

New Approaches in Mortality Modelling and Forecasting

*A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy*

by

Ugofilippo Basellini

Interdisciplinary
Centre on Population Dynamics | **SDU** 

University of Southern Denmark
Faculty of Health Sciences
Institute of Public Health
Interdisciplinary Centre on Population Dynamics

Odense, Denmark
February 2020

Academic Advisors

Dr. **Carlo Giovanni Camarda**, Ph.D.
Mortality, Health and Epidemiology Unit
Institut national d'études démographiques

Associate Professor **James Oeppen**, M.A.
Interdisciplinary Centre on Population Dynamics
Faculty of Business and Social Sciences
University of Southern Denmark

Associate Professor **Vladimir Canudas-Romo**, Ph.D.
School of Demography
Australian National University, Canberra, Australia

Professor **Annette Baudisch**, Ph.D.
Interdisciplinary Centre on Population Dynamics
Faculty of Business and Social Sciences
University of Southern Denmark

Assessment Committee

Associate Professor **Fanny Janssen**, Ph.D.
Department of Demography
Faculty of Spatial Sciences
University of Groningen

Dr. **Iain D. Currie**, Ph.D.
Department of Actuarial Mathematics and Statistics
School of Mathematical and Computer Sciences
Heriot-Watt University

Professor **Mauro Laudicella**, Ph.D. (chair)
Department of Public Health
Faculty of Health Sciences
University of Southern Denmark, Denmark

Contents

Acknowledgements	vii
English summary	ix
Danish summary	xi
Publications	xiii
List of abbreviations	xv
1 Introduction	1
1.1 Preliminaries	2
1.2 Mortality modelling	8
1.3 Mortality forecasting	12
1.4 Aims of the thesis	15
1.5 Ch. 2: location-scale models in demography	15
1.6 Ch. 3: modelling and forecasting adult distributions	16
1.7 Ch. 4: modelling and forecasting age-at-death distributions	18
1.8 Ch. 5: forecasting cohort age-at-death distributions	19
1.9 Ch. 6: smoothing, decomposing and forecasting mortality rates	20
1.10 Discussion & Outlook	21
1.11 Reproducibility	23
2 Location-Scale Models in Demography: A Useful Re-parameterization of Mortality Models	25
2.1 Introduction	28
2.1.1 Parametric mortality models	28
2.1.2 Short history of parametric mortality models	29
2.1.3 Aims	30
2.2 Methods	30
2.2.1 Life-table functions	30
2.2.2 The location-scale family of mortality models	31
2.2.3 Gompertz and the location-scale family	33
2.2.4 Other parametric mortality models	34
2.2.5 Data and estimation procedure	36
2.3 Illustrations	36
2.3.1 Parameter interpretation: shifting and compression	36
2.3.2 Correlation and bias of parameter estimates	40

2.4	Discussion and conclusion	44
2.5	Appendix	45
2.5.1	Weibull and the log–location–scale family	45
2.5.2	Location–scale functional form of twelve parametric models of mortality	46
2.5.3	Derivation of the best fitting parametric model	47
2.5.4	Sec. 2.3.1: additional results	47
2.5.5	Decomposition of mortality changes into location and scale effects	49
3	Modelling and forecasting adult age-at-death distributions	51
3.1	Introduction	54
3.2	Methods	57
3.2.1	The Segmented Transformation Age-at-death Distributions Model	57
3.2.2	The standard distribution	59
3.2.3	Data and estimation of the standard distribution	60
3.2.4	Estimation and forecast of the STAD parameters	62
3.3	Results	63
3.3.1	Observed versus fitted data	63
3.3.2	Out-of-sample validation	64
3.3.3	Forecasts to 2040	67
3.4	Discussion and conclusion	70
3.5	Appendix	74
3.5.1	Deviance, Effective Dimension and BIC	74
3.5.2	Sec. 3.3.1: additional results	74
4	A Three-component Approach to Model and Forecast Age-at-death Distributions	77
4.1	Introduction	80
4.2	Methods	82
4.2.1	Mortality functions	82
4.2.2	Data and mortality decomposition	83
4.2.3	Modelling component-specific distributions	85
4.2.4	Estimating and forecasting the 3C-STAD parameters	88
4.3	Results	90
4.3.1	Out-of-sample validation	90
4.3.2	Forecast to 2050	92
4.4	Discussion	94
5	An age-at-death distribution approach to forecast cohort mortality	99
5.1	Introduction	102
5.2	Background	103
5.3	Methods & Data	105
5.3.1	The C-STAD model	105
5.3.2	Data	107
5.3.3	The standard distribution	108
5.3.4	Estimation and forecast of the C-STAD parameters	110
5.4	Results	112

5.4.1	Out-of-sample validation of the C-STAD model	112
5.4.2	Mortality developments for Swedish and Danish females, cohorts 1835–1970	114
5.5	Discussion	116
5.6	Appendix	119
5.6.1	Deviance residuals and bootstrap death counts	119
5.6.2	Section 5.3: additional results	119
6	Smoothing, decomposing and forecasting mortality rates	125
6.1	Background	128
6.2	Data and methods	130
6.2.1	The Lee-Carter model	131
6.2.2	The Three-Component smooth Lee-Carter model	134
6.3	Results	139
6.4	Discussion	144
6.5	Conclusions	147
Bibliography		149

Acknowledgements

This dissertation has been developed and written during my three-year appointment as a joint Ph.D. student at the Institut national d'études démographiques (INED) and at the Interdisciplinary Centre on Population Dynamics (CPop) in the Department of Public Health of the University of Southern Denmark.

Writing a Ph.D. thesis is a significant challenge that I surely would have not completed if it wasn't for the extraordinary people that have accompanied me along the journey. First of all, I want to thank all the friends that I have had the luck of meeting during my life. Friends from the Junior Tennis Club, Liceo Scientifico Severi, Università Bocconi, CPop (and former MaxO), MaxNetAging Research School, the European Doctoral School of Demography and INED: I have been blessed to meet and spend part of my life with you. Among so many, a special thanks is in order to my best friends, Alessandro and Matteo: you have taught me the real definition of friendship, and I am grateful for having you in my life.

Moreover, I wish to thank my mentors and supervisors that have shared their wisdom and knowledge with me. Your teachings, academics and not, have made me the man I am today, and I will always remember what you have done for me. Finding the right path after the graduate studies can be very hard, and I believe there exists no one who can be so genuinely kind, supportive and always ready to help as Marco Bonetti. Thank you for being there for me, with your unlimited enthusiasm, so many times. The first work experiences can also be significant challenges in someone's life, unless your team director is Tim Crayford: I have never felt more valued than during the years spent together at Just Retirement. Your exceptionally high opinion of me has pushed me to work harder, and you have taught me how to work with others. Next, Ph.D. supervisors are sometimes, and unfortunately, known for complicating a doctoral student's path. Giancarlo Camarda, you have definitely gone too much in the other direction: your constant guidance, help, support and kindness made my Ph.D. such an enjoyable and unforgettable journey. Since I am your first doctoral student, I can tell you something you haven't heard before: you couldn't have been a better supervisor than this. Last, but not least, I extend my deepest gratitude to all the other mentors that have helped me during these last few years: Vladimir Canudas-Romo, Jim Oeppen, Annette Baudish and Jim Vaupel.

Finally, the most special thanks are reserved to my family. To my parents Elena and Aldo: I will be forever thankful to you for raising me and Carlo in the most exceptional manner; for always being there and making us feeling loved in each moment of our lives. Thank you especially for being supportive of each of my own choices, even those that put some physical distance between us. To my brother Carlo: the countless hours spent playing together during our childhood are my absolute favourite memories. You are a truly special brother and surely one of my best friends. To my grandparents Michela and Gustavo: thank you for all your love and support during my life, you have done so much for our family. Lastly, a very special thanks to Marília: your everyday kindness and love stimulate me to be a better man. I am extremely lucky to have you at my side: with you, life is so much more beautiful.

English summary

Mortality modelling and forecasting are deeply rooted in demographic and actuarial sciences. Models to describe mortality patterns over age and time have long been used and developed since John Graunt ([1662](#)) introduced one of the first models of mortality, the life table. Forecasts of mortality have also been produced for many years: the first examples trace back to the beginning of the twentieth century, when English actuaries started to measure the financial burden of unanticipated longevity improvements on insurance and pension providers' reserves.

Today, the study of human mortality still occupies a central role in demographic and actuarial analyses. Most of the attention received by this area of research has been stimulated by two pressing challenges faced by modern societies: population ageing and longevity risk. According to the latest World Population Prospects, virtually every country of the world is experiencing growth in the number and proportion of older persons, resulting from continuous mortality and fertility declines ([United Nations, 2019](#)). Furthermore, the demographic transition has been impacting both public and private pension systems, whose retirement liabilities lie between \$60 and \$80 trillions in developed economies due to unexpected mortality improvements ([Michaelson and Mulholland, 2014](#)). Funding public policies and retirement products for the elderly becomes increasingly difficult as working-age populations shrink and dependency ratios increase worldwide.

The enormous size of unexpected public and private retirement liabilities is the result of overly conservative forecasts of mortality during most recent decades. Despite the great advances in the field of mortality forecasting, including the shift from deterministic to stochastic approaches, currently and widely used methods have repeatedly failed to anticipate the sustained rate of mortality improvements observed in many low-mortality countries. The need for novel models that can predict longevity improvements more accurately than established methodologies is evident and timely. Therefore, this dissertation aims to bring new insights to the analysis and forecasting of human mortality by introducing novel statistical methods that offer different perspectives on mortality developments.

This dissertation comprises six chapters, five of which are studies that have been devised to address this goal. Each study takes the form of a research manuscript, which has been published or submitted to scientific journals; furthermore, routines for reproducing the results presented in

the thesis have been made publicly available. The first chapter introduces the basic notions and measures employed in the study of human mortality, reviews the main contributions in the history of modelling and forecasting mortality, and provides a short overview of the five studies developed in the thesis. In Chapter 2, we illustrate a general framework for modelling adult mortality that reconciles the well-known laws of mortality into a single flexible family. Re-parameterizing mortality models in terms of the proposed location-scale family has two important advantages: the model's parameters have a direct demographic interpretation, and their estimation is more precise due to their lower correlation.

From the third to the fifth chapters, the attention is shifted from mortality rates to age-at-death distributions as an alternative, yet informative (and neglected), function for modelling and forecasting human mortality. Chapter 3 proposes a relational approach to model and forecast adult mortality by transforming the age-axis of a standard distribution of deaths. The proposed Segmented Transformation Age-at-death Distributions (STAD) model successfully captures mortality developments over age and time, and its forecasts are more accurate and optimistic than those obtained with the seminal Lee-Carter (LC) model ([Lee and Carter, 1992](#)) and its extensions. The STAD model is further employed and generalized in the following two chapters. In Chapter 4, the methodology is extended to the entire age-range. The age-pattern of mortality is first smoothly decomposed into three independent components that operate upon childhood, middle and old ages (as originally proposed by [Thiele, 1871](#)). The three components are then modelled and forecast with specialized versions of the STAD model. The resulting forecasts are shown to be more accurate and optimistic than those of traditional and well-established models. Chapter 5 presents a generalization and application of the STAD methodology for modelling and forecasting cohort mortality data. Models developed to forecast cohort data are very scarce in the literature, and our proposed approach allows us to precisely complete the mortality experience of partially observed cohorts. Finally, Chapter 6 proposes a new extension of the influential LC model that overcomes some of its known drawbacks. Working in a penalized composite link framework, we simultaneously smooth and decompose the mortality pattern into three independent components, which are modelled, estimated and forecast within an LC smooth framework. Fitted and forecast mortality profiles do not show the jaggedness typically displayed by the LC model; furthermore, mortality rates can vary more flexibly across age and time, as they result from a combination of three component-specific schedules of mortality changes.

