shows the structural parameters of the external causes.

- For the youngest ages, the parameter α in Figure 4.5.6 top right panel seems to experience a decreasing trend which ends at age 63 and then increases.
- The parameter β in Figure 4.5.6 top left panel is positive. It first decreases with age until 75 and then increases.
- The parameter κ in Figure 4.5.6 bottom left panel is first decreasing until 1992 and then increases linearly starting 2000. This implies that mortality from external causes initially decreased over time before increasing. However, the β parameters in Figure 4.5.6 top left panel are colse to zero for the ages between 60 to 80, generating little variations of mortality over time for those ages.
- The parameter ρ in Figure 4.5.6 bottom right panel increases until 1999 before it stagnates.

Other causes

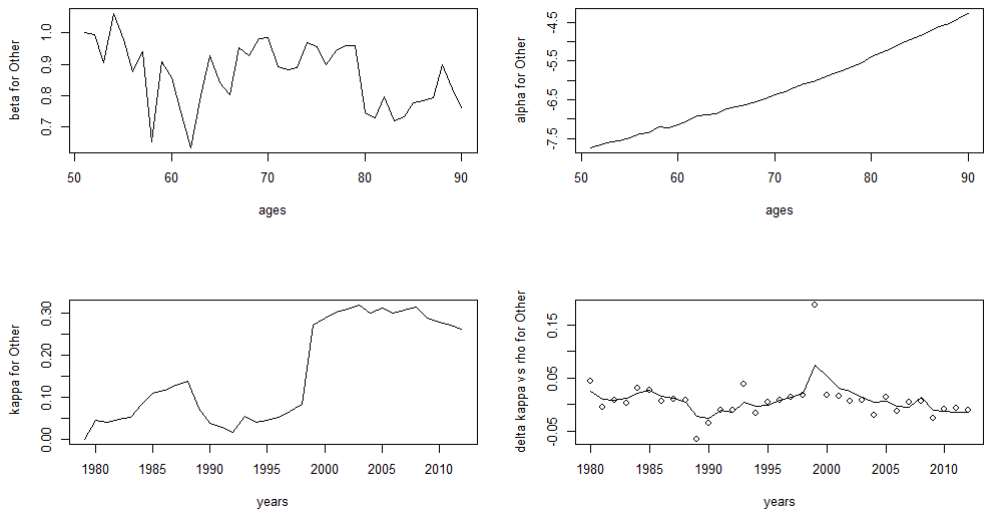


FIGURE 4.5.7 – *Structural parameters related to the Other causes of death.*

Figure 4.5.7 presents the structural parameters related to the Other causes of death.

- The parameter β in Figure 4.5.7 top left panel is positive and fluctuates dramatically.
- The parameter α in Figure 4.5.7 top right panel is increasing, linearly with age, illustrating the increasing probability of dying from Other causes while aging.
- The parameter κ in Figure 4.5.7 bottom left panel seems to have experienced a jump at the time of the ICD nomenclature changed. This jump aside, the parameter appears to be almost stationary. As the corrections of the multivariate time series may be difficult to process, we choose to not modify the data. The conterpart of this decision is the lack of information providing by the model for this particular cause of death.
- The parameter ρ in Figure 4.5.7 bottom right panel is close to 0, indicating a stagnation in the Other causes mortality over time.

Cause of death	Cancer	Metabolic	Mental	Circulatory	Respiratory	External	Others
Variance	4.62×10^{-5}	5.07×10^{-4}	1.14×10^{-5}	1.11×10^{-4}	1.68×10^{-5}	1.31×10^{-3}	1.10×10^{-3}

TABLE 4.5.2 – *Diagonal of the matrix variance* Σ_η

4.5.2 Stochastic dependency between cause-specific mortalities

Structural correlation analysis

We present the results regarding the covariance structure of the state variable dynamics. We denote the correlation matrices associated to the random variables η and ζ respectively $Corr_\eta$ and $Corr_\zeta$. These matrices allow us to introduce a stochastic dependency between the causes of death among the cause-specific error terms. The parameters η_t and ζ_t represent stochastic shocks on the state variables : the first shock represents a level variation of the parameter κ_t while the second shock represents a random change on the trajectory of the parameter κ_t.

We first present structural correlation analysis of η_t.

We observe in 4.5.2 that depending on the observed cause, the variance of the associated parameter η_t^i varies significantly according the cause of death. From the largest variance to the smallest, we obtain the following order : External, Other, Metabolic, Circulatory, Cancer, Respiratory and Mental. This means that level changes in the parameter κ_t were most observed for External causes while Mental diseases were less affected by this type of stochastic variation. Figure 4.5.8 displays the correlation structure of η_t.

We observe that shocks on Cancer are positively correlated with shocks on Circulatory and Other diseases. Shocks on Metabolic pathology are positively correlated with shocks affecting Circulatory and Respiratory diseases, and negatively associated with shocks on Mental diseases. Shocks on Mental pathologies are negatively correlated with shocks on Circulatory and Respiratory diseases. Shocks on Circulatory pathologies are positively correlated with shocks on respiratory and Other causes. Shocks on Respiratory diseases are positively correlated with shocks on External causes of death.

In consequence, the observation of a positive shock on the parameter κ_t associated with Circulatory diseases over the observation period and not related to a change in trajectory is likely to be concomitant with a decrease in the parameter κ_t associated with Mental diseases and an increase in the parameter κ_t associated with respiratory diseases. Taking into account the respective values of the para-

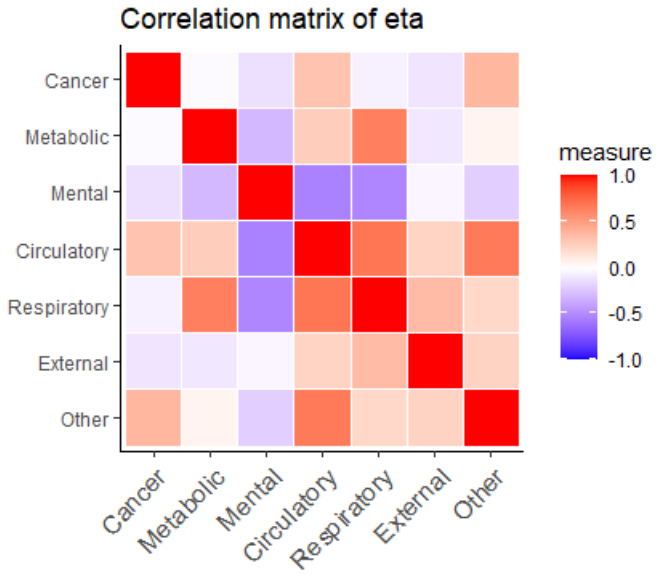


FIGURE 4.5.8 – *Heatmap of the matrix* $Corr_\eta$

Cause of death	Cancer	Metabolic	Mental	Circulatory	Respiratory	External	Others
Variance	2.43×10^{-5}	1.60×10^{-4}	1.76×10^{-6}	2.94×10^{-5}	8.19×10^{-7}	2.04×10^{-4}	2.26×10^{-4}

TABLE 4.5.3 – *diagonal of* Σ_ζ

meter κ_t for each of the causes, we obtain an increase in mortality for Circulatory, Mental diseases (from around 65 years of age) and an increase in mortality for respiratory diseases.

We now turn to the description of the variance structure covariance of the shocks ζ_t affecting the drift parameter ρ_t.

We notice in 4.5.3 that depending on the observed cause, the variance of the associated parameter ζ_t^i varies a lot. We note that the variance is smaller than the variance associated to the error terms η_t^i. From the largest variance to the smallest, we have : Other, External, Metabolic, Cancer, Circulatory, Mental and Respiratory. Figure 4.5.9 displays the correlation structure of ζ_t. We observe that the causes order is almost the same than for the η variance.

We observe that trajectory shocks for Cancer are positively correlated with trajectory shocks for respiratory diseases and negatively associated with trajectory shocks for Mental and external diseases. Shocks on metabolic pathologies are positively correlated with shocks affecting Circulatory diseases, and negatively associated with shocks on Mental diseases. Shocks on Mental pathologies are negatively correlated with shocks on Circulatory and Respiratory diseases. Shocks on circulatory pathologies are positively correlated with shocks on respiratory and Other causes. Shocks to respiratory diseases are positively correlated with shocks

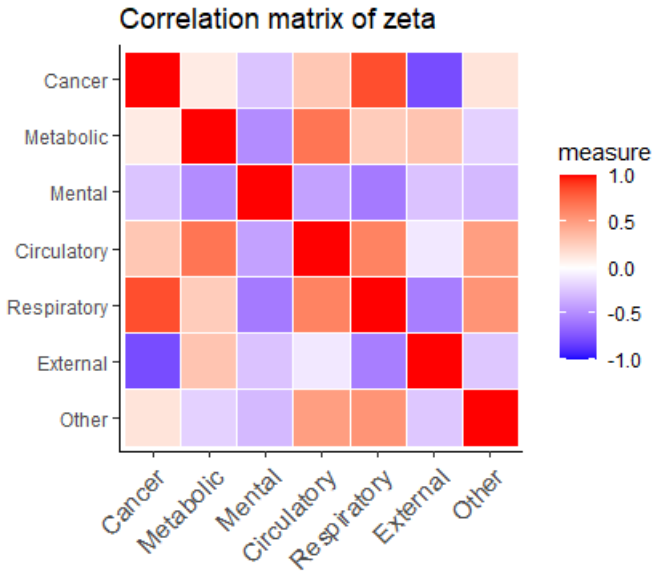


FIGURE 4.5.9 – *Heatmap of the matrix* $Corr_\zeta$

to Other causes of death and negatively correlated with shocks to external causes of death. Shocks on external diseases are negatively correlated with shocks affecting the category Other causes of death.