Danish summary

Dødelighedsmodellering og -prognoser er dybt forankret i demografiske og aktuariske viden-skaber. Modeller til at beskrive dødelighedsmønstre over alder og tid er længe blevet brugt og udviklet, siden John Graunt ([1662](#)) introducerede en af de første modeller for dødelighed, dødelighedstavlen. Prognoser for dødelighed er blevet udarbejdet igennem mange år: De første eksempler kan spores tilbage til begyndelsen af det tyvende århundrede, hvor engelske aktuarer begyndte at måle den økonomiske byrde ved uventede forbedringer af levetiden på forsikrings- og pensionsudbyderes reserver.

I dag indtager analysen af dødelighed for mennesker en central rolle i demografiske og aktuariske analyser. Det meste af opmærksomheden, som dette forskningsområde får, kommer fra to presserende udfordringer det moderne samfund står over for: aldring af befolkningen og levetidsrisiko. I henhold til de seneste verdensbefolkningsprognoser vil praktisk talt alle lande i verden opleve vækst i antallet og andelen af ældre, som følge af kontinuerlige stigninger i den for-ventede levealder og forringet fertilitet ([United Nations, 2019](#)). Den demografiske transition har indflydelse på både offentlige og private pensionssystemer i udviklede økonomier, hvis pensions-forpligtelser, på grund af uventede forbedringer i dødeligheden, ligger mellem 60 og 80 billioner dollars ([Michaelson and Mulholland, 2014](#)). Finansiering af offentlige aktiviteter og pension-sprodukter til ældre bliver stadig vanskeligere, når befolkningsandelen i den erhvervsaktive alder falder og forsøgerkvoten stiger over hele verden.

De enorme uventede offentlige og private pensionsforpligtelser er resultatet af for konservative prognoser for dødelighed i de seneste årtier. På trods af mange store fremskridt inden for døde-lighedsprognoser, herunder skiftet fra en deterministisk til en stokastisk tilgang, har nuværende og bredt anvendte metoder gentagne gange undladt at forudse de konstante forbedringer i døde-ligheden, som er observeret i mange lande med lav dødelighed. Behovet for nye modeller, der kan forudsige forbedringer af levetiden mere nøjagtigt end etablerede metoder er indlysende og påkrævet. Derfor har denne afhandling til formål at bringe ny viden til analysen og forudsigelsen af menneskelig dødelighed ved at introducere nye statistiske metoder, der tilbyder forskellige perspektiver på dødelighedsudviklingen.

Denne afhandling består af seks kapitler, hvoraf fem er analyser, er udført til at nå dette mål. Hver analyse har form af et forskningsmanuskript, der er blevet offentliggjort eller sendt til

videnskabelige tidsskrifter; endvidere er metoderne til at replicere resultaterne, der er præsenteret i afhandlingen, gjort offentligt tilgængelige. Det første kapitel introducerer de grundlæggende forestillinger og mål, der er anvendt i studiet af menneskelig dødelighed, gennemgår de vigtigste bidrag i historien om modellering og forudsigelse af dødelighed og giver en kort oversigt over de fem studier, der er udført i afhandlingen. I kapitel 2 illustrerer vi en generel ramme for modellering af dødelighed for voksne, der forener velkendte dødelighedslove i en enkelt fleksibel familie. Re-parametrering af dødelighedsmodeller med hensyn til den foreslæde location-scale familie har to vigtige fordele: Modellens parametre har en direkte demografisk fortolkning, og deres estimering er mere præcis på grund af deres lavere korrelation.

Fra tredje til femte kapitel flyttes opmærksomheden fra dødelighed til fordelingen af dødsfald som en alternativ, men alligevel informativ (og forsømt) funktion til modellering og forudsigelse af menneskelig dødelighed. Kapitel 3 foreslår en relationel tilgang til at modellere og forudsige dødelighed blandt voksne ved at omdanne aldersaksen for en standardfordeling af dødsfald. Den foreslæde STAD-model modellerer succesfuldt dødelighedsudviklingen over alder og tid, og dens fremskrivninger er mere nøjagtige og optimistiske end dem, der opnås med den banebrydende Lee-Carter (LC) model ([Lee and Carter, 1992](#)) og dennes udvidelser. STAD-modellen anvendes også og generaliseres i de følgende to kapitler. I kapitel 4 udvides modellen til at inkluderer alle aldre. Aldersmønsteret på dødelighed er først dekomponeret i tre glatte og uafhængige komponenter, der opererer med barndom, middel-alder og alderdom (oprindeligt foreslæt af [Thiele, 1871](#)). De tre komponenter modelleres og fremskrives derefter med specialiserede versioner af STAD-modellen. Prognoser opnået med denne STAD-modellen viser sig at være mere nøjagtige og optimistiske end traditionelle og veletablerede modeller. Kapitel 5 præsenterer en generalisering og anvendelse af STAD-metodologien til modellering og prognoser af kohorte dødelighedsdata. Modeller, der er udviklet til at forudsige kohortedata, er meget få i litteraturen, og vores foreslæde fremgangsmåde giver os mulighed for præcist at fuldføre livsforløbet af delvist observerede kohorter. Endelig foreslår kapitel 6 en ny udvidelse af den indflydelsesrige LC-model, der afhjælper nogle af dens kendte problemer. Med Penalized Composite Link model nedbryder vi dødeligheds mønsteret i tre uafhængige komponenter, der er modelleret, estimert og fremskrevet inden for en glat LC-struktur. Estimerede og fremskrevne dødelighedsprofiler viser ikke den skarphed, der typisk opstår af LC-modellen; endvidere kan dødelighedsændringer variere mere fleksibelt over alder og tid, da de er resultatet af en kombination af tre komponentspecifikke skemaer for mortalitetsforbedringer.

Publications

Manuscripts included in this dissertation

Paper I

Basellini, U., Canudas-Romo, V. and Lenart, A. (2019). Location-Scale Models in Demography: A Useful Re-parameterization of Mortality Models. *European Journal of Population*, **35**(4), 645–673. DOI: [10.1007/s10680-018-9497-x](https://doi.org/10.1007/s10680-018-9497-x)

Paper II

Basellini, U. and Camarda, C.G. (2019). Modelling and forecasting adult age-at-death distributions. *Population Studies*, **73**(1), 119–138. DOI: [10.1080/00324728.2018.1545918](https://doi.org/10.1080/00324728.2018.1545918)

Paper III

Basellini, U. and Camarda, C.G. A Three-component Approach to Model and Forecast Age-at-death Distributions. In Mazzuco, S., and Keilman, N. (eds.), *Developments in Demographic Forecasting*, Springer. Forthcoming.

Paper IV

Basellini, U., Kjærgaard, S., and Camarda, C.G. An age-at-death distribution approach to forecast cohort mortality. Manuscript under review.

Paper V

Camarda, C.G., and Basellini, U. Smoothing, decomposing and forecasting mortality rates. Manuscript under review.

Other co-authored works during the PhD, not included in the dissertation

Paper VI

Bergeron-Boucher, M.P., Kjærgaard, S., Pascariu, M., Aburto, J.M., Alvarez, J., Basellini, U., Rizzi, S. and Vaupel, J.W. Alternative forecasts of Danish life expectancy. In Mazzuco, S., and Keilman, N. (eds.), *Developments in Demographic Forecasting*, Springer. Forthcoming.

Paper VII

Bonetti, M., Gigliarano, C., and Basellini, U. The Gini concentration index for the study of survival. In Mukhopadhyay, N., and Sengupta, P.P. (eds.), *Gini: An Edited Volume on the Occasion of Gini Centenary*, New Delhi: Serials. Forthcoming.

Paper VIII

Sutter, A., Barton, S., Sharma, M.D., Basellini, U., Hosken, D., and Archer, C.R. (2018). Senescent declines in elite tennis players are similar across the sexes. *Behavioral Ecology*, **29**(6), 1351–1358. DOI: [10.1093/beheco/ary112](https://doi.org/10.1093/beheco/ary112)

Paper IX

Archer, C.R., Basellini, U., Hunt, J., Simpson, S.J., Lee, K.P., and Baudish, A. (2018). Diet has independent effects on the pace and shape of aging in *Drosophila melanogaster*. *Biogerontology*, **19**(1), 1–12. DOI: [10.1007/s10522-017-9729-1](https://doi.org/10.1007/s10522-017-9729-1)

Paper X

Riffe, T., Nepomuceno, M., and Basellini, U. Mortality modeling. Manuscript under review as invited chapter for the *Encyclopedia of Gerontology and Population Aging* (eds. D. Gu and M.E. Dupre), section Mortality (P. Gerland).

Paper XI

Aburto, J.M., Villavicencio, F., Basellini, U., Kjærgaard, S., and Vaupel, J.W. Dynamics of life expectancy and lifespan equality. Manuscript under review.

List of abbreviations

3C-sLC	Three-Component smooth Lee-Carter model
3C-STAD	Three-Component Segmented Transformation Age-at-death Distributions model
AB	Relative absolute bias
AFT	Accelerated failure time model
ARIMA	Autoregressive integrated moving average model
BIC	Bayesian Information Criterion
BMS	Booth-Maindonald-Smith model
BDV	Brouhns-Denuit-Vermunt model
C-STAD	Cohort Segmented Transformation Age-at-death Distributions model
CLM	Composite Link Model
CODA	Compositional Data Analysis
DSS	Dawid-Sebastiani score
ECP	Empirical coverage probability
ED	Effective dimension
GG	Gamma-Gompertz model
GLM	Generalized Linear Model
IWLS	Iterative re-Weighted Least Squares algorithm
LC	Lee-Carter model
LLS	Log–location–scale family of models
LM	Lee-Miller model
LS	Location–scale family of models
HMD	Human Mortality Database
HU	Hyndman-Ullah model
HUrob	Robust Hyndman-Ullah model
HUw	Weighted Hyndman-Ullah model
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MinGEV	Minimal Generalized Extreme-Value model
MLE	Maximum likelihood estimator
OLS	Ordinary least-squares
PI	Prediction intervals
RMSE	Root mean square error
RWD	Random walk with drift model
SSE	Sum of Smooth Exponentials model
STAD	Segmented Transformation Age-at-death Distributions model
SVD	Singular-value decomposition
VAR	Vector autoregressive model

Chapter 1

Introduction

“Fra la precisa descrizione di un fatto del passato e l’accurata previsione del futuro non c’è alcuna relazione, ma è ovvio che l’una tende a rendere più credibile l’altra. [There is no connection between the precise description of past events and the accurate prediction of the future, but obviously the one lends credibility to the other].”

Tiziano Terzani (1995), *Un indovino mi disse*

Mortality modelling and forecasting have a long history in demographic and actuarial analysis. The origin of mortality modelling as a research field, and of demography as a scientific discipline, is generally identified with the figure of John Graunt. In 1662, Graunt presented his *Natural and Political Observations Made upon the Bills of Mortality* to the Royal Society of London, the treatise that is considered to have laid the foundations of modern demography. Since then, many efforts have been devoted to the search for an encompassing model that could describe the age-pattern of human mortality accurately and parsimoniously.

Mortality forecasting is a relatively more recent endeavour. Although forecasts of mortality can be traced back at least to the beginning of the twentieth century, research on mortality forecasting flourished during the last three decades. A new ensemble of mortality models, characterised by a common reliance on sound statistical and stochastic methods, has been proposed with the explicit goal of forecasting longevity improvements.

Life expectancy at birth, which measures the average number of years that a newborn will live (if age-specific mortality rates remain constant over time, Preston et al., 2001), has been increasing worldwide since around 1800 (Riley, 2001). The rise of human longevity is undoubtedly one of the most remarkable achievements of modern societies (Oeppen and Vaupel, 2002); however, improvements in survival as well as declines in fertility have generated a global process of population ageing. According to the United Nations, populations in virtually all countries and areas of the world are growing older, with persons over age 65 being the fastest-growing age group (United Nations, 2019). As a result, public and private sectors face increasing challenges to provide adequate pension products and elderly health care.

The difficulty of funding retirement products has become so pervasive and pressing in our ageing societies that the financial industry has coined the term “longevity risk” when referring to this issue. Longevity risk is defined as the risk that retirees could live longer (on average) than expected. From a financial perspective, this is a risk whenever an institution provides payments depending on how long individuals are going to live (Blake et al., 2014): if many (more than average) people live longer than their estimated life expectancy, the institution will face greater monetary outflows than its planned reserves. Small differences between the realized and previously projected lifespans of pensioners become highly magnified in the financial market: the potential size of the global longevity risk market has recently been estimated to lie between

\$60 and \$80 trillion ([Michaelson and Mulholland, 2014](#)). As such, the need for innovative models that can predict the future course of mortality more accurately than previous approaches is evident and timely; according to [Janssen \(2018\)](#) and [Bengtsson et al. \(2019\)](#), this need is greater than ever before.

The main goal of this dissertation is to propose innovative statistical methods that can provide novel insights into the analysis and forecasting of human mortality. Demographic and actuarial studies of mortality typically focus on whole populations, or large sub-populations, where only a handful of covariates are known (generally age, time and sex). Similarly, the methods proposed in this dissertation are primarily developed for studying mortality at the aggregate level.

This introductory Chapter is organised as follows. Section 1.1 introduces the structure and the source of data that we use throughout this dissertation, as well as the main functions and measures typically employed in mortality studies. Sections 1.2 and 1.3 review the main historical developments and contributions to the fields of mortality modelling and forecasting. The specific aims of the thesis are presented in Section 1.4. From Section 1.5 to 1.9, we provide a short introduction to the Chapters of this dissertation. Section 1.10 discusses the methodologies presented in this thesis and their related results, and it outlines some possible directions for future work. Finally, Section 1.11 provides the links and information for using the routines and reproducing the results of this dissertation.

1.1 Preliminaries

Mortality data

An understanding of the data that are typically employed in mortality studies is a necessary first step to performing research on mortality modelling and forecasting. This study starts by describing the general structure of mortality data, as well as the specific source of data that we use throughout this dissertation.

The graphical illustration of mortality data is a convenient way to understand its structure. The principal tool for visualizing this type of data is the Lexis diagram ([Lexis, 1875](#)), a graph that depicts the life histories of different individuals through so-called life-lines. A Lexis diagram displays the three main dimensions along which mortality is primarily measured, namely age, period (i.e. calendar year) and (birth) cohort. These dimensions are related by the well-known identity $cohort = period - age$. In a Lexis diagram, age and period are represented on the y - and x -axis, respectively, while cohort corresponds to left-to-right diagonals.

The left panel of Figure 1.1 shows a simplified Lexis diagram for a population closed to migration, in which several life-lines are shown for two different birth cohorts. All life-lines start from the x -axis, and they move along diagonals with slope equal to one as time passes and individuals age. The lines stop once the individuals experience death, which is represented by a red point in the graph. Censoring and truncation are not considered in this example.

When the research question concerns the analysis of mortality patterns and developments of rather large populations, exact life-lines for millions of individuals are rarely known. It is therefore customary and convenient to group together individual life-lines within the same lower

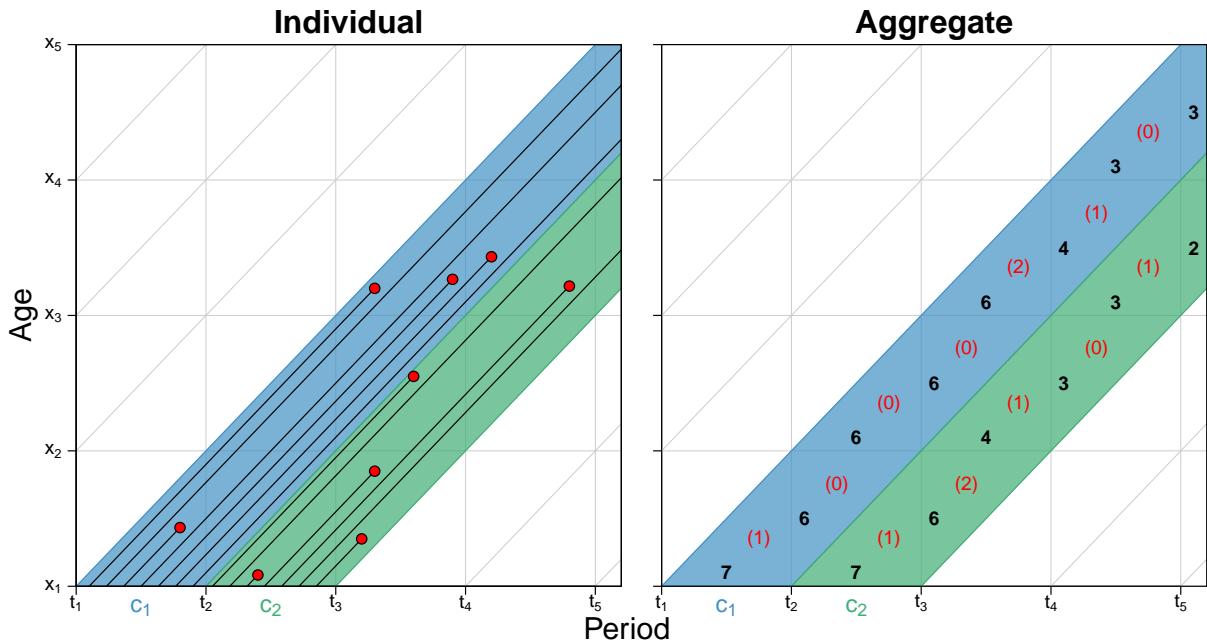


Figure 1.1: Two simplified Lexis diagrams that illustrate individual life-lines (left panel) and aggregate mortality data (right panel) for the cohorts c_1 and c_2 (blue and green areas, respectively). In the left panel, red points represent the death of the individual. The age and period dimensions correspond to the y - and x -axis, respectively.

Source: Own elaborations, inspired by Figure 1.1 in Camarda (2008).

and upper triangles of the diagram. This allows one to efficiently summarize death and population data at the aggregate level by age, period and cohort. The right panel of Figure 1.1 shows a Lexis diagram in which aggregate mortality data have been computed from the left panel.

Clearly, the availability of aggregate mortality data is paramount for the purpose of this dissertation. Traditional sources of such datasets are censuses, official vital statistics and population registries, which are routinely collected in several developed countries. In the year 2000, the University of California at Berkeley (United States) and the Max Planck Institute for Demographic Research (Rostock, Germany) initiated a far-sighted collaborative project called the [Human Mortality Database](#) (HMD, 2019), with the aim of providing researchers, policy makers and others with a unique source of accurate, homogenised and comparable national data on birth, death and population counts for several countries. Today, the HMD provides free access to detailed, consistent and high quality mortality data for 41 different countries or areas via the Internet ([Barbieri et al., 2015](#); [Wilmoth et al., 2019](#)). The database has gained increasing attention during recent years, becoming one of the main sources of data for mortality studies: up to August 2019, almost 3500 journal articles have been published relying on the HMD¹.

Data in this dissertation are thus uniquely retrieved from the HMD. For all the manuscripts in this thesis except Chapter 5, we perform mortality analysis on the traditional age-period perspective depicted in Figure 1.1. As such, we derive from the HMD observed death counts $y_{x,t}$ and central exposures to the risk of death $e_{x,t}$ at age x and time t . Let bold capital and small letters denote matrices and vectors, respectively. Data are arranged into two matrices $\mathbf{Y} = (y_{x,t})$ and $\mathbf{E} = (e_{x,t})$, each of dimensions $m \times n$, where: (i) rows are classified by m single ages at death,

¹source: <https://www.mortality.org/Public/HMD-Publist.pdf>

$\mathbf{x}' = [x_1, \dots, x_m]$, and (ii) columns are classified by n single years, $\mathbf{t}' = [t_1, \dots, t_n]$. Visually, the matrices can be shown as follows:

$$\mathbf{Y} = \begin{bmatrix} y_{x_1,t_1} & y_{x_1,t_2} & \dots & y_{x_1,t_n} \\ y_{x_2,t_1} & y_{x_2,t_2} & \dots & y_{x_2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{x_m,t_1} & y_{x_m,t_2} & \dots & y_{x_m,t_n} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} e_{x_1,t_1} & e_{x_1,t_2} & \dots & e_{x_1,t_n} \\ e_{x_2,t_1} & e_{x_2,t_2} & \dots & e_{x_2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{x_m,t_1} & e_{x_m,t_2} & \dots & e_{x_m,t_n} \end{bmatrix}.$$

In Chapter 5, we shift the focus of the analysis to the age-cohort perspective. The structure of the available mortality data is different and characterized by unobserved data for some elements of \mathbf{Y} and \mathbf{E} . A full description of the structure of age-cohort mortality data is provided in Section 5.2 of Chapter 5.

The matrices \mathbf{Y} and \mathbf{E} allow us to compute one of the principal measures of mortality, namely central death rates, which are often denoted age-specific mortality rates. Specifically, the rate $m_{x,t}$ at age x and time t is computed by dividing the corresponding death counts by exposures, i.e. $m_{x,t} = y_{x,t}/e_{x,t}$. Death rates for all ages and years can then be arranged in the matrix $\mathbf{M} = (m_{x,t})$. As a practical illustration, Figure 1.2 displays the available mortality data for Swedish females in 2017.

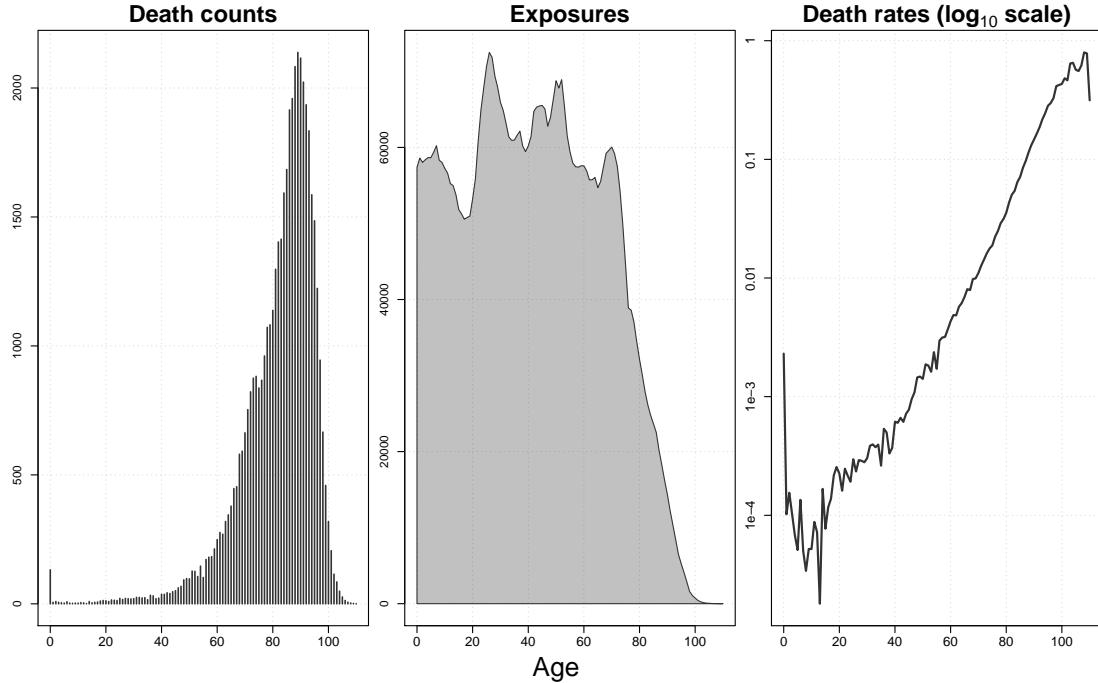


Figure 1.2: Observed death counts, exposures to the risk of death and central death rates (on \log_{10} scale) for Swedish females at ages 0–110+ in 2017.