Thus, the observation of a positive shock on the parameter ρ_t affecting the Circulatory diseases over the observation period is likely to be concomitant with a decrease in the parameter ρ_t associated with External causes of death and an increase in the parameter ρ_t associated with respiratory diseases. This increase generates higher parameter κ_t over the following periods for deaths due to Circulatory and Respiratory diseases and a decrease for those due to external causes. This phenomenon increases mortality due to Circulatory and Respiratory diseases and decreases mortality due to external causes.

Illustration of the causes of death dependency

Due to the cross-effects between the different parameters κ_t, ρ_t and β_x, the interpretation, in terms of mortality by cause, of the correlation matrices developed in Section 4.5.2 is difficult. In the following section, we discuss the time dependency of the causes of death in mortality realizations. For this purpose, simulations are ran over the five-years period from 2013 to 2017. The number of deaths is then aggregated by ten-years age bands : 51-60, 61-70, 71-80 and 81-90 on which we compute the empirical covariance structure of dying by cause of death resulting from the simulations. We denote $\widehat{Corr}_\mu^{range}$, the inter-cause correlation matrix of the forces of mortality. Figures 4.5.10, 4.5.11, 4.5.12 and 4.5.13 present the heat-

map of the respective correlation matrices $\widehat{Corr}_{\mu}^{51-60}$, $\widehat{Corr}_{\mu}^{61-70}$, $\widehat{Corr}_{\mu}^{71-80}$ and $\widehat{Corr}_{\mu}^{81-90}$.

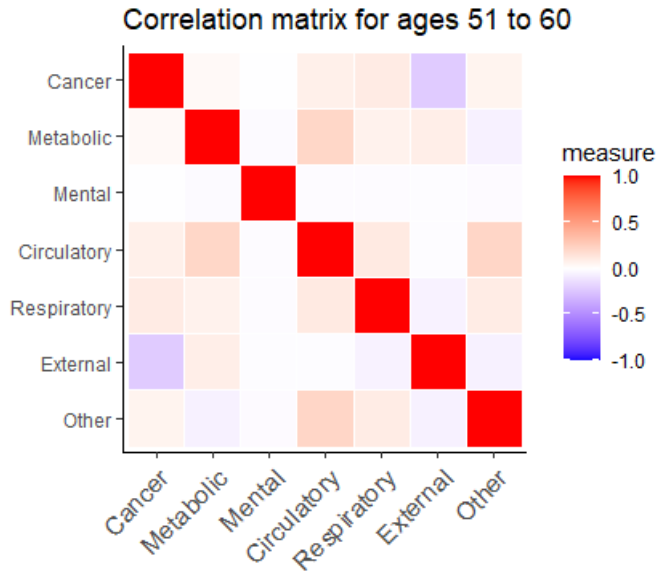


FIGURE 4.5.10 – *Heatmap of the matrix* $\widehat{Corr}_{\mu}^{51:60}$

Figure 4.5.10 presents the correlation heatmap for age group 51-60. We observe that the deaths from Cancer have a weak negative correlation with the External deaths. It means that during the period 2013-2017, for the age band 51-60, when mortality related to Cancer decreased, some of the deaths have been transferred to the external causes, while a decline of the general mortality is observed. Mortality from Metabolic diseases is positively correlated with mortality from Circulatory diseases. For this age group, no significant covariance is identified between mortality from Mental diseases and mortality from other cause of death. Deaths related to Circulatory diseases are positively correlated with deaths classified as Other causes. The other correlations are not very important.

Figure 4.5.11 displays the correlation heatmap between deaths by cause for age group 61-70. We observe that the deaths from Cancer are positively correlated with deaths related to Respiratory diseases. The correlation with the External causes and the other causes is near to 0. No significant covariance is identified between deaths from Mental diseases and deaths related to other causes. We notice an increase of the correlations between Metabolic, Circulatory, Respiratory and Other causes compared to the previous age band 51-60 in Figure 4.5.10. No other important is observed.

Figures 4.5.12 and 4.5.13 show the correlation structure $\widehat{Corr}_{\mu}^{71:80}$ and $\widehat{Corr}_{\mu}^{81:90}$.

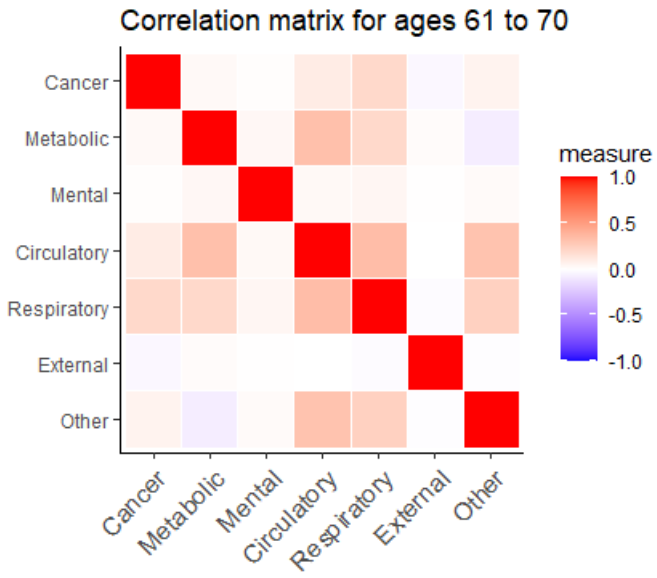


FIGURE 4.5.11 – *Heatmap of the matrix* $\widehat{Corr}_{\mu}^{61:70}$

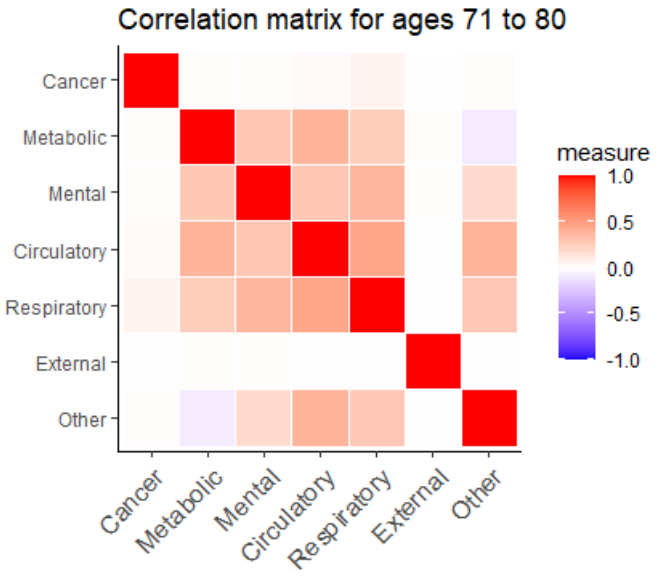


FIGURE 4.5.12 – *Heatmap of the matrix* $\widehat{Corr}_{\mu}^{71:80}$

Cancer and External causes no longer present correlations with other pathologies. We observe that a group of causes having highly correlated covariance has emerged. The deaths related to Metabolic, Mental, Circulatory and Respiratory diseases are highly positively correlated. The other correlations are not very important.

According to the description of the previous correlation structures, we notice that the deaths from External diseases and Cancer seem to be weakly correlated with other pathologies over the four age groups observed. The only exception for Cancer concerns External causes in the 51-60 age group and Respiratory diseases

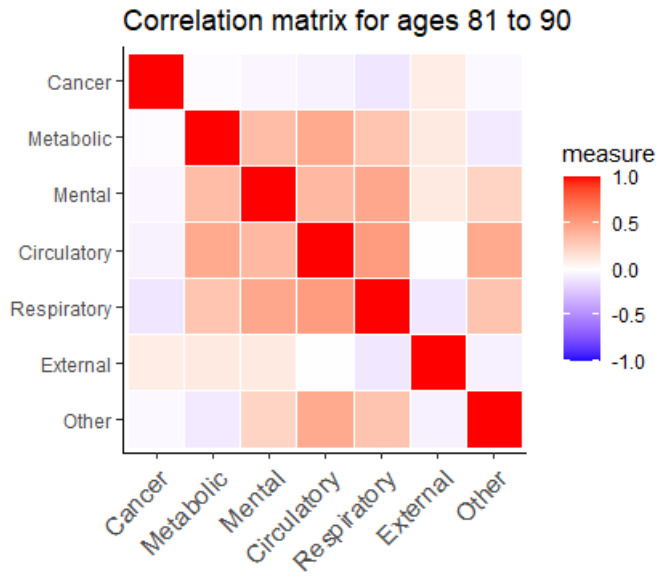


FIGURE 4.5.13 – *Heatmap of the matrix* $\widehat{Corr}_{\mu}^{81:90}$

in the 61-70 age group. It should be noted that among American women, the most deadly cancer in the last decade was lung cancer. Since a large proportion of these cancers are linked to smoking in the same way as certain Respiratory pathologies, it is not surprising that there is a positive correlation between the two pathologies due to the common risk factor. It should be noted that during the period of adjustment of the correlation parameters, the mean age at death was between 65 and 70 years, which supports the positive correlation on this age group specifically.

4.5.3 Mortality forecasting

Mortality forecast can be achieved using several approaches. The classical Lee-Carter model proposes to extrapolate the time varying parameter κ_t to the horizon $t_{n_y} + h$ using a random walk with drift. One thus obtains a parameter $\widehat{\kappa}_{t_{n_y}+h}$ and the resulting forecasted force of mortality $\widehat{\mu}_{x,t_{n_y}+h} = exp\left(\alpha_x + \beta_x \widehat{\kappa}_{t_{n_y}+h}\right)$. This approach has been used in Brouhns et al (2002)[20]. In the following comparisons, we apply this extrapolation on the parameters resulting from the state-space model and named this model A1. It has the advantage of maintaining log-linear growth, which is reassuring from a practical point of view. Nevertheless, this forecast is biased as already mentioned in Section 4.2. In our case, the extrapolation of the parameter κ is done by considering the last drift parameter $\rho_{t_{n_y}}$ calibrated by the model.