Source: [Human Mortality Database \(2019\)](#).

It is worth mentioning here that, in all manuscripts of this dissertation, we always compute death rates from the matrices \mathbf{Y} and \mathbf{E} instead of using the life-table death rates provided by the HMD. The former are indeed observed rates, whereas the latter are smoothed at older ages “*to obtain an improved representation of the underlying mortality conditions*” ([Wilmoth et al., 2019](#), p. 34).

Mortality measures and functions

Having described the structure and source of data used in this dissertation, here we introduce the measures and functions that are typically used in mortality analysis.

Modelling and forecasting human mortality is a branch of demography that is closely intertwined with the analysis of survival data in statistics. We therefore present the measures and functions employed in the study of human mortality linking the terminologies used in survival analysis (derived from [Klein and Moeschberger, 2003](#)) with those used in demography (derived from [Preston et al., 2001](#)).

Survival analysis focuses on the study of durations until the occurrence of a specific event of interest. For the purposes of this dissertation, the event of interest is death, although it could be any other event, such as marriage, conception, onset of a disease and so forth. The duration until the specified event is denoted with the random variable X , which is typically continuous and non-negative. For the time being, let us think of X as the age dimension of the mortality data just introduced, and let us temporarily drop the time (calendar year) dimension to ease notation. There are three functions that characterize the distribution of X : (i) the probability density function $f(x)$, which is the unconditional probability of death at age x , (ii) the survival function $S(x) = \Pr(X > x)$, which is the probability of an individual surviving to age x , and (iii) the hazard rate (function) $h(x)$, which is the instantaneous rate of death at age x , conditioned upon surviving to that age, or formally:

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{\Pr(x \leq X < x + \Delta x | X \geq x)}{\Delta x}. \quad (1.1)$$

As such, $h(x)\Delta x$ can be interpreted as the approximate probability that an individual of age x dies in the next instant of time, given that she survived until x .

Figure 1.3 shows an example of the three mortality functions of the Gompertz model, one of the most frequently employed parametric models of human mortality, which will be formally introduced in Section 1.2. For our purposes here, the graphs are useful to illustrate the properties of the three functions. The left panel shows that the density is a non-negative function, whose area under the curve is equal to one, i.e. $\int_0^\infty f(x) dx = 1$. The central panel displays the survival function, which is monotonically decreasing and bounded between 0 and 1, with $S(0) = 1$ and $\lim_{x \rightarrow \infty} S(x) = 0$. Finally, the right panel shows the hazard function, which is also non-negative at all ages, i.e. $h(x) \geq 0 \forall x$.

An important characteristic of the three mortality functions is that they are uniquely related to each other: it is possible to derive any two of them by knowing the third one, without the need of additional information. Specifically, the relationship that exists between the survival and density functions is:

$$S(x) = \int_x^\infty f(t) dt \Leftrightarrow f(x) = -\frac{dS(x)}{dx} \quad (1.2)$$

Moreover, the hazard and the survival are related by:

$$S(x) = \exp \left[- \int_0^x h(t) dt \right] \Leftrightarrow h(x) = -\frac{d \ln [S(x)]}{dx} \quad (1.3)$$

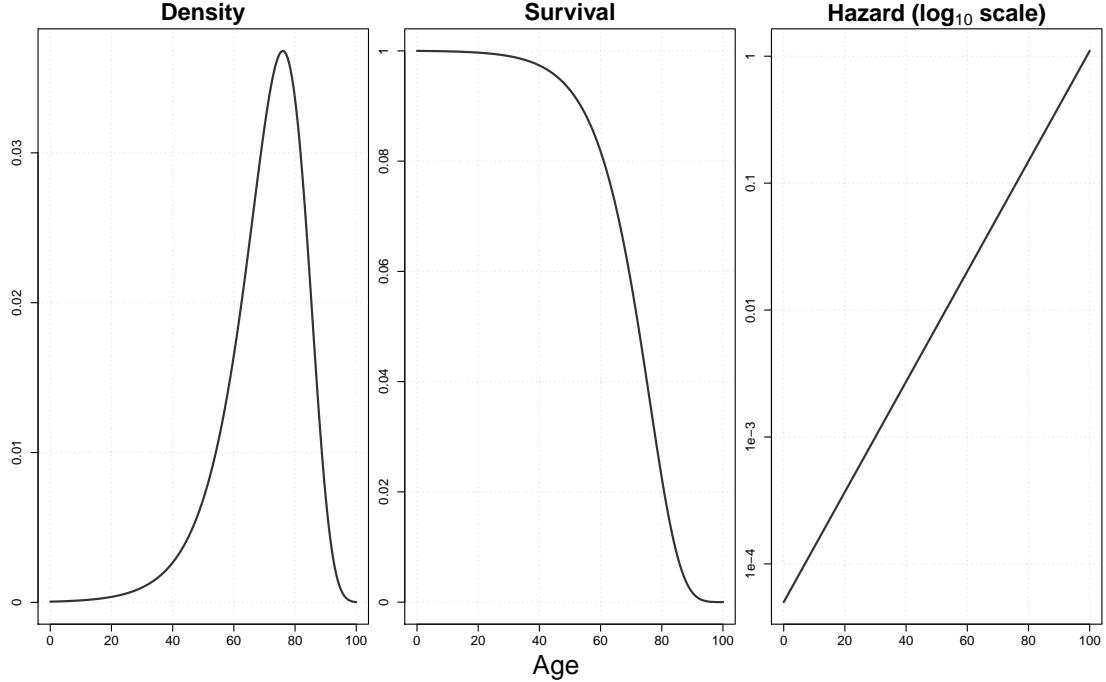


Figure 1.3: An illustration of the density, survival and hazard functions of the Gompertz model for human mortality (cf. Section 1.2).

Source: Own elaborations.

Combining Eq. (1.2) and (1.3) shows the complementarity of the three mortality functions, which can be expressed in terms of each other as:

$$f(x) = h(x) S(x). \quad (1.4)$$

In demography, the three mortality functions are known by the notation and terminology of the life table, and although having different names, they are identical to those used in survival analysis. Let us maintain the continuous notation, before moving to the discrete formulation intrinsic to life tables. Demographers and actuaries generally denote: (i) the density function as the age-at-death distribution $d(x)$, (ii) the survival function as the probability of surviving from birth to age x , $\ell(x)$, and (iii) the hazard function as the force of mortality $\mu(x)$.

Mortality is a continuous process, and the functions just introduced are well suited to capture the inherent nature of the risk of death, which constantly changes over age and time. As discussed in Section 1.1, mortality data are however collected only at particular ages and years. For this reason, the life table has long been used in mortality analysis to illustrate age-specific information about the survival and progressive extinction of a birth cohort. Given its tabular nature, a life table displays mortality functions only at particular ages (or age groups). A discrete notation is therefore appropriate to introduce the other life-table functions.

In addition to ages x and central death rates m_x , a life table comprises several additional columns. First, d_x and ℓ_x (the discrete-time equivalents of $d(x)$ and $\ell(x)$) denote the number of deaths at age x and the number of survivors to age x , respectively. Next, a_x denotes the mean number of person-years lived by those dying in the interval starting with x ; $q_x = d_x / \ell_x$ corresponds to the probability of dying in the age group x , and $p_x = 1 - q_x$ the complementary

survival probability; L_x denotes the number of person-years lived in the cohort for the age group x .

The last column of the life table deserves special attention due to its extensive use in mortality analysis. This column is the life expectancy at age x , e_x , which measures the average number of additional years that a survivor at age x is expected to live beyond that age. Specifically, it is computed as $e_x = \sum_{a=x}^{\omega} L_a / \ell_x$, where ω is the highest age attained in the life table, and it can be expressed in continuous notation as:

$$e(x) = \frac{\int_x^{\omega} \ell(a) da}{\ell(x)}. \quad (1.5)$$

Life expectancy at birth, e_0 , is equivalent to the mean age at death of the life-table cohort, and as such, it is a measure of central tendency of the age-at-death distribution. This measure is practically omnipresent in studies of population health, because it conveniently summarizes the mortality pattern of the population into a single number.

However, life expectancy at birth alone does not capture all the relevant features of the distribution of deaths. Indeed, a specific mean value can originate from very different shapes of the underlying distribution. As such, scholars have recently started to go beyond the analysis of “central longevity indicators” (mean, median and modal age at death, [Cheung et al., 2005](#)) by analysing different summary measures of the distribution.

In particular, the concentration of lifespan distributions in human populations, typically referred to as lifespan inequality, disparity or variability in the demographic literature, has received significant attention during recent decades. Inequality in length of life is indeed the most fundamental of all inequalities, as every other inequality is conditional upon being alive ([van Raalte et al., 2018b](#)).

Several indices have been proposed to measure lifespan inequality, which have been shown to be highly correlated with each other (see, for example, [Wilmoth and Horiuchi, 1999](#); [Vaupel et al., 2011](#); [van Raalte and Caswell, 2013](#); [Colchero et al., 2016](#)). In this dissertation, we employ two measures of lifespan inequality: in Chapters 3-5, we use the Gini concentration index $G(x)$ at age x , also known as Gini coefficient, which is defined as ([Hanada, 1983](#); [Shkolnikov et al., 2003](#)):

$$G(x) = 1 - \frac{1}{e(x) [\ell(x)]^2} \int_x^{\omega} [\ell(t)]^2 dt. \quad (1.6)$$

In Chapter 6, we use the average number of life years lost at age x ([Vaupel and Canudas-Romo, 2003](#)):

$$e^{\dagger}(x) = \frac{\int_x^{\omega} e(t) d(t) dt}{\ell(x)}. \quad (1.7)$$

Both measures have been employed to measure lifespan inequality within and between populations (see, e.g., [Hanada, 1983](#); [Shkolnikov et al., 2003](#); [Smits and Monden, 2009](#); [Shkolnikov et al., 2011a](#); [Vaupel et al., 2011](#); [van Raalte and Caswell, 2013](#); [Gigliarano et al., 2017](#); [Aburto and van Raalte, 2018](#)) and to evaluate mortality forecasts ([Bohk-Ewald et al., 2017](#); [Diaz et al., 2018](#); [Basellini and Camarda, 2019b](#); [Camarda, 2019](#)). For additional details on the Gini coefficient, please refer to Section 5.4.