Another approach consists in extrapolating the model by considering future years as missing data. State-space models have been adapted to interpolate or extrapolate when inputs are lacking. Such an extrapolation method can be found in Launay et al (2013)[84]. The SMC algorithm is applied on the extrapolated period by omitting the density function as observations are not available. It is also necessary that the importance function is independent of the observations. In the context of bootstrap filtering, this condition is validated.

For the extrapolated years $t > t_{n_y}$ is therefore as follows :

(1) draw X_t^j with the distribution $q(x_t|X_{t_1:t-1})$ for $j = 1, ..., M$,

(2) compute $\widetilde{w}_t^j = w_{t-1}^j \frac{p(X_t^j|X_{t-1}^j)}{q(X_t^j|X_{t_1:t-1}^j)}$, and normalize $w_t^j = \frac{\widetilde{w}_t^j}{\sum_k \widetilde{w}_t^k}$.

This method has the advantage of being unbiased, but on the other hand, the trend obtained may not be log-linear and may even increase unrealistically when the variance of the model is too high. Thereafter, we refer to this forecasting approach as A2. Concerning the estimation of confidence intervals, Brouhns et al (2005) [19] suggest a bootstrapping approach which turns out to be analogous to the extension of the state-space model by SMC from a single realization of the parameter $\kappa_{t_{n_y}}$. In the following of this application, we compare the two approaches named previously as A1 and A2 as well as the standard LC model applied on each cause separately, named hereafter A3. The calibration period ranges from 1979 to 2012 and we compare the extrapolated series to the observed deaths over the period 2013-2017. In order to evaluate the adjustment, we use the sum of the Poisson deviances of the deaths :

$$d^i \propto \sum_{x,t} D_{x,t}^i \log\left(\frac{\widehat{\mu}_{x,t}^i}{\mu_{x,t}^i}\right) - (D_{x,t}^i - E_{x,t}\widehat{\mu}_{x,t}^i) \tag{5.1}$$

More details on age and period accuracy by cause of death are provided in Appendix B. We can observe in these figures that the standard Lee-Carter have an important bias for all the ages while the two variants of the state-space models are much more accurate. We provide in table 4.5.4 a synthesis of the result by cause of death.

We observe that in 6 out of 7 causes of death as well as for the resulting aggregated mortality, approach A2 outperforms the two other approaches in terms of deviance. We also observe that the model A3, the standard Lee-Carter, leads always to the worst extrapolation among the three approaches. For some causes of death like Circulatory, the error obtained is particularly important compared to

Cause of death	Model A1	Model A2	Model A3
Cancer	6487	**6442**	4087732
Metabolic	**493**	**493**	911055
Mental	6000	**4351**	425984
Circulatory	8644	**6460**	5793081
Respiratory	2340	**2165**	1517111
External	**627**	694	569906
Other	3414	**3059**	1677223
All-causes	11853	**10765**	14777327

Table 4.5.4 – *Extrapolation accuracy of the three models ratio by cause and method.*

the other approaches. These large deviations from the observed mortality lead to large resulting aggregated mortality deviance. In another hand, the adaptability of trends offered by the state-space model leads to much more reliable results.

4.6 Conclusion

In this paper, we propose a general framework to forecast mortality by cause of death. The approach consists in a Poisson Lee-Carter state-space model with a variance-covariance structure allowing time dependency between causes and trend shift. We provide a comprehensive part on the estimation process for this kind of model, including smoothing, filtering and EM algorithm. The appendix A provides details on the calibration specificities for the Poisson model developed in this chapter. The model have been applied on the US female population for 7 different causes of death. We show on the one hand that the state-space model as defined in this paper is an interesting approach to take into account the observed time dependency structure between the causes of death, which may then be submitted to the interpretation of medical experts. On the other hand, its accuracy is higher than the standard Lee-Carter model, which we explain by the inclusion of trend changes in the model. Further research may be conducted regarding the cause-specific mortality projections, for example a possible extension of our approach could include dependent cohort effect. In addition, comparisons of other reconciliation methods applied to the state-space lee-carter model would be interesting. The time series obtained may be reconciled using the approach developed in Hyndman et al. (2011)[73] or transformed according to the Li and Lu (2019)[88] to retrieve the net force of mortality. We plan to improve this model by adding cohort effetcts which to eliminate potential biases affecting the estimation for causes like circulatory and mental. The question of the robustness of the parameter is

also an important issue we would like to adress. Comparisons between the model coefficients are difficult because of chosen constraints, and different constraints could allows a better interpretability of the model and more robust forecast.

Appendix

A Details on EM algorithm

A.1 $\widehat{Q}_1$ maximisation

We start by defining the probability measures related to the maximization of $\widehat{Q}_1$. The variable X_{t_1} is normally distributed as $N\left(\bar{x}_{t_1}, \Sigma_{\rho_{t_1}}\right)$. We remind that $X_{t_1} = \left(\kappa'_{t_1}, \rho'_{t_1}\right)'$ and assume independence between the two variables. It is thus equivalent to consider for each of the state variables κ_{t_1} and ρ_{t_1} the structural parameters $\bar{\kappa}_{t_1}$ and $\Sigma_{\kappa_{t_1}}$ in one side and $\bar{\rho}_{t_1}$ and $\Sigma_{\rho_{t_1}}$ in the other side. As the parameters of the Lee-Carter model from which this model is derived must be constrained to ensure identifiability, we can apply to the parameter κ_{t_1} one of these constraints assuming that $\kappa_{t_1} = 0$. Thus, we only have to estimate the parameters associated to ρ_{t_1}, i.e. $p_\theta\left(x_{t_1}\right) = p_\theta\left(\rho_{t_1}\right)$. The multivariate parameter ρ_{t_1} being Gaussian, we have :

$$p_\theta(\rho_{t_1}) = \frac{1}{\sqrt{\det(2\pi\Sigma_{\rho_{t_1}})}} \exp\left[-\frac{1}{2}(\rho_{t_1} - \overline{\rho}_{t_1})'\Sigma^{-1}_{\rho_{t_1}}(\rho_{t_1} - \overline{\rho}_{t_1})\right], \tag{1.1}$$

which in log-scale becomes :

$$log(p_\theta(\rho_{t_1})) = -\frac{1}{2}log\left(\det\left(\Sigma_{\rho_{t_1}}\right)\right) - \frac{1}{2}(\rho_{t_1} - \overline{\rho}_{t_1})'\Sigma^{-1}_{\rho_{t_1}}(\rho_{t_1} - \overline{\rho}_{t_1}) + C, \tag{1.2}$$

where C is a constant independent of θ.

We denote by $\rho^{(j)}_{t_1}$ the j-th simulation of ρ_{t_1} obtained from the smoothing algorithm. Using the weights $w^i_{t|t_{n_y}}$ from (3.14) and noticing they sum up to one,

a closed form can be derived for the parameters $\Sigma_{\rho_{t_1}}$ and ρ_{t_1} :

$$\begin{aligned}\widehat{Q}_1 &= \sum_{j=1}^{M} w^j_{1|t_{n_y}} \log p_\theta \left(\rho^{(j)}_{t_1}\right) \\ &= -\frac{1}{2}\sum_{j=1}^{M} w^j_{1|t_{n_y}} \left[log\left(\det\left(\Sigma_{\rho_{t_1}}\right)\right) + (\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})'\Sigma^{-1}_{\rho_{t_1}}(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})\right] \\ &= -\frac{1}{2} log\left(\det\left(\Sigma_{\rho_{t_1}}\right)\right) - \frac{1}{2}\sum_{j=1}^{M} w^j_{1|t_{n_y}} (\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})'\Sigma^{-1}_{\rho_{t_1}}(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1}).\end{aligned} \tag{1.3}$$

This function is concave as a positive weighted sum of concave functions. In consequence, the maximum can be obtained by equating the gradient to zero. We consider first the part of the gradient related to $\overline{\rho}_{t_1}$. The derivatives with respect to the mean parameter for the simulation $\rho^{(j)}_{t_1}$ is :

$$\frac{\partial log(p_\theta(\rho^{(j)}_{t_1}))}{\partial \overline{\rho}_{t_1}} = \Sigma^{-1}_{t_1}(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1}). \tag{1.4}$$

Then weighting these derivative and summing over j, we obtain :

$$\frac{\partial \sum_{j=1}^{M} w^j_{1|t_{n_y}} log(p_\theta(\rho^{(j)}_{t_1}))}{\partial \overline{\rho}_{t_1}} = \Sigma^{-1}_{t_1}\left(\sum_{j=1}^{M} w^j_{1|t_{n_y}} \rho^{(j)}_{t_1} - \overline{\rho}_{t_1}\right). \tag{1.5}$$

By setting this equation equal to zero we obtain the solution :

$$\overline{\rho}_{t_1} = \sum_{j=1}^{M} w^j_{1|t_{n_y}} \rho^{(j)}_{t_1}, \tag{1.6}$$

which is unique because of the concavity of the objective function. The derivative for a simulation $\rho^{(j)}_{t_1}$ with respect to the variance parameter $\Sigma_{\rho_{t_1}}$ is :

$$\frac{\partial log(p_\theta(\rho^{(j)}_{t_1}))}{\partial \Sigma_{\rho_{t_1}}} = -\frac{1}{2}\left(\Sigma^{-1}_{\rho_{t_1}} - \Sigma^{-1}_{\rho_{t_1}}(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})'\Sigma^{-1}_{\rho_{t_1}}\right). \tag{1.7}$$

It results that the weighted sum by $w^j_{1|t_{n_y}}$ of the partial derivatives is :

$$\frac{\partial \widehat{Q}_1}{\partial \Sigma_{\rho_{t_1}}} = -\frac{1}{2}\left(\Sigma^{-1}_{\rho_{t_1}} - \Sigma^{-1}_{\rho_{t_1}}\left[\sum_{j=1}^{M} w^j_{1|t_{n_y}} (\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})'\right]\Sigma^{-1}_{\rho_{t_1}}\right), \tag{1.8}$$

which we equalize to zero to obtain :

$$\Sigma_{\rho_{t_1}} = \sum_{j=1}^{M} w^j_{1|t_{ny}} (\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})(\rho^{(j)}_{t_1} - \overline{\rho}_{t_1})'. \tag{1.9}$$

A.2 $\widehat{Q}_2$ maximization

Before tackling the issue of maximizing $\widehat{Q}_2$, we notice that the weights obtained in equation (4.13) also sum up to one and that $\widehat{Q}_2$ depends on the sub-vector $(\text{vec}(\Sigma_\eta)', \text{vec}(\Sigma_\zeta)')'$ of the parameter vector θ. In this section, we study the function $log\left(p_\theta\left(x_t|x_{t-1}\right)\right)$ and its derivative, which can be calculated using the dynamic of $X_t = (\kappa'_t, \rho'_t)'$. Hence, we obtain $p_\theta(x_t|x_{t-1}) = \mathcal{N}\left(x_t; (Ax_{t-1}, \Sigma_\epsilon)\right)$ where $\Sigma_\epsilon = \text{blockdiag}(\Sigma_\eta, \Sigma_\zeta)$ and $\mathcal{N}\left(x; (\mu, \Sigma)\right)$ denotes the density of a Gaussian of mean μ, variance Σ taken at point x.
We observe that :

$$\begin{aligned} log\left(p_\theta(x_t|x_{t-1})\right) &= log\left(p_\theta(\kappa_t|x_{t-1})\right) + log\left(p_\theta(\rho_t|x_{t-1})\right) \\ &= log\left(\mathcal{N}\left(\kappa_t; (A_{sup}x_{t-1}, \Sigma_\eta)\right)\right) + log\left(\mathcal{N}\left(\rho_t; (A_{inf}x_{t-1}, \Sigma_\zeta)\right)\right). \end{aligned} \tag{1.10}$$

Because each of the parameters κ_t and ρ_t depends on separate parts of the vector θ, and due to the independence between these two variables, the associated functions can be maximized separately. Therefore, we only develop the maximization procedure for the parameter Σ_η associated to the dynamic of κ_t, the estimation phase for Σ_ζ being analogous. Similar arguments than those developed in Section A.1 allows us to write the objective function depending on Σ_η as :

$$\widehat{Q}_2 = -\frac{t_{n_y} - 1}{2} log\left(\det\left(\Sigma_\eta\right)\right) - \frac{1}{2} \sum_{t=t_1}^{t_{n_y}-1} \sum_{i,j=1}^{M} w^{ij}_{t|t_{n_y}} \left(\kappa^{(j)}_{t+1} - A_{sup}x^i_t\right)' \Sigma_\eta^{-1} \left(\kappa^{(j)}_{t+1} - A_{sup}x^i_t\right) + C, \tag{1.11}$$

with C being not related to Σ_η. The function $\widehat{Q}_2$ is then concave regarding Σ_η, as the sum of positively weighted concave functions. We can now obtain the estimator of Σ_η by equating the gradient to zero. The differential for a couple of observations $\kappa^{(j)}_{t+1}$ and x^i_t with respect to the variance matrices is :

$$\frac{\partial log(p(\kappa_t|x_{t-1}))}{\partial \Sigma_\eta} = -\frac{1}{2}\left(\Sigma_\eta^{-1} - \Sigma_\eta^{-1}(\kappa^{(j)}_{t+1} - A_{sup}x^i_t)(\kappa^{(j)}_{t+1} - A_{sup}x^i_t)'\Sigma_\eta^{-1}\right). \tag{1.12}$$

We then obtain

$$\frac{\partial \widehat{Q}_2}{\partial \Sigma_\eta} = -\frac{t_{n_y}-1}{2}\Sigma_\eta^{-1} - \frac{1}{2}\Sigma_\eta^{-1}\sum_{t=t_1}^{t_{n_y}-1}\sum_{i,j=1}^{M} w_{t|t_{n_y}}^{ij}\left(\kappa_{t+1}^{(j)} - A_{sup}x_t^i\right)\left(\kappa_{t+1}^{(j)} - A_{sup}x_t^i\right)'\Sigma_\eta^{-1}. \tag{1.13}$$

In consequence the estimator is :

$$\widehat{\Sigma}_\eta = \frac{1}{t_{n_y}-1}\sum_{t=t_1}^{t_{n_y}-1}\sum_{i,j=1}^{M} w_{t|t_{n_y}}^{ij}\left(\kappa_{t+1}^{(j)} - A_{sup}x_t^i\right)\left(\kappa_{t+1}^{(j)} - A_{sup}x_t^i\right)'. \tag{1.14}$$

A.3 $\widehat{Q}_3$ maximization

The remaining parameters cannot be obtained through a closed formula, therefore we have to employ numerical optimization algorithm like Newton-Raphson. In order to simplify the estimation, we provide the gradient function of $\widehat{Q}_3$. The objective function depends on the space variable realization and the state variable $x_t = (\kappa_t', \rho_t')'$. In this model, the space variable Y_t vector is $(D_{x,t}^i)_{x,i}$: the number of deaths at age x, year t and due to cause i. As all the components of this vector conditionally on x_t are independent, we have the following likelihood :

$$p(Y_t|x_t) = \prod_x \prod_i \mathscr{P}\left(\lambda_{x,t}^i; exp\left(\alpha_x^i + \beta_x^i\kappa_t^i\right)E_{x,t}\right), \tag{1.15}$$

where $\mathscr{P}(D_{x,t}^i; \lambda_{x,t}^i)$ refers to the Poisson probability associated with the realisation $D_{x,t}^i$ and the parameter $\lambda_{x,t}^i$, i.e. $\mathscr{P}(D_{x,t}^i; \lambda_{x,t}^i) = \frac{\left(\lambda_{x,t}^i\right)^{D_{x,t}^i}}{D_{x,t}^i!}exp\left(-\lambda_{x,t}^i\right)$. The logarithmic transformation of $p(Y_t|x_t)$ results in :

$$log(p(Y_t|X_t)) = \sum_x \sum_i D_{x,t}^i\left(\alpha_x^i + \beta_x^i\kappa_t^i\right) - E_{x,t}exp\left(\alpha_x^i + \beta_x^i\kappa_t^i\right) + Cst\left(E_{x,t}, D_{x,t}^i\right). \tag{1.16}$$

We then obtain :

$$\begin{aligned}
\frac{\partial log(p(Y_t|X_t))}{\partial \alpha_x^i} &= D_{x,t}^i - E_{x,t}exp\left(\alpha_x^i + \beta_x^i\kappa_t^i\right) = D_{x,t}^i - E_{x,t}\lambda_{x,t}^i, \\
\frac{\partial log(p(Y_t|X_t))}{\partial \beta_x^i} &= D_{x,t}^i\kappa_t^i - E_{x,t}exp\left(\alpha_x^i + \beta_x^i\kappa_t^i\right)\kappa_t^i = \kappa_t^i\frac{\partial log(p(Y_t|X_t))}{\partial \alpha_x^i}.
\end{aligned} \tag{1.17}$$

We denote by $\kappa_t^{i,(j)}$ the j^{th} simulation of κ_t^i. By summing over all simulations and years, we obtain :

$$\begin{aligned}
\frac{\partial \widehat{Q}_3}{\partial \alpha_x^i} &= \sum_{t=t_1}^{t_{n_y}} D_{x,t}^i - \sum_{t=t_1}^{t_{n_y}} E_{x,t} \sum_{j=1}^{M} w_{1|t_{n_y}}^j exp\left(\alpha_x^i + \beta_x^i \kappa_t^{i,(j)}\right), \\
\frac{\partial \widehat{Q}_3}{\partial \beta_x^i} &= \sum_{j=1}^{M} w_{1|t_{n_y}}^j D_{x,t}^i \kappa_t^{i,(j)} - \sum_{j=1}^{M} w_{1|t_{n_y}}^j E_{x,t} exp\left(\alpha_x^i + \beta_x^i \kappa_t^{i,(j)}\right) \kappa_t^{i,(j)}.
\end{aligned} \tag{1.18}$$

This gradient function is then included in the optimization process to maximize the cost function $\widehat{Q}_3$.

B Residuals of the models

B.1 Residuals of the A1 model

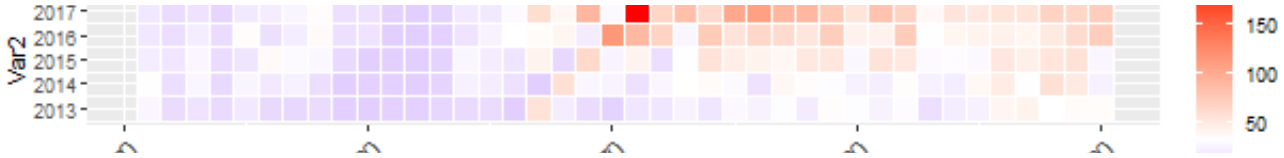


Figure B.1 – *Deviance of the state-space LC with simple extrapolation by age and year for cancer*

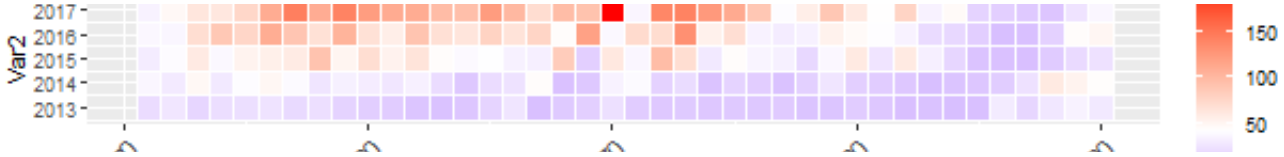


Figure B.2 – *Deviance of the state-space LC with simple extrapolation by age and year for circulatory*