Finally, two important assumptions that will be used throughout this dissertation should be mentioned here. The first is a central assumption in demographic and actuarial analyses of mortality: the piece-wise constant hazard model. Specifically, the force of mortality $\mu_{x,t}$ is assumed to remain constant over each year of age (from age x to $x+1$) and over each calendar year (from year t to $t+1$). This assumption implies that: (i) $\mu_{x,t}$ approximates the force of mortality at exact age $x + \frac{1}{2}$ and exact time $t + \frac{1}{2}$, and (ii) central death rates $m_{x,t}$ are the maximum likelihood estimators (MLEs) of the force of mortality (Currie, 2016). We refer the interested reader to Chapter 4 of Alho and Spencer (2005) for the derivation of the MLE in the case of n individuals with independent and identically distributed lifespans following the exponential distribution.

Secondly, following the important work of Brillinger (1986), we assume that times of birth and lifetimes in the population are stochastic, with the former following a Poisson process, and the latter being independent of each other, independent of the birth process, and corresponding to the force of mortality $\mu_{x,t}$. These assumptions imply that the observed number of deaths $y_{x,t}$ in the population are realizations of the random variable $Y_{x,t}$, which follows a Poisson process with expected value equal to the product of exposure and force of mortality (see Brillinger, 1986, p. 700):

$$Y_{x,t} \sim \mathcal{P}(e_{x,t} \mu_{x,t}). \quad (1.8)$$

This assumption is particularly useful for estimation purposes. Let the force of mortality $\mu(\boldsymbol{\theta})_{x,t}$ be a function of age, time and a vector of parameters $\boldsymbol{\theta}$. Then, for each time period t , the parameters can be estimated by maximising the Poisson log-likelihood:

$$\ln \mathcal{L}(\boldsymbol{\theta} | y_{x,t}, e_{x,t}) \propto \sum_x [y_{x,t} \ln (\mu(\boldsymbol{\theta})_{x,t}) - e_{x,t} \mu(\boldsymbol{\theta})_{x,t}] . \quad (1.9)$$

where the logarithm of the exposures is generally referred to as an “offset” in the context of Generalized Linear Models (GLMs, McCullagh and Nelder, 1989).

In practice, the Poisson assumption implies that the relative importance of central death rates is given by the number of deaths occurring at that age. As such, the parameters of a mortality model will be estimated to maximise the model’s goodness-of-fit at those ages where most of the deaths occur. Figure 1.4 displays an example of the weights and relative importance of the death rates shown in the right panel of Figure 1.2. Figure 1.4 clearly shows that rates at ages 80–90 are the main target of a mortality model embedded in a Poisson setting, whereas rates at young ages (excluding age 0) carry less weight.

1.2 Mortality modelling

According to Coale et al. (1983), Graunt’s celebrated *Observations* (1662) marked the first investigation into the regularities of the mortality age-pattern. By analysing the bills of mortality in London (i.e. weekly accounts of burials and christenings collected by various parishes), Graunt listed both the number of deaths and survivors by decades for a cohort of 100 persons, thereby introducing the first example of one of the most important tools in demography – the life table.

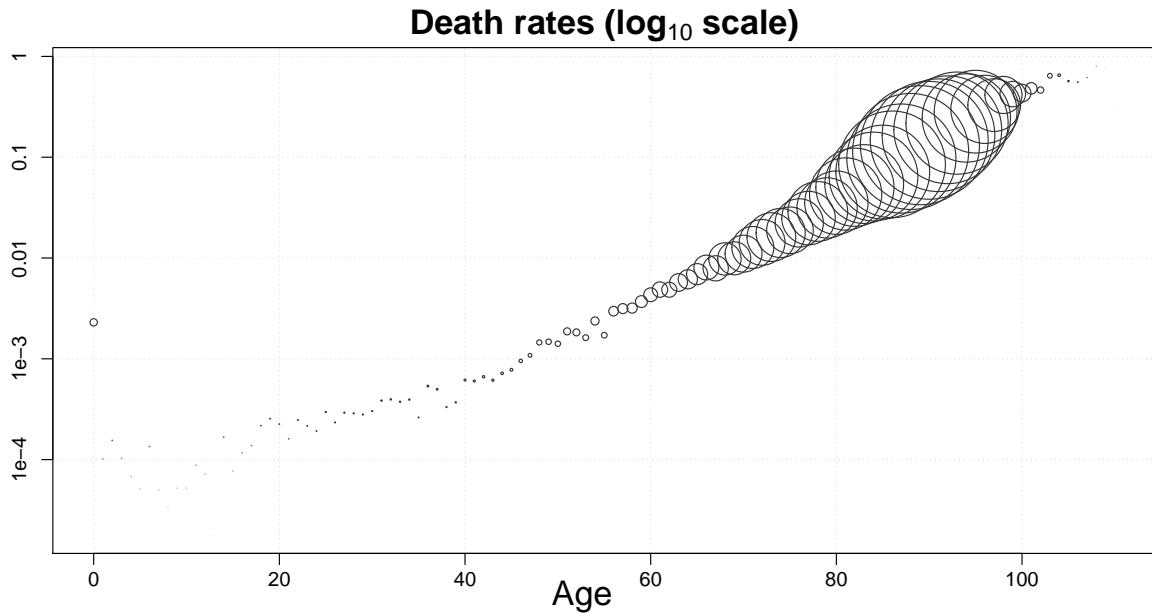


Figure 1.4: Observed central death rates (on \log_{10} scale) for Swedish females at ages 0–110+ in 2017. Circles are proportional to the number of deaths at the corresponding age.

Source: Own elaborations based on data from the [Human Mortality Database \(2019\)](#).

Graunt's life table was based on rather limited data, as London lacked reliable census data at the time, and the bills of mortality did not specify the ages of the deceased. Graunt complained of these deficiencies, and tried to estimate age-specific death rates indirectly by computation ([Cole, 1957](#)). For this reason, some scholars regard the work of [Halley \(1693\)](#), which was based on the more accurate register data of the city of Breslau in 1687–1691, as the first example of the modern life table. From around 1750, several other efforts were directed towards the production of life tables, which took a central place in population studies ([Coale et al., 1983](#)).

Modern life tables expressed in single years of age comprise 500 to 1000 numbers, depending on the number of conventional columns printed. However, much of this information is redundant ([Keyfitz, 1982](#)). For this reason, several efforts have been directed towards the specification of a “mortality law”, i.e. a parametric function of age, which can capture the relevant features of the age-pattern of human mortality using only a few parameters.

There are several reasons why parametric models of mortality have received so much attention during the last two centuries ([Keyfitz, 1982; Congdon, 1993](#)):

- to smooth the data. Mortality rates of adjacent age groups sometimes display random fluctuations, mostly due to small sample sizes. A smooth parametric curve eliminates these uneven variations, thereby offering a more plausible representation of the underlying mortality pattern.
- to describe mortality rates of several different age groups parsimoniously, by using only a few parameters. This property further allows one to derive mortality rates at any given age group (i.e. interpolation).
- to aid inferences from incomplete data. When vital registers of a population are not available, one can infer the mortality pattern by using the partial information at hand and a specific class of models (e.g. relational models, model life tables).

- to facilitate comparisons of mortality. A model's parameters summarize the mortality conditions of the population analysed. As such, differences in the mortality patterns can be more easily detected by comparing the estimated parameters of different populations.
- to construct life tables. It is often forgotten that the life-table construction relies on a parametric model of mortality, namely the piece-wise constant hazard model (cf. Section 1.1).
- to facilitate mortality forecasting. Time-series extrapolations of the model's parameters easily allow one to obtain future mortality schedules.

An early attempt to describe the age-pattern of mortality with a mathematical formula goes back to [DeMoivre \(1725\)](#), who hypothesised that “the number of lives existing at any age is proportional to the number of years intercepted between the age given and the extremity of old age” ([DeMoivre, 1725](#), cited in [Smith and Keyfitz \(1977\)](#), p. 273). In terms of the mortality functions introduced in the previous Section, DeMoivre's formula can be expressed as $\ell(x) = \ell(0) \left(1 - \frac{x}{\omega}\right)$, where $\ell(0)$ is the life-table radix, and ω the maximum attainable age in the population. Without loss of generality, let the life-table radix be equal to one. The force of mortality of the one-parameter DeMoivre's law thus follows:

$$\mu(x) = \frac{1}{\omega - x}. \quad (1.10)$$

A century later, one of the most prominent laws of mortality was contributed by Benjamin Gompertz. By analysing life tables, Gompertz observed a recurrent pattern, a geometric increase in mortality at adult ages: “this law of geometrical progression pervades, in an approximate degree, large portions of different tables of mortality” ([Gompertz, 1825](#), p. 514). As such, Gompertz's law of mortality can be expressed by an exponential increase of mortality with age:

$$\mu(x) = a e^{bx}, \quad (1.11)$$

where the two parameters $(a, b) > 0$ represent the level of the force of mortality at the starting age of analysis (typically, age 30), and the rate of ageing, respectively. An example of the Gompertz law is shown in Figure 1.3 of the previous Section. The force of mortality (on a logarithmic scale) increases linearly with age, and the model's parameters a and b correspond to the intercept and slope of the linear model, respectively.

The two-parameter Gompertz law has proved to be a remarkably good model for different populations and epochs ([Forfar, 2004](#)): it is an almost universal pattern that has been found to apply (over appropriate age ranges) in many countries during the last 170 years ([Thatcher et al., 1998](#)). Nevertheless, two limitations have been highlighted throughout the years: (i) the model does not include excess accident mortality at young adult ages and (ii) the model does not capture the deceleration of mortality at advanced ages ([Manton et al., 1993; Vaupel et al., 1998](#)).

The underestimation of mortality at young adult ages was identified and addressed a few decades after Gompertz by [Makeham \(1860\)](#), who suggested modifying Gompertz's law by adding an additional parameter to capture the constant risk of death from all causes which does not

depend on age. Makeham's law of mortality is thus given by:

$$\mu(x) = c + a e^{bx}. \quad (1.12)$$

Logistic parameterizations have instead been proposed to account for the overestimation of mortality at advanced ages. These models are characterized by an upper bound for the force of mortality, which decelerates at higher ages following a mortality plateau (i.e. a horizontal asymptote). An early attempt in this direction was contributed by Perks, who proposed a generalization of Makeham's law because: "it was found that the Makeham curves ran much too high at the older ages" (Perks, 1932, p. 13). Nowadays, one of the most used logistic-type mortality law is the Kannisto model:

$$\mu(x) = \frac{a e^{bx}}{1 + a e^{bx}}. \quad (1.13)$$

Analysing thirteen countries with a sufficiently long time-series of reliable data, Thatcher et al. (1998) found that logistic curves fit the mortality pattern at old ages at least as well as, and usually better than, any other mortality models. Hence, the Human Mortality Database (2019) adopted the Kannisto model to smooth death rates above age 80 in life-table construction (see Wilmoth et al., 2019, p. 34).

All the mortality laws described thus far have been proposed to model the adult age-pattern of mortality. However, the non-monotonic shape of human mortality has long been observed and recognized by demographers and actuaries. The Danish astronomer and actuary Thiele (1871) contributed one of the first attempts to model the entire age-pattern of mortality: in particular, he suggested describing human mortality with three different groups that operate principally, or almost exclusively, upon childhood, middle and old ages, respectively. Formally, he suggested the following expression for the force of mortality:

$$\mu(x) = \mu_1(x) + \mu_2(x) + \mu_3(x), \quad (1.14)$$

where the force of mortality $\mu(x)$ at age x is additively decomposed into three independent components, $\mu_1(x)$, $\mu_2(x)$, and $\mu_3(x)$. For mortality in childhood, $\mu_1(x)$, Thiele suggested a negative exponential function; for middle ages, $\mu_2(x)$, he suggested a negative log-parabolic curve, while for higher ages he adopted Gompertz's law. It should be further highlighted that Thiele's contribution pioneered decomposition modelling by providing one of the first examples of competing-risk hazard models.

In recent decades, several attempts have been made towards modelling the whole pattern of mortality, all inspired by the work of Thiele. Among the most relevant contributions, Siler (1979) proposed a five-parameter model with an age-independent constant hazard for the term $\mu_2(x)$ in Eq. (1.14). Although originally developed to portray animal mortality, the model has been used for human populations as well (see, e.g., Siler, 1983; Gage, 1993; Canudas-Romo and Schoen, 2005; Canudas-Romo, 2008; Bergeron-Boucher et al., 2015). Furthermore, Heligman and Pollard (1980) followed Thiele's parameterization more closely and proposed an eight-parameter model to portray human mortality. Kostaki (1992), in turn, suggested extending the Heligman-

Pollard model by adding an additional parameter, which improves the fit to mortality data at younger adult ages. Moreover, [de Beer and Janssen \(2016\)](#) proposed a ten-parameter model that directly captures the shifting and compression dynamics of mortality changes. Finally, [Mazzuco et al. \(2018\)](#) introduced a six-parameter mixture distribution function to model the age-at-death distribution, which has been extended with the inclusion of two additional parameters by [Zanotto \(2017\)](#). For an additional short review of parametric mortality models, please refer to Subsection 2.1.2.

1.3 Mortality forecasting

The laws of mortality introduced in the previous Section aim to describe the age-pattern of mortality for a population at a specific point in time, or for a specific cohort of individuals. This class of mortality models is generally referred to as age-models, as the age dimension is the only explanatory variable used in the model. Another approach to modelling mortality, typically stimulated by the goal of forecasting future mortality conditions, is given by including the period dimension (i.e. calendar years) in the modelling process. This approach has generated the class of age-period models, which are widely employed in mortality forecasting. This Section reviews the most important contributions to this field of research.

In a major review paper on demographic forecasting, [Booth \(2006\)](#) distinguishes three approaches to forecasting demographic processes: extrapolation, expectation and explanation. Extrapolation is the most common approach in current demographic forecasting, and the majority of statistical offices and international organizations employ extrapolative forecasting methods ([Booth and Tickle, 2008; Stoeldraijer et al., 2013](#)). As such, here we focus on this forecasting approach.

Forecasts of mortality can be traced back at least to the beginning of the twentieth century, when English actuaries started to assess the financial effects of longevity improvements on the reserves of pension providers ([Pollard, 1987](#)). For example, the London Institute of Actuaries' annuitants tables of 1924 were produced by extrapolating age-specific death probabilities observed between 1900 and 1920 to an ultimate level of mortality ([Anderson and Dow, 1948](#), pp. 183–193).

Until the 1980s, the methods used to forecast mortality were relatively simple and involved a fair degree of subjective judgement ([Booth and Tickle, 2008](#)). An excellent review of the methods in use at that time can be found in [Pollard \(1987\)](#), who distinguished six types of projections by: (i) extrapolation of mortality rates at selected ages, (ii) reference to a law of mortality, (iii) reference to model life tables, (iv) reference to another more advanced population, (v) reference to an optimal life table, and (vi) causes of death.

It is however during the last three decades that mortality forecasting has prospered owing to the introduction and development of stochastic methodologies to project mortality. Much of the success enjoyed by the field has been stimulated by the seminal contribution of [Lee and Carter \(1992\)](#). Lee and Carter proposed an elegant and powerful methodology to model and forecast mortality based on a log-bilinear form for central death rates $m_{x,t}$. Formally, the Lee-Carter

(LC) model can be expressed as:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \quad (1.15)$$

where α_x describes the average shape of age-specific mortality, β_x the rate of mortality improvement at age x , and κ_t the general level of mortality at time t . The $\varepsilon_{x,t}$ are error terms that reflect residual age-specific historical influences not captured by the model. Since the model is underdetermined, two standard constraints are used to ensure model identification, namely $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$.

The LC model is today the best known and widely employed forecasting methodology. Three main reasons explain why the model has gained so much attention: (i) the functional form is not complex, as a matrix of (logged) age-specific mortality rates over time is summarized by two age-specific parameters and one time-varying index, (ii) forecasting is greatly simplified, as mortality forecasts are derived from the projection of the single time-index, typically using a random walk model with drift, and (iii) the model is probabilistic, thereby allowing the derivation of prediction intervals of mortality rather than single deterministic point forecasts.

Since its introduction almost thirty years ago, different limitations of the model have been highlighted, and new extensions have been proposed to overcome them. In its original form, the LC parameters are estimated by ordinary least-squares (OLS) using singular value decomposition. In addition, the first estimates of $\hat{\kappa}_t$ are adjusted so that the fitted deaths of the model match the number of observed deaths in each year t . The main drawback of the OLS approach is that the errors are assumed to be homoskedastic and normally distributed, which is an unrealistic assumption for human mortality ([Alho, 2000](#)). Indeed, the logarithm of the central death rates is more variable at older than at younger ages because of the smaller number of deaths ([Brouhns et al., 2002](#)). To overcome this limitation, [Brouhns et al. \(2002\)](#) embedded the LC model in a Poisson setting, which is a more realistic assumption for human mortality characterized by a heteroscedastic error structure.

A second relevant limitation of the LC model is the assumption of a constant rate of age-specific mortality improvement over time ([Lee and Miller, 2001](#)). The assumption has proven wrong in low-mortality countries during recent decades: rates of mortality improvement have declined at infant and childhood ages, and they have increased at older ages ([Kannisto et al., 1994](#); [Vaupel et al., 1998](#); [Wilmoth and Horiuchi, 1999](#)). As a result, it has been shown that the model tends to underpredict future gains in life expectancy ([Lee and Miller, 2001](#)). To overcome this drawback, [Li et al. \(2013\)](#) proposed rotating the β_x schedule for long-term projections. This rotation captures the observed deceleration of mortality improvements at younger ages, and the acceleration at older ages.

Third, the LC fitted and forecast life tables are typically irregular, characterized by a high degree of jaggedness in the mortality age-profile. This is a consequence of the estimated model's parameters, which are generally volatile (especially the β_x schedule). To overcome this limitation, smoothing techniques have been implemented within the LC framework. [Renshaw and Haberman \(2003c\)](#) suggested smoothing the estimated series of α_x and β_x using parametric or non-parametric methods. [Hyndman and Ullah \(2007\)](#) start by smoothing the observed data using a functional data approach ([Ramsay and Silverman, 2005](#)). Furthermore, [Delwarde et al.](#)

(2007) and Currie (2013) proposed estimating the LC parameters by maximising a penalized log-likelihood function, which results in smooth parameter estimates and, in turn, regular forecast age-patterns.

In addition to the already presented extensions, four other improvements of the LC model deserve mention, as they have received considerable attention and have been used in a number of publications. First, Lee and Miller (2001) suggested the second-step adjustment of the time-index to match exactly the observed value of life expectancy at birth e_0 . This allows one to reduce the jump-off bias, which can be particularly severe in the early years of the forecasts (Lee and Miller, 2001). Second, Booth et al. (2002) adjusted the time-index to reproduce the age-at-death distribution, and they further suggested a procedure to optimally restrict the fitting period in the presence of a non-linear trend in the time index. Third, Renshaw and Haberman (2006) extended the model to account for cohort effects. Fourth, Hyndman and Ullah (2007) proposed employing a greater number of principal components to capture additional dimensions of change in mortality rates. In addition to this, the authors start by smoothing the observed data, and they allow for greater flexibility in the selection of the time-series model of the time parameters.

Two main alternative approaches to the LC methodology have been introduced to model and forecast age-specific mortality functions. First, Cairns et al. (2006) introduced a two-factor model for the logit of the probability of death for older age mortality (above age 60). The model and its extensions have stimulated considerable interest (see, e.g., Cairns et al., 2009, 2011; Dowd et al., 2010a,b), and they have become a benchmark model in the actuarial literature. Second, Currie et al. (2004) proposed smoothing the mortality surface with a tensor product of B -spline bases and smoothing parameters over ages and years (Eilers and Marx, 1996). Forecasts are obtained by extrapolating the fitted surface over the time direction. This model has been widely employed for projecting mortality in England & Wales by the Continuous Mortality Investigation (see, e.g., CMI Committee, 2007). Camarda (2019) recently introduced a promising extension of the model that constrains mortality forecasts to lie within acceptable age profiles and time trends, thereby removing the unreasonable projections that may arise from the original model.

A different approach to the forecasting problem has also gained interest in the last decade. In particular, rather than specifying a model for age-specific mortality functions, some scholars have investigated and modelled the evolution of life expectancy. Among the pioneers of this approach, Torri and Vaupel (2012) exploited the linearity of the trend in e_0 in low-mortality countries, and proposed forecasting the best-practice life expectancy and the gap between a given population and the best-practice level. Moreover, Raftery et al. (2013) introduced a Bayesian hierarchical model to produce probabilistic projections of e_0 for all countries. The methodology is currently used to produce the population projections published biennially in the World Population Prospects of the United Nations (Raftery et al., 2014). Furthermore, Pascariu et al. (2018) proposed a model to jointly forecast female and male life expectancy at any age (e_x) based on the gaps between the best-practice and female e_x on one side, and the gap between female and male e_x on the other.

Finally, some recent efforts have been directed towards the inclusion of coherence as an additional factor in mortality forecasting. Motivated by the observed global convergence in

mortality levels ([Wilson, 2001](#)), [Li and Lee \(2005\)](#) proposed modifying the LC model to account for the similarities within a group of populations when projecting mortality. This approach avoids the situation that differences among sub-populations (such as sexes or regions within a country) implausibly increase in long-run forecasts. Other notable approaches that have explicitly included coherence in mortality forecasts are the models proposed by [Hyndman et al. \(2013\)](#), [Janssen et al. \(2013\)](#) and [Bergeron-Boucher et al. \(2017, 2018\)](#).

1.4 Aims of the thesis

The principal goal of this dissertation is to provide new contributions to the field of mortality modelling and forecasting by proposing innovative statistical methods that can offer novel insights into the study and forecasting of human mortality. More specifically, four main aims are pursued within this thesis:

- Aim 1:** introduce a family of models that can reconcile and generalize the most common parametric models of mortality. The proposed family of location–scale models achieve this goal, and it is further characterised by desirable properties from a demographic and statistical perspective (Chapter 2).
- Aim 2:** propose a new paradigm in mortality forecasting that departs from the conventional approach of modelling and projecting death rates or probabilities. Specifically, age-at-death distributions are used for studying and forecasting mortality changes. Relational age-at-death distribution models provide a parsimonious and efficient approach to pursue this goal (Chapters 3, 4 and 5).
- Aim 3:** investigate the rather unexplored area of forecasting mortality from the age-cohort perspective, which allows one to complete the mortality profiles of partially observed cohorts (Chapter 5).
- Aim 4:** introduce a novel dimension in mortality forecasting, namely the decomposition of the age-pattern into childhood, early-adulthood and senescent components of mortality (Chapters 4 and 6).

In the following Sections, we introduce the five papers that comprise this dissertation. Each paper corresponds to a Chapter of the thesis. The studies take the form of research manuscripts, which have been published or submitted to scientific journals and one book.

For the purpose of this Introduction, we describe the Chapters presenting their overall goals, the rationale behind the methodological approaches and the main results, without going deeply into the mathematical details. This should provide a first accessible introduction of the manuscripts to a broad audience. We refer interested readers to the following Chapters for the full presentation of the manuscripts.