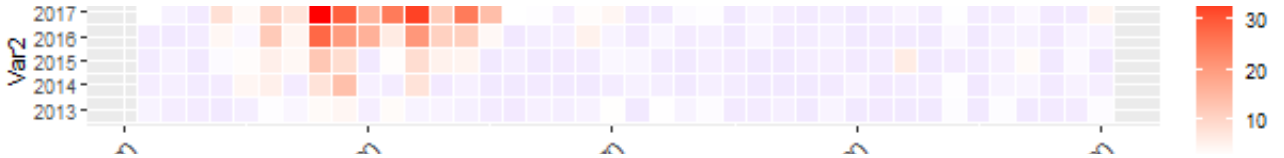


Figure B.3 – *Deviance of the state-space LC with simple extrapolation by age and year for external*

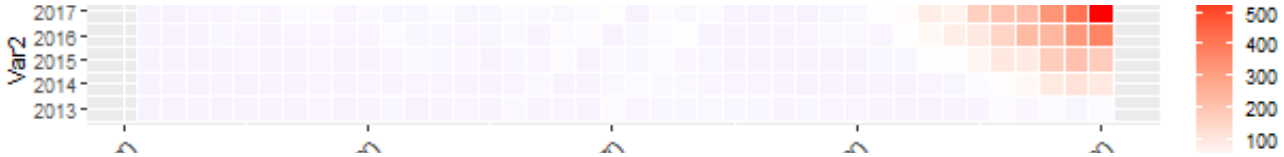


Figure B.4 – *Deviance of the state-space LC with simple extrapolation by age and year for mental and nervous*

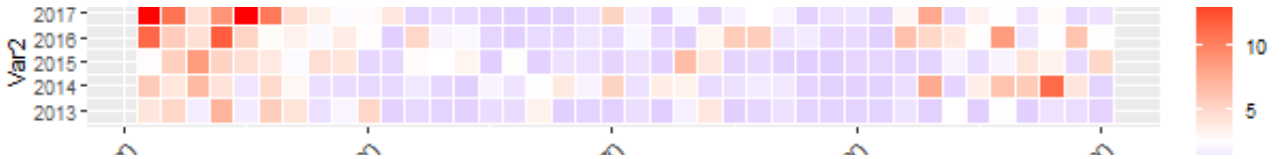


Figure B.5 – *Deviance of the state-space LC with simple extrapolation by age and year for metabolic*

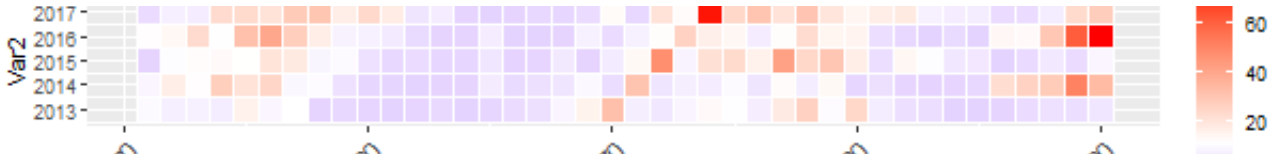


Figure B.6 – *Deviance of the state-space LC with simple extrapolation by age and year for respiratory*

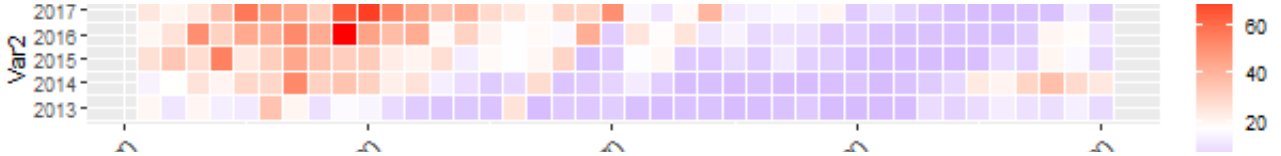


Figure B.7 – *Deviance of the state-space LC with simple extrapolation by age and year for other causes*

B.2 Residuals of the A2 model

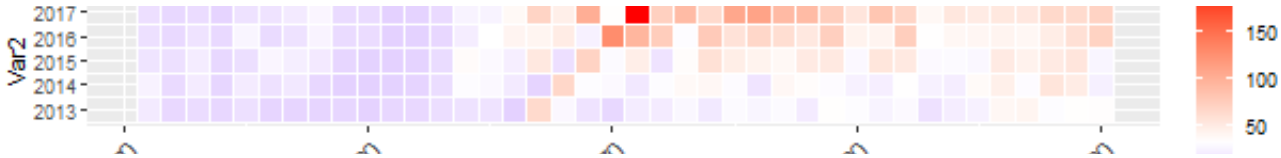


Figure B.8 – *Deviance of the original state-space LC extrapolation by age and year for cancer*

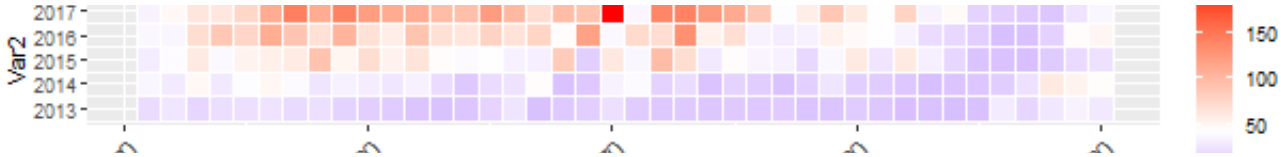


Figure B.9 – *Deviance of the original state-space LC extrapolation by age and year for circulatory*

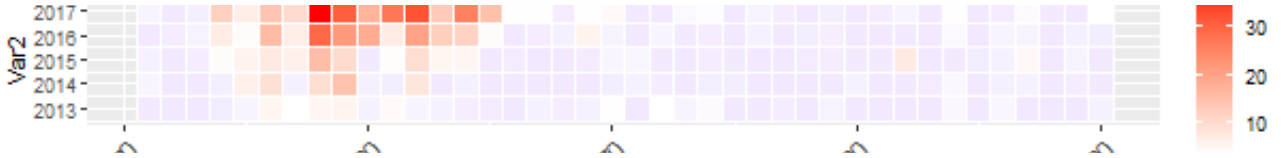


Figure B.10 – *Deviance of the original state-space LC extrapolation by age and year for external*

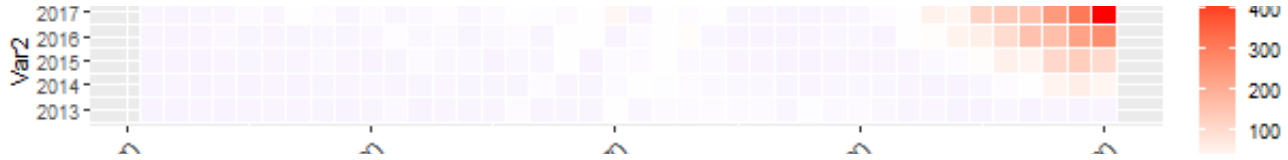


Figure B.11 – *Deviance of the original state-space LC extrapolation by age and year for mental and nervous*

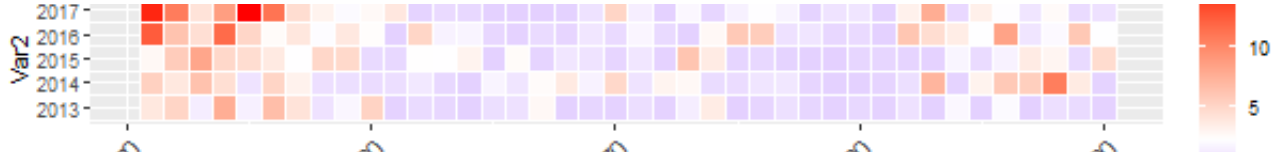


Figure B.12 – *Deviance of the original state-space LC extrapolation by age and year for metabolic*

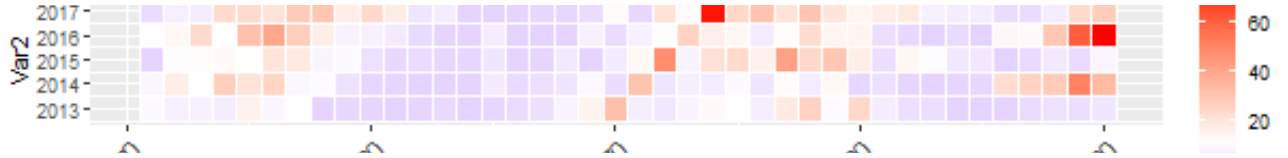


Figure B.13 – *Deviance of the original state-space LC extrapolation by age and year for respiratory*

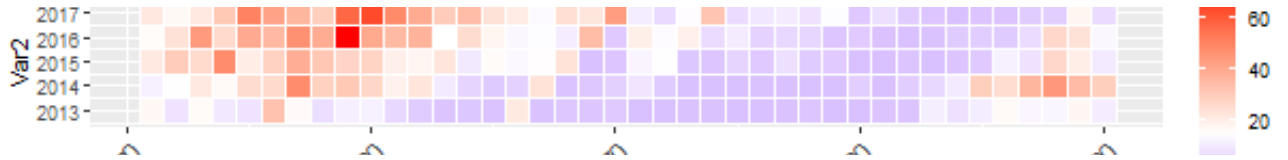


Figure B.14 – *Deviance of the original state-space LC extrapolation by age and year for other causes*

B.3 Residuals of the A3 model

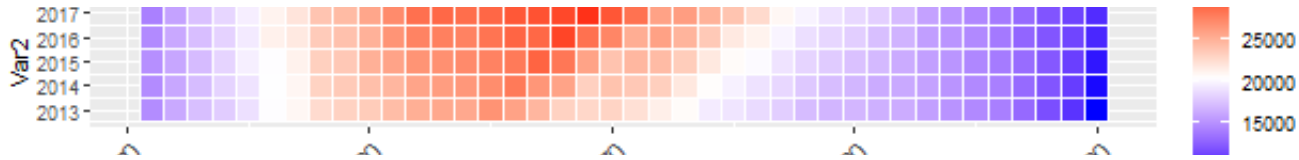


Figure B.15 – *Deviance of the standard LC extrapolation by age and year for cancer*