1.5 Ch. 2: location–scale models in demography

As introduced in Section 1.2, the search for a law of mortality has a long tradition in demographic and actuarial analysis: several parametric models have been proposed over the years to describe the age-pattern of human mortality. However, few efforts have been directed towards

the search for an encompassing family that could reconcile many of these models within a single framework. One notable attempt in this direction is the “Gompertz–Makeham formula of type (r, s) ” contributed by [Forfar et al. \(1988\)](#): this class of models comprises, among others, the laws of [Gompertz \(1825\)](#), [Makeham \(1860\)](#), Barnett ([CMI Committee, 1974](#)) and Wilkie ([CMI Committee, 1976](#)), and it was proposed with the aim of graduating mortality tables.

A general family of models that contains several laws of mortality as special cases has at least two appealing features. First, such a family has an inherent feature of flexibility, which is a desirable property for modelling the mortality patterns of different populations. Populations, defined as a collection of persons alive at a specified point in time who meet certain criteria ([Preston et al., 2001](#)), typically differ by their composition of sex, birth cohort, marital and social status, lifestyles, health and regional geography. As such, a flexible family of models may be better suited to capturing the singular features of the population analysed. Second, the generalisation of different laws of mortality within a single framework is a rather elegant result in itself, which can better highlight the relationships existing between different models.

For these reasons, in Chapter 2 we propose employing location–scale (LS) distributions, a well-known family of models in survival analysis and reliability theory, as an overarching family for modelling adult mortality in human populations. It is indeed possible to show that several laws introduced in Section 1.2, as well as others, can be re-parameterized in terms of the LS framework. Please refer to Tables 2.1 and 2.3 in the following Chapter for a comprehensive list of models belonging to the LS family.

In addition to its flexibility, the two-parameter LS family is characterized by desirable properties. From a demographic perspective, the location and scale parameters of the family satisfy the requisites of familiarity and interpretability ([Bell, 1997](#)). Parametric models are generally employed to facilitate analysis and comparisons of mortality patterns, and the LS parameters provide useful demographic insights on the mortality developments of the populations analysed. Specifically, the two parameters capture the shifting and compression dynamics of mortality changes observed in most developed countries during the last century (see, e.g., [Fries, 1980](#); [Kannisto, 2001](#); [Bongaarts, 2005](#); [Janssen and de Beer, 2019](#)).

Furthermore, LS parameters have a second advantage from a statistical perspective. In general, parameters’ estimates of mortality laws in their traditional formulation are highly correlated (cf. Figure 2.6 in Chapter 2, and see, for example, [Missov et al., 2015](#)). This high correlation is problematic because it can reduce the parameters’ interpretability and introduce a source of bias in the estimation procedure. Re-parameterizing mortality models in terms of the LS family significantly reduces this correlation, both between and within populations (cf. Figures 2.7 and 2.8). The interpretability of the LS parameters is therefore enhanced, and the bias in their estimation reduced (cf. Table 2.2).

1.6 Ch. 3: modelling and forecasting adult distributions

The great majority of the proposed and established approaches to model and forecast human mortality (cf. Section 1.3) is based on central death rates, the maximum likelihood estimators of the force of mortality. Given that mortality is characterized by three complementary functions,

it is somewhat surprising that the survival function and age-at-death distributions have been almost disregarded for the forecast of mortality.

One of the main reasons why mortality rates have been the predominant choice for studying and forecasting mortality is that they readily depict mortality developments over age and time. A simple plot of the evolution of mortality rates for a given population, as shown for Swedish females between 1950 and 2017 in the left panel of Figure 1.5, provides a direct representation of the changes in the mortality age-pattern. The downward shift of the mortality curve clearly portrays the reduction in mortality at all ages during the period analysed.

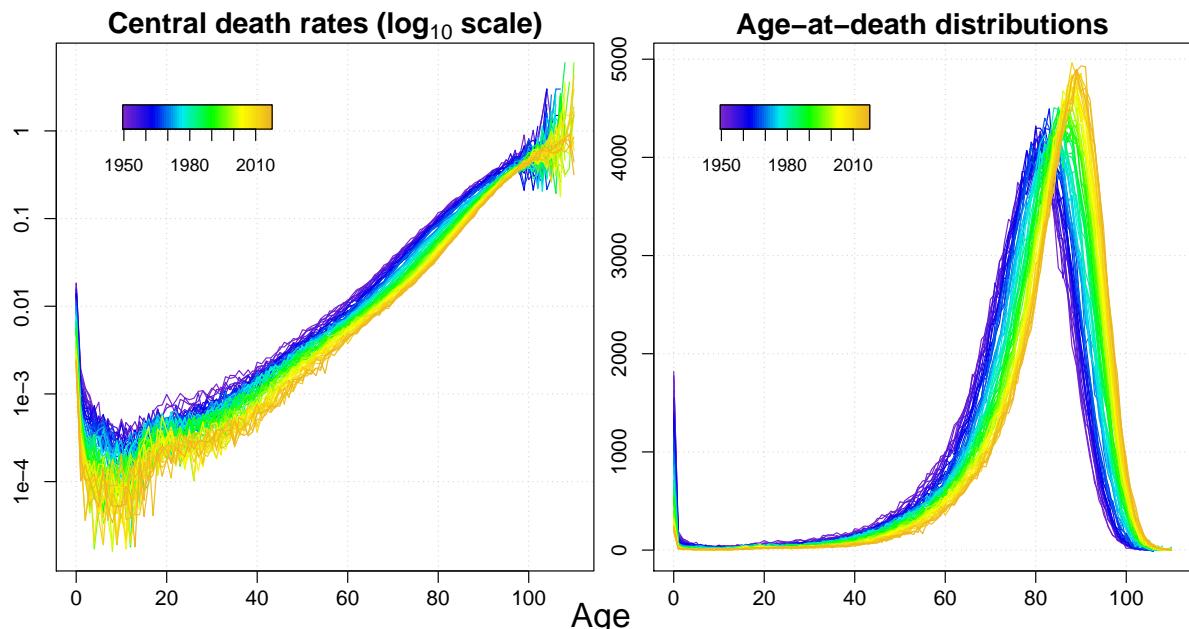


Figure 1.5: Evolution of central death rates (on \log_{10} scale, left panel) and age-at-death distributions (right panel) for Swedish females during 1950–2017.

Source: Own elaborations based on data from the [Human Mortality Database \(2019\)](#).

Age-at-death distributions provide a different yet informative perspective on mortality developments that cannot be directly inferred from the analysis of mortality rates. The right panel of Figure 1.5 shows the evolution of the age-at-death distributions corresponding to the rates shown in the left panel. The shifting and compression dynamics of mortality captured by the location-scale family (Cf. Chapter 2) are detectable from the changes in the age-at-death distribution. Specifically, the graphs clearly show the shift of the distribution towards older ages, accompanied by a decrease in the overall variability.

As such, age-at-death distributions are perfectly placed to answer two key questions in mortality studies: (Q1) how long do we live on average? and (Q2) how variable is the age at which we die? ([Ouellette and Bourbeau, 2011](#)). Indeed, the distribution of deaths provides key insights on the “central longevity indicators” (mean, median and modal age at death, [Cheung et al., 2005](#); [Canudas-Romo, 2010](#), Q1), as well as on the variability of lifespans (Q2), an area of research that has recently received increasing attention due to its relevant implications for public health ([van Raalte et al., 2018b](#)). As a consequence, the distribution of deaths has been subject to increasing interest in mortality analysis during the most recent years (see, e.g., [Ouellette and Bourbeau, 2011](#); [Mazzuco et al., 2018](#); [Keilman, 2019](#)).

Despite being well suited to analysing mortality developments, only a few efforts have been made to leverage age-at-death distributions for modelling and forecasting mortality (see, e.g., [Oeppen, 2008](#); [Oeppen and Camarda, 2013](#); [Bergeron-Boucher et al., 2017](#); [Pascariu et al., 2019](#)). For these reasons, Chapter 3 introduces an age-at-death distribution model to forecast adult mortality. The proposed approach is a relational model, where a time-unvarying standard (or reference) distribution is transformed over time to capture the observed mortality changes. The transformation takes the form of a segmented linear function, hence we denote the model Segmented Transformation Age-at-death Distributions (STAD). The STAD model is parsimonious and efficient: its three parameters allow us to successfully capture adult mortality developments over age and time, and to disentangle the shifting and compression dynamics of mortality change. Moreover, mortality forecasts can be derived from the extrapolation of the three parameters using standard time-series models.

In the context of projecting mortality in high-longevity countries, the STAD point forecasts are generally more accurate than those of the benchmark [Lee and Carter \(1992\)](#) model and its extensions, and STAD prediction intervals are well calibrated. Furthermore, STAD forecasts until 2040 are more optimistic and characterized by stronger shifts and smaller compression than those of the methodologies based on mortality rates.

1.7 Ch. 4: modelling and forecasting age-at-death distributions

Demographers are generally interested in mortality patterns comprising the full age range. Knowledge of the entire mortality curve is required to derive widely used summary measures that are not conditioned upon surviving to some ages. For example, life expectancy at birth can only be computed when mortality data are available at all ages.

As such, it is desirable to extend the STAD methodology to produce forecasts for the entire mortality pattern. This is not a straightforward task: using the STAD model (cf. Chapter 3) from age 0 instead of age 30 significantly reduces fitting and forecasting accuracy. The reason for this lies in the shape of the human mortality pattern. Since the seminal contribution of [Thiele \(1871\)](#), demographers decompose the mortality age-profile into three different groups operating principally, or almost exclusively, upon juvenile, younger adult and older adult ages. The STAD model is specifically designed to study and forecast the Senescent component of mortality (i.e. mortality from age 30 onwards), and a plain extrapolation of the standard distribution disregards the first two mortality components, resulting in a loss of accuracy.

Chapter 4 proposes a two-step approach to forecast mortality at all ages using age-at-death distributions and the STAD framework. The procedure is based on: (i) a non-parametric decomposition of the mortality pattern into three independent components corresponding to Childhood, Early-Adulthood and Senescence, respectively, and (ii) modelling and forecasting changes in the three component-specific distributions using specialized versions of the STAD. As such, we call our model Three-Component Segmented Transformation Age-at-death Distributions (3C-STAD). Overall mortality developments are obtained by combining the three components, and mortality forecasts are derived from parameters' extrapolation using standard time series models.

We compare forecasts of the 3C-STAD model with those of three other well-known approaches to forecasting mortality in two high-longevity countries by sex. The 3C-STAD fore-

casts generally outperform those of other models, as they are more accurate in terms of both point forecasts and prediction intervals. In addition to this, three main findings emerge from forecasting mortality until 2050: (i) the 3C-STAD model forecasts more optimistic longevity improvements than approaches based on mortality rates, (ii) projected mortality schedules of the 3C-STAD are smooth, lacking the jaggedness and irregularities found in other models, and (iii) the 3C-STAD forecast age-at-death distributions are characterized by greater shift and less compression than those of other approaches.

1.8 Ch. 5: forecasting cohort age-at-death distributions

In addition to being based on central death rates, a second common characteristic of the recently proposed approaches to forecast mortality (cf. Section 1.3) is that they focus on the age-period perspective, i.e. the goal of the model is to forecast mortality patterns in future years. A different approach to mortality forecasting, which has been largely overlooked and unexplored in the actuarial and demographic literature, consists in shifting from the age-period to the age-cohort perspective: here, the goal of the model is the completion of the mortality experience of non-extinct (i.e. partially observed) cohorts.

Models to forecast cohort mortality are rather few in the literature: three models have been recently proposed with this explicit purpose ([Chiou and Müller, 2009](#); [Zanotto and Mazzucco, 2017](#); [Rizzi et al., 2019](#)), and the 2D P -spline of [Currie et al. \(2004\)](#) has been used for completing cohorts in England & Wales ([CMI Committee, 2007](#)). The main reason for the scarcity of efforts in this direction is the heavy data demands that such models require. However, this issue is reduced when the analysis is restricted to adult mortality ([Booth, 2006](#)).