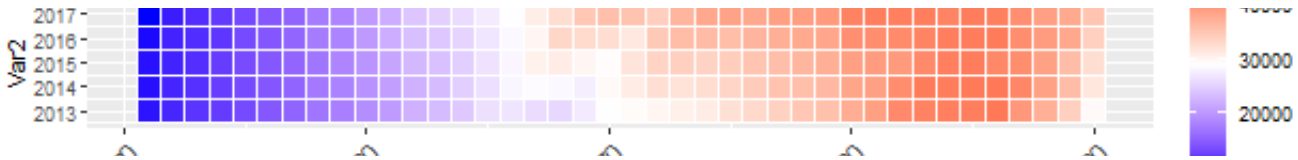


Figure B.16 – *Deviance of the standard LC extrapolation by age and year for circulatory*

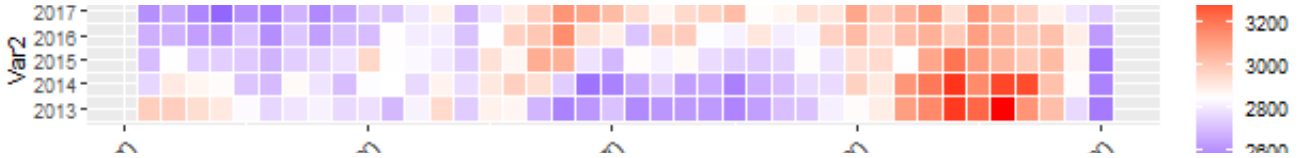


Figure B.17 – *Deviance of the standard LC extrapolation by age and year for external*

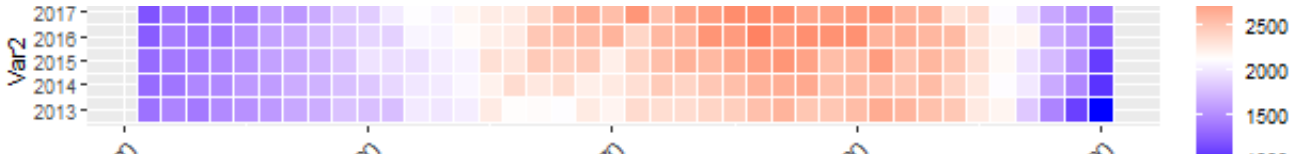


Figure B.18 – *Deviance of the standard LC extrapolation by age and year for mental and nervous*

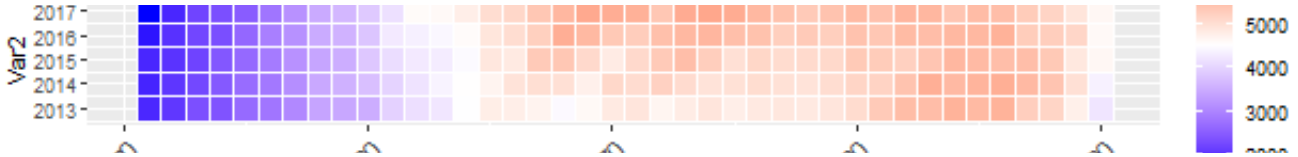


Figure B.19 – *Deviance of the standard LC extrapolation by age and year for metabolic*

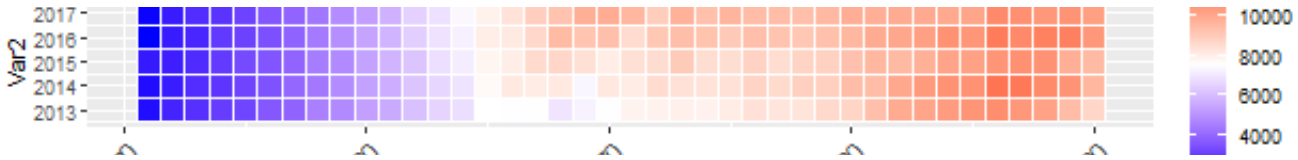


Figure B.20 – *Deviance of the standard LC extrapolation by age and year for respiratory*

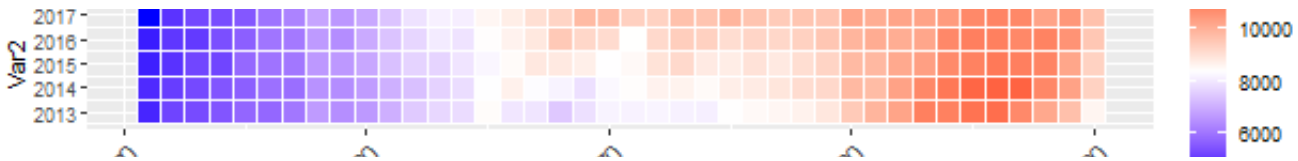


Figure B.21 – *Deviance of the standard LC extrapolation by age and year for other causes*

Chapitre 5

Conclusion

Nous avons présenté dans ce travail de thèse trois apports à la théorie de la modélisation de la mortalité par cause de décès. Dans la première partie, il a été question d'extrapolation de la mortalité par cause aux âges avancés. Nous avons proposé une approche associant la robustesse des modèles d'extrapolation pour la mortalité toutes causes confondues et la souplesse des P-splines. Une analyse attentive de la vraisemblance nous a permis de scinder le problème en deux parties : la première relative à la mortalité toutes causes et la seconde ayant trait à la contribution des causes à la mortalité générale. En supposant des paramètres différents pour ces deux parties, nous avons montré que le problème revenait d'un côté à trouver le modèle d'extrapolation aux grands âges le plus performant pour la mortalité toutes causes et de l'autre à projeter finement les proportions relatives de chacune des causes de décès par rapport à la mortalité globale. Dans le cadre de la régression des contributions des causes à la mortalité, nous avons avancé deux nouveaux modèles multinomiaux inspirées de la théorie des compositional data analysis développée par Aitchison (1982)[2]. Nous avons appliqué cette approche sur la population féminine américaine afin d'étudier les différences entre chacune des combinaisons modèle multinomial et méthode de fermeture.

Dans la deuxième étude, nous avons proposé une méthode permettant de grouper les caractéristiques (âge, sexe, cause de décès, lieux, ...) de dynamiques similaires du point de vue d'une modélisation par Lee-Carter. Les tendances ont été identifiées en cherchant à minimiser l'erreur d'ajustement obtenue par l'application du modèle de Lee-Carter sur un nombre défini de groupes. Cette classification non-supervisée visait à remplacer la distinction arbitraire qui précède d'ordinaire la calibration des modèles de Lee-Carter. Cet algorithme s'apparente à une méthode de partitionnement K-centroid, dans laquelle les variables sont les éléments de la série temporelle de mortalité. Nous avons montré que cette méthode, appelée

K-Lee-Carter, est une variante des K-centroid pour une distance définie comme la corrélation de Pearson pondérée. L'algorithme qui en résulte a permis d'accroître le caractère explicatif du modèle et d'identifier des tendances de mortalité similaires, lesquelles peuvent par exemple éclairer sur les facteurs de risques communs à certaines caractéristiques. Deux autres versions de l'algorithme ont été proposées afin de lisser les résultats et d'éviter à deux séries de mortalité jugées d'habitude proches de se retrouver dans deux groupes différents. Là encore une application à la population américaine a été proposée, et notamment sur la mortalité par cause des femmes. Il a été montré que pour la période considérée, le modèle des K-Lee-Carter conduit à de meilleurs résultats tant du point de vue de l'ajustement que de celui de la prédiction.

Dans la troisième partie de cette thèse, nous avons proposé un modèle État-Espace permettant de traiter trois problèmes fondamentaux dans la projection de la mortalité par cause, à savoir : les changements de tendances, la présence d'un biais dans les prévisions de mortalité, et l'intégration d'une structure de dépendance entre causes de décès. Afin de rester dans un cadre connu, nous avons employé pour chacune des causes le modèle de LC-Poisson tel que présenté dans Brouhns et al. (2002)[20]. Les paramètres temporels du modèle de Lee-Carter ont été considérés comme des variables d'état auxquelles nous avons attribué une structure de dépendance multivariée afin d'intégrer la dépendance temporelle entre les causes. Enfin, le problème des changements de tendances des paramètres Kappa du modèle a été pris en considération grâce à l'ajout d'une dynamique au paramètre de pas ρ. La calibration du modèle reprend la méthode fréquentiste proposée par Schön (2011)[123], laquelle consiste à estimer les paramètres du modèle au moyen d'un algorithme Expectation-Maximization. Nous avons adapté cette méthode au modèle de projection de mortalité, et avons fourni pour un certain nombre de paramètres, des formules fermées pour accroître la vitesse de calcul. Une application à la population féminine américaine entre 1979 et 2012 a été proposée. Les structures de dépendance obtenues ont été détaillées, ainsi que leurs impacts sur la dépendance entre décès par cause. À la suite de tests, nous avons pu observer que pour la période considérée, le modèle proposé était très nettement supérieur à l'approche classique consistant à extrapoler chacune des causes séparément au moyen d'un modèle de Lee-Carter usuel. Ainsi en plus d'indications potentiellement éclairantes quant à la structure de dépendance temporelle des causes de décès, le modèle s'est avéré fournir des prévisions intéressantes.

Au-delà des apports propres de chacun de ces travaux, il nous est apparu que

l'aspect le plus enrichissant de cette thèse résidait dans l'interaction potentielle des différentes approches. Ainsi après avoir extrapolé la mortalité aux grands âges pour différentes causes, il serait possible de la diviser en groupes de dynamique homogène pour ensuite appliquer un modèle État-Espace permettant d'inclure la dépendance entre groupes. C'est dans cette optique que ces travaux ont été réalisés et des applications croisant les différentes approches devraient voir le jour prochainement.

Bibliographie

[1] O. Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6 :701–726, 1978.

[2] J. Aitchison. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2) :139–177, 1982.

[3] D. Alai, S. Arnold (-Gaille), and M. Sherris. Modelling cause-of-death mortality and the impact of cause-elimination. *Annals of Actuarial Science*, 9(1) :167–186, 2015.

[4] J. Alho. Stochastic methods in population forecasting. *International journal of forecasting*, 6 :521–30, 1991.

[5] J. Alho and B.D. Spencer. A population forecast as a database : Implementing the stochastic propagation of error. *Journal of Official Statistics*, 7 :295–310, 1991.

[6] L. Alkema, D. Chou, D. Hogan, S. Zhang, A.B. Moller, A. Gemmill, D. Ma Fat, T. Boerma, M Temmerman, C. Mathers, and L Say. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030 : a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group. *The Lancet*, 387(10017) :462–474, 2016.

[7] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, 1993.

[8] P. Andersson and M. Lindholm. Mortality forecasting using a lexis-based state-space model. *Annals of Actuarial Science*, pages 1–30, 2020.

[9] A. Arnold (-Gaille) and M. Sherris. Causes-of-death mortality : What do we know on their dependence ? *North American Actuarial Journal*, 19(2) :116–128, 2015.

[10] S. Arnold (-Gaille) and M. Sherris. Forecasting mortality trends allowing for cause-of-death mortality dependence. *North American Actuarial Journal*, 17 :273–282, 2013.

[11] S. Arnold-Gaille and M. Sherris. International cause-specific mortality rates : New insights from a cointegration analysis. *ASTIN Bulletin*, 46(1) :9–38, 2016.

[12] G. Athanasopoulos, R. Ahmed, and R. Hyndman. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25(1) :146–166, 2009.

[13] W. Bell and B. Monsell. Using principal components in time series modeling and forecasting of age-specific mortality rates. 1991.

[14] M.P. Bergeron-Boucher, V. Canudas-Romo, J. Oeppen, and J. Vaupel. Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37 :527–566, 2017.

[15] D. Blake, A. Cairns, K. Dowd, and A. Kessler. Still living with mortality : The longevity risk transfer market after one decade. *SSRN Electronic Journal*, 24, 2018.

[16] J. Bongaarts. Trends in senescent life expectancy. *Population Studies*, 63(3) :203–213, 2009.

[17] H. Booth, R. Hyndman, L. Tickle, and P. De Jong. Lee-Carter mortality forecasting : A multi-country comparison of variants and extensions. *Demographic Research*, 15 :289–310, 2006.

[18] G. Box and G. Jenkins. *Time Series Analysis Forecasting And Control*. Holden-Day, 1970.

[19] N. Brouhns, M. Denuit, and I. Keilegom. Bootstrapping the Poisson log-bilinear model for mortality projection. *Scandinavian Actuarial Journal*, pages 212–224, 2005.

[20] N. Brouhns, M. Denuit, and J. Vermunt. A poisson log-bilinear regression approach to the construction of projected life tables. *Insurance : Mathematics and Economics*, 31 :373–393, 2002.

[21] A. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty : Theory and calibration. *Journal of Risk & Insurance*, 73(4) :687–718, 2006.

[22] D. Cairns, A.and Blake, K. Dowd, G. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from england & wales and the united states. *North American Actuarial Journal*, 13 :1–35, 2009.

[23] E. Cannan. The probability of a cessation of the growth of population in england and wales during the next century. *The Economic Journal*, 5(20) :505, 1895.

[24] P. Carracedo, A. Debón, A. Iftimi, and F. Montes. Detecting spatio-temporal mortality clusters of European countries by sex and age. *International journal for equity in health*, 17 :38, 2018.

[25] J. Carriere. Dependent decrement theory (with discussion). *Transactions of the Society of Actuaries*, 46 :45–74, 1994.

[26] M. Cervellati and U. Sunde. Human capital formation, life expectancy, and the process of development. *American Economic Review*, 95(5) :1653–1672, 2005.

[27] C. L. Chiang. *Introduction to stochastic processes in biostatistics*. John Wiley and Sons, New York, 1st edition, 1968.

[28] A.J. Coale and Ellen Kisker. Defects in data on old-age mortality in the United States : New procedures for calculating mortality schedules and life tables at the highest ages. *Asian & Pacific Population Forum*, 4 :1–31, 1990.

[29] J. Cossman, R. Cossman, W. James, C. Campbell, C. Troy, Blanchard, and A. Cosby. Persistent clusters of mortality in the United States. *American journal of public health*, 97–12 :2148–50, 2007.

[30] D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2) :187–220, 1972.

[31] E.M. Crimmins. The changing pattern of american mortality decline, 1940-77, and its implications for the future. *Population and Development Review*, 7(2) :229–254, 1981.

[32] I. Currie. On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4) :356–383, 2016.

[33] I. Currie, M. Durbán, and P. Eilers. Smoothing and forecasting mortality rates. *Statistical Modelling*, 4 :279–298, 2004.

[34] P. Day. *A new history of social welfare*. Sixth edition. Boston, MA : Pearson/Allyn and Bacon, 2009.

[35] C. de Boor. *A Practical Guide to Spline*, volume 27. Applied Mathematical Sciences, New York : Springer, 1978.

[36] P. De Jong and L. Tickle. Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, 13(1) :1–18, 2006.

[37] A. Debón, L. Chaves, S. Haberman, and F. Villa. Characterization of between-group inequality of longevity in european union countries. *Insurance : Mathematics and Economics*, 75, 2017.

[38] A. Debón, F. Martinez Ruiz, and F. Montes. A geostatistical approach for dynamic life tables : The effect of mortality on remaining lifetime and annuities. *Insurance : Mathematics and Economics*, 47 :327–336, 2010.

[39] A. Delwarde and M. Denuit. *Construction de tables de mortalité périodiques et prospectives*. Economica, 2005.

[40] A. DeMoivre. *Annuities on lives : Or, the valuation of annuities upon any number of lives as also of reversions*. William Person, 1725.

[41] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39 :1–38, 1977.

[42] M. Denuit and A.C. Goderniaux. Closing and projecting lifetables using log-linear models. *Bulletin of the Swiss Association of Bulletin of the Swiss Association of Actuaries*, (1) :29–48, 2005.

[43] P. Deprez, P. V. Shevchenko, and M. V. Wüthrich. Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7(2) :337–352, 2017.

[44] D. Dimitrova, S. Haberman, and V. Kaishev. Dependent competing risks : Cause elimination and its impact on survival. *Insurance : Mathematics and Economics*, 53 :464–477, 2013.

[45] C. Do and S. Batzoglou. What is the expectation maximization algorithm ? *Nature biotechnology*, 26 :897–9, 2008.

[46] A. L. Dolinar, J. Sambt, and S. Cerne. Clustering eu countries by causes of death. *Population Research and Policy Review*, 38 :157—-172, 2019.

[47] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*. Oxford University Press, 2001.

[48] C. Dutang. Some explanations about the IWLS algorithm to fit generalized linear models. 2017.

[49] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal. Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, 35(3) :279–300, 2003.

[50] P. Eilers and B. Marx. Flexible Smoothing with B-splines and Penalties. *Statistical Science*, 11(2) :89–102, 1996.

[51] R. Elandt-Johnson. Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. *Scandinavian Actuarial Journal*, 1976(1) :37–51, 1976.

[52] D. Evans, N.A. Gillespie, and N.G. Martin. Biometrical genetics. *Biological Psychology*, 61(1) :33 – 51, 2002.

[53] D. Faraggi and R. Simon. A neural network model for survival data. *Statistics in Medicine*, 14(1) :73–82, 1995.

[54] R. A. Fisher. On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85(1) :87–94, 1922.

[55] F. Flici. Closing-out the Algerian life tables : for more accuracy and adequacy at old-ages. 2016.

[56] Centers for Disease Control and Prevention. Available at https://www.cdc.gov/ (data downloaded on 2019-03-02).

[57] K. J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. McGaughey, M. A. Pletcher, A. E. Smith, K. Tang, C.W. Yuan, J. C. Brown, J. Friedman, J. He, K. R. Heuton, M. Holmberg, D. J. Patel, P. Reidy, A. Carter, K. Cercy, A. Chapin, D. Douwes-Schultz, T. Frank, F. Goettsch, P. Y. Liu, V. Nandakumar, M. B. Reitsma, V. Reuter, N. Sadat, R. J. D. Sorensen, V. Srinivasan, R. L. Updike, H. York, A. D. Lopez, R. Lozano, S. S. Lim, A. H. Mokdad, S. E. Vollset, and C. J. L. Murray. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death : reference and alternative scenarios for 2016–40 for 195 countries and territories. *The Lancet*, 392(10159) :2052 – 2090, 2018.

[58] M.C. Fung, G. Peters, and P. Shevchenko. A unified approach to mortality modelling using state-space framework :characterisation, identification, estimation and forecasting. *Annals of Actuarial Science*, 11(2) :343–389, 2017.

[59] L. Gavrilov and N. Gavrilova. Mortality measurement at advanced ages : A study of the social security administration death master file. *North American actuarial journal*, 15 :432–447, 2011.

[60] N. Gavrilova, L. Gavrilov, and V. Krut'ko. Mortality trajectories at exceptionally high ages : A study of supercentenarians. *Living to 100 monograph*, 2017, 2017.

[61] C. Gessert, B. Elliott, and I. Haller. Dying of old age : An examination of death certificates of minnesota centenarians. *Journal of the American Geriatrics Society*, 50 :1561–5, 2002.

[62] F. Girosi and G. King. Understanding the Lee-Carter mortality forecasting method. 2007.

[63] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 4th edition, 2013.

[64] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115 :513–583, 1825.

[65] L. Gordon and R. Olshen. Tree-structured survival analysis. *Cancer treatment reports*, 69(10) :1065–9, 1985.

[66] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140 :107 – 113, 1993.

[67] J. Graunt. *Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality*, pages 11–20. Springer Berlin Heidelberg, Berlin, Heidelberg, 1977.

[68] S. Gravina. Epigenetic factors in aging and longevity. *Pflügers Archiv : European journal of physiology*, 459 :247–58, 2009.

[69] E. Halley. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw ; with an attempt to ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal Society of London*, 17(196) :596–610, 1693.

[70] L. Heligman and J. Pollard. The age pattern of mortality. *Journal of the Institute of Actuaries*, 107, 1980.

[71] HMD. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 2019-03-02).

[72] B. Honore and A. Lleras-Muney. Bounds in competing risk models and the war on cancer. *Econometrica*, 74 :1675–1698, 2006.

[73] R. Hyndman, R. Ahmed, G. Athanasopoulos, and H.L. Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9) :2579–2589, 2011.

[74] S. M. Jazwinski. Longevity, Genes, and Aging. *Science*, 273(5271) :54–59, 1996.

[75] V. Kaishev, D. Dimitrova, and S. Haberman. Modelling the joint distribution of competing risks survival times using copula functions. *Insurance : Mathematics and Economics*, 41 :339–361, 2007.

[76] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D) :35–45, 1960.

[77] V. Kannisto, J. Lauritsen, A.R. Thatcher, and J. Vaupel. Reductions in mortality at advanced ages : Several decades of evidence from 27 countries. *Population and Development Review*, 20(4) :793–810, 1994.

[78] A. Kantas, N.and Doucet, S. Singh, J. Maciejowski, and N. Chopin. On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3) :328–351, 2015.

[79] E.L. Kaplan and P. Meier. Nonparametrics estimates for incomplete observations. *Journal of the American Statistical Association*, 53 :457–480, 1958.

[80] O. Kempthorne. *An Introduction to Genetic Statistics*. Iowa State University Press ; 1st Edition (January 1, 1969), 1958.

[81] T. Kirkwood. Deciphering death : A commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370, 2015.

[82] S. Kjaergaard, Y. Ergemen, M. Kallestrup-Lamb, J. Oeppen, and R. Lindahl-Jacobsen. Forecasting causes of death by using compositional data analysis : the case of cancer deaths. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 68, 2019.

[83] A. Kogure and Y. Kurachi. A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance : Mathematics and Economics*, 46 :162–172, 2010.

[84] T. Launay, A. Philippe, and S. Lamarche. On particle filters applied to electricity load forecasting. *Journal de la Société Française de Statistique*, 154, 2013.

[85] R. Lee. The lee-carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4 :80–91, 2000.

[86] F. Leisch and B. Grün. Extending standard cluster algorithms to allow for group constraints. 2020.

[87] H. Li, H. Li, Y. Lu, and A. Panagiotelis. A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance : Mathematics and Economics*, 86 :122–133, 2019.

[88] H. Li and Y. Lu. Modeling cause-of-death mortality using hierarchical Archimedean copula. *Scandinavian Actuarial Journal*, 2019(3) :247–272, 2019.

[89] N. Li and R. Lee. Coherent mortality forecasts for a group of population : An extension of the Lee–Carter method. *Demography*, 42 :575–94, 2005.

[90] N. Li, R. Lee, and P. Gerland. Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6) :2037–2051, 2013.

[91] Y. Liu and J. Siu. It's all in the hidden states : A longevity hedging strategy with an explicit measure of population basis risk. *Insurance : Mathematics and Economics*, 70, 07 2016.

[92] Y. Liu and J. Siu. The locally linear Carins-Blake-Dowd Model : a note on Delta-Nuga hedging of longevity risk. *ASTIN Bulletin*, 47 :1–73, 2016.

[93] P. Lyu, A. De Waegenaere, and B. Melenberg. A multi-population approach to forecasting all-cause mortality using cause-of-death mortality data. *North American Actuarial Journal*, 0(0) :1–36, 2020.

[94] J. MacQueen. Proceedings of the fifth berkeley symposium on mathematical statistics and probability, volume 1 : Statistics. pages 281–297. University of California Press, 1967.

[95] W. Makeham. On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6) :301–310, 1860.

[96] P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989.

[97] R. McNown and A. Rogers. Forecasting mortality : A parameterized time series approach. *Demography*, 26 :645–60, 1989.

[98] R. McNown and A. Rogers. Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8 :413–432, 1992.

[99] F. Meslé, J. Vallin, and Z. Andreyev. Mortality in europe : The divergence between East and West. *Population (English Edition, 2002-)*, 57 :157, 2002.

[100] A. Michaelson and J. Mulholland. Strategy for increasing the global capacity for longevity risk transfer : Developing transactions that attract capital markets investors. *The Journal of Alternative Investments*, 17 :18–27, 2014.

[101] T.K. Moon. The expectation-maximization algorithm. *Signal Processing Magazine, IEEE*, 13 :47 – 60, 1996.

[102] M. Moreno-Betancur, H. Sadaoui, C. Piffaretti, and G. Rey. Survival analysis with multiple causes of death extending the competing risks model. *Epidemiology (Cambridge, Mass.)*, 28, 2016.

[103] W. Nelson. Hazard plotting for incomplete failure data. *Journal of Quality Technology*, 1 :27–52, 1969.

[104] B. Nielsen and J. Nielsen. Identification and Forecasting in the Lee-Carter Model. *SSRN Electronic Journal*, 2010.

[105] OECD. *Health at a Glance 2019*. 2019.

[106] J. Oeppen. Coherent forecasting of multiple-decrement life tables : a test using Japanese cause of death data. 2008.

[107] J. Paparrizos and L. Gravano. k-shape : Efficient and accurate clustering of time series. 2015.

[108] C. Pedroza. A Bayesian forecasting model : Predicting U.S. male mortality. *Biostatistics (Oxford, England)*, 7 :530–50, 2006.

[109] I. Permanyer and N. Scholl. Global trends in lifespan inequality : 1950-2015. *PLOS ONE*, 14 :e0215742, 2019.

[110] J.H. Pollard. Projection of age-specific mortality rates. *Population Bulletin of the United Nations*, (21/22) :55–69, 1987.

[111] J.H. Pollard. Cause of death and expectation of life : some international comparisons. pages 269–291, 1990.

[112] R. Prentice, J. Kalbfleisch, A.V. Peterson, N. Flournoy, V. Farewell, and N. Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34 4 :541–54, 1978.

[113] H. Putter, M. Fiocco, and R. Geskus. Tutorial in biostatistics : competing risks and multi-state models. *Statistics in medicine*, 26 11 :2389–430, 2007.

[114] Lee R. and L. Carter. Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*, 87(419) :659–671, 1992.

[115] C. Redondo Lourés and A.J.G. Cairns. Cause of death specific cohort effect in U.S. mortality. 2021.

[116] A.E. Renshaw and S. Haberman. A cohort-based extension to the lee-carter model for mortality reduction factors. *Insurance : Mathematics and Economics*, 38 :556–570, 2006.

[117] R. Richman and M. Wuthrich. A neural network extension of the lee-carter model to multiple populations. *SSRN Electronic Journal*, 2018.

[118] F. Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6) :386–408, 1958.

[119] M. Roser, E. Ortiz-Ospina, and H. Ritchie. Life expectancy. *Our World in Data*, 2013.

[120] R. Rusconi. National annuity markets : Features and implications. *OECD, Directorate for Financial and Enterprise Affairs, OECD Working Papers on Insurance and Private Pensions*, 2008.

[121] J. Saboia. Modeling and forecasting populations by time series : The swedish case. *Demography*, 11 :483–492, 1974.

[122] A. Schwarzkopf, R. Tersine, and J. Morris. Top-down versus bottom-up forecasting strategies. *International Journal of Production Research*, 26(11) :1833–1843, 1988.

[123] T. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47 :39–49, 2011.

[124] A. Thatcher, J. Vaupel, and V. Kannisto. *The force of mortality at ages 80 to 120*. Odense University Press, Odense, Denmark, 1998.

[125] C.C.L. Tsai and E.S. Cheng. Incorporating statistical clustering methods into mortality models to improve forecasting performances. *Insurance : Mathematics and Economics*, 99 :42–62, 2021.

[126] A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1) :20–22, 1975.

[127] J. Vallin, E. Andreev, F. Meslé, and V. Shkolnikov. Geographical diversity of cause-of-death patterns and trends in russia. *Demographic Research*, 12 :323–380, 2005.

[128] J. Vaupel, K. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3) :439–454, 1979.

[129] J. Vaupel, Z. Zhang, and A. van Raalte. Life expectancy and disparity : An international comparison of life table data. *BMJ open*, 1 :e000128, 2011.

[130] A. Villegas and S. Haberman. On the modeling and forecasting of socioeconomic mortality differentials : An application to deprivation and mortality in England. *North American Actuarial Journal*, 18(1) :168 – 193, 2014.

[131] M. Wagner. Cointegration analysis with state space models. *AStA Advances in Statistical Analysis*, 94 :273–305, 2010.

[132] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl. Constrained K-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, page 577–584, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 2001.

[133] M. Wall, J. Huang, J. Oswald, and D. McCullen. Factors associated with reporting multiple causes of death. *BMC medical research methodology*, 5 :4, 2005.

[134] S. Wickramasuriya, G. Athanasopoulos, and R. Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114 :1–45, 2018.

[135] J. Wilmoth. Are mortality projections always more pessimistic when disaggregated by cause of death ? *Mathematical Population Studies*, 5(4) :293–319, 1995.

[136] B. Xie, W. Pan, and X. Shen. Penalized model-based clustering with cluster-specific diagonal covariance matrices and grouped variables. *Electronic Journal of Statistics*, 2 :168–212, 2008.

[137] S. Yang, J. Yue, and H.C. Huang. Modeling longevity risks using a principal component approach : A comparison with existing stochastic mortality models. *Insurance : Mathematics and Economics*, 46 :254–270, 2010.

[138] B. Zethelius, L. Berglund, and J. Sundström. Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *Journal of Vascular Surgery*, 48 :500, 2008.

[139] M. Zheng and J. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82(1) :127–138, 1995.