Nonetheless, completing the mortality experience of partially observed cohorts is often an important goal for insurance companies, which are typically interested in the mortality developments of clients born in different cohorts. In such cases, cohort projections are obtained by first forecasting mortality in a period fashion, and then by extracting the cohort patterns from the diagonals of the forecast Lexis surface. Despite being widely used, this approach seems rather counter-intuitive and inefficient, and it can generate implausible prediction intervals ([van Raalte et al., 2018a](#)).

Furthermore, analyses of cohort mortality have an important advantage over those based on the period perspective. Cohort mortality developments are actually observed, whereas period ones are based on the assumption of unchanged mortality rates intrinsic to (period) life tables. Survival in real birth cohorts may thus differ from survival in the hypothetical situation of constant mortality rates because of: (i) tempo effects, (ii) cohort effects and (iii) selection (for a full discussion, see [Borgan and Keilman, 2019](#), pp. 90–92). Indeed, analyses of age-cohort data have provided different insights into mortality developments than studies based on the age-period perspective (see, e.g., [Goldstein and Wachter, 2006](#); [Shkolnikov et al., 2011b](#); [Borgan and Keilman, 2019](#); [Keilman, 2019](#); [Nepomuceno and Canudas-Romo, 2019](#)).

Chapter 5 contributes a novel methodology to forecast mortality from the age-cohort perspective. The approach is based on adult age-at-death distributions and a generalization of STAD framework (cf. Chapter 3). As such, we denote our proposed model Cohort Segmented Transformation Age-at-death Distributions (C-STAD). Unlike the case of age-period data, the original

STAD model does not provide a reasonable fit to adult cohort mortality. The main reason is the significant reduction of young adult mortality during the period analysed (cf. Subsection 5.3.2), mostly due to improvements in sanitary environment, public hygiene and nutrition ([McKeown, 1976](#)). However, the generalization of the segmented transformation of the STAD provides an efficient solution to improve the goodness-of-fit of the C-STAD model.

Our analyses show that the C-STAD is able to successfully capture mortality developments across ages and cohorts, and that it can precisely complete the mortality experience of partially observed cohorts. Specifically, the C-STAD methodology is more accurate than: (i) the 2D P -spline model applied to age-cohort data, and (ii) the conventional approach of forecasting age-period mortality with the [Lee and Carter \(1992\)](#) model, and then extracting cohort patterns from the projected Lexis surface.

1.9 Ch. 6: smoothing, decomposing and forecasting mortality rates

In the last chapter of this dissertation, we propose another methodology to model and forecast the entire pattern of mortality from the conventional age-period perspective. Unlike Chapter 4, here we adopt the standard approach encountered in the literature, i.e. we model and forecast central death rates.

Despite its limitations, the seminal model of [Lee and Carter \(1992\)](#) has been widely adopted by international agencies and private companies, making it the most used mortality forecasting model of modern times. As more thoroughly discussed in Section 1.3, the three main shortcomings of the model are: (i) the assumption of normality for the error terms, (ii) the jaggedness of the fitted and forecast mortality rates, and (iii) the assumption of a constant rate of age-specific mortality improvements over time. The numerous extensions of the LC model introduced in recent years have been proposed to deal with one drawback at a time. However, none of them addressed these limitations altogether.

In Chapter 6, we propose a generalization of the LC model that captures the complex nature of the mortality age-pattern and simultaneously overcomes the limitations highlighted in the literature. Specifically, the proposed approach: (i) is framed within a Poisson setting; (ii) enforces smoothness in the model parameters, and hence in the outcomes, and (iii) addresses the drawback of the fixed rate of mortality improvement by decomposing mortality into Childhood, Early-Adulthood and Senescent components. Each component is modelled and simultaneously estimated with a smooth variant of the LC model. We hence denote our approach Three-Component smooth Lee-Carter (3C-sLC) model.

We present the results of fitting and forecasting mortality with the 3C-sLC in four populations, and we compare the outcomes with those obtained from a smooth improved version of the Lee-Carter (LC) model. These analyses show that the 3C-sLC fits the observed mortality data better than the smooth LC model. The increased goodness-of-fit is a direct consequence of the mortality decomposition intrinsic to our proposed approach: mortality developments of the 3C-sLC are described by the combination of the three sets of component-specific LC parameters, which enhance the power and flexibility of the model.

Moreover, the increased flexibility of the 3C-sLC translates into wider prediction intervals. This is a desirable outcome that directly addresses the criticized narrowness of the LC prediction intervals ([Alho, 1992](#)). Finally, specific knowledge on forecast patterns of mortality allows us to perform hypothetical exercises, such as estimating the potential gains in life expectancy derived from the elimination of one or more mortality components.

1.10 Discussion & Outlook

Despite having a long history in demographic and actuarial analysis, mortality modelling and forecasting is still a very active area of research. Numerous models have been recently proposed to study and project mortality developments over age and time. While significant advances have been made, there still is room for improvement: academics and practitioners are far from achieving a consensus on the most appropriate way to model and forecast human mortality. In addition, currently and widely used forecasting approaches have repeatedly failed to anticipate the sustained rate of mortality improvements observed in many low-mortality countries, generating enormous liabilities for public and private sectors.

The goal of this dissertation was to provide a contribution to the literature on mortality modelling and forecasting by proposing new statistical approaches that can bring novel insights to the analysis of mortality developments. Specifically, the five papers that comprise this thesis directly address the four research questions outlined in Section 1.4, namely: (i) propose a general framework for modelling human mortality, (ii) develop a novel paradigm to forecast mortality that is based on age-at-death distributions, (iii) investigate the unexplored age-cohort dimension in mortality forecasting, and (iv) include the decomposition perspective into mortality forecasting.

Studies on mortality developments for large populations are generally based on age-specific death rates. Among others, one of the main reasons for this “rate-centric” perspective is that mortality rates easily allow one to capture the change in the risk of death over age, due to their conditioning on the survivors to this particular age ([Camarda, 2008](#)). As a consequence, the historical development of mortality laws has focused on the force of mortality, i.e. the continuous counterpart of mortality rates.

The location-scale (LS) family of models proposed in Chapter 2 reconciles several of these traditional laws of mortality within a unique framework, and it offers desirable properties with respect to their traditional parameterization. Clearly, the LS family is not omni-comprehensive: several existing laws fall outside the scope of the family, whose focus is restricted to the adult mortality pattern. However, the family provides researchers with a new flexible tool for modelling mortality patterns and studying mortality dynamics, and it elegantly joins several laws of mortality within a common framework.

A by-product of the rate-centric perspective discussed is that age-at-death distributions have been neglected for forecasting the future course of mortality. This is rather surprising, because the distribution of deaths provides meaningful demographic insights on mortality developments. The STAD model and its 3C-STAD extension, covered in Chapters 3 and 4, develop a novel paradigm whereby mortality forecasting is based on age-at-death distributions. These models

enlarge the toolbox of demographers, actuaries and, more generally, researchers interested in mortality projections for specific populations.

A very recent study has shown that the choice of different life-table functions impacts the resulting forecasts, as modelling rates and probabilities of death leads to more pessimistic forecasts than using survival probabilities, life-table deaths and life expectancy ([Bergeron-Boucher et al., 2019](#)). This is consistent with our findings: forecasts based on the STAD methodology and its extensions are more optimistic than those based on the [Lee and Carter \(1992\)](#) model and its variants.

[Chapter 5](#) investigates a second unexplored territory in mortality projections, namely cohort mortality forecasting. Cohort forecasts of mortality are interesting for practical purposes: completing the survival experience of partially observed cohorts allows researchers to anticipate demographic developments of real birth cohorts, whose mortality is actually observed and not based on the assumption of constant mortality rates (as for synthetic cohorts of period life tables). Studies of cohort mortality are indeed widespread in the literature, and actuaries are often interested in mortality forecasts of specific cohorts. As such, the proposed C-STAD model can benefit the academic and research community with one of the few approaches to forecasting age-cohort mortality data.

Finally, [Chapter 6](#) proposes a novel extension of the seminal and widely employed [Lee and Carter \(1992\)](#) model that overcomes its principal limitations. This approach, as well as the 3C-STAD model of [Chapter 4](#), introduces a highly innovative perspective into mortality forecasts, namely the decomposition of the mortality pattern into three independent components: Childhood, Early-Adulthood and Senescent mortality. Forecast trends can therefore be analysed in terms of these components: for example, it is possible to decompose longevity increases into the contribution of each component.

Despite the significant advances achieved in mortality forecasting, European demographic forecasts have not become more accurate over the most recent years ([Keilman, 2008](#)). Although forecasts in the social sciences are unlikely to be as accurate as those in the physical sciences, better separation of signals from noise can lead to improvements in forecast accuracy ([Makridakis et al., 2019](#)). The methods proposed in this thesis are a step in this direction: forecasts obtained with the STAD, 3C-STAD and C-STAD models have been shown to produce more accurate point forecasts than other projection methodologies in several out-of-sample validation exercises.

A second important aspect of forecasting concerns prediction intervals. Until recently, little attention was paid to forecast distributions, or measures of forecast distribution accuracy ([Makridakis et al., 2019](#)). However, the uncertainty surrounding point forecasts is equally important as the point forecast itself. Prediction intervals indeed allow for a more thorough comparison of different forecasting methodologies ([Chatfield, 2000](#)). This feature is often neglected in the demographic literature, where the focus is generally restricted to point forecast accuracy. In Chapters 3 and 4, we assessed and compared the prediction interval accuracy of the STAD and 3C-STAD models, finding increased precision when compared to the more traditional forecasting approaches. As such, this dissertation stimulates the consideration of this important procedure.

Several directions of future research can currently be foreseen. Despite enabling reproducibility and implementation of the routines developed in this thesis (cf. Section 1.11), a freely available R package ([R Development Core Team, 2019](#)) in the Comprehensive R Archive Network (CRAN) would allow researchers to more easily employ the methodologies presented in this dissertation. The development of such package is planned for future work.

In light of the recent efforts towards the production of coherent forecasts for multiple (sub)-populations, the STAD framework could be extended to achieve this goal. For example, a coherent standard distribution could be derived from combining the mortality experience of different populations, and the population-specific parameters could be forecast within a multivariate time-series setting. This would allow one to include coherence as an additional factor in age-at-death distribution forecasts.

Furthermore, the STAD methodology could be enlarged outside the analysis of all-cause mortality. Cause-specific age-at-death distributions could be analysed and forecast with the STAD model or its modifications. For example, it could be possible to distinguish between smoking and non-smoking related mortality, as in [Janssen et al. \(2013\)](#). In addition to this, fertility and migration data could be analysed and forecast with adaptations of the STAD approach.

Finally, one limitation of the 3C-STAD and C-STAD models that could benefit from future work concerns computation speed. Generating results from the two models, specifically the prediction intervals, requires more time than other widespread forecasting approaches. Increased computer power will accelerate these procedures, but more specific solutions could be developed.

1.11 Reproducibility

All manuscripts of the thesis have been written with the goal of complying to the good practices of open science, namely code sharing and reproducibility of results. Routines for implementing the models presented in this thesis, which have all been developed in the free R software ([R Development Core Team, 2019](#)), are already available online for the articles that have been published and accepted for publication, and they will be made public upon eventual acceptance of the two manuscripts under review.

Specifically, routines for fitting the location-scale (LS) family of models presented in Chapter 2 can be found at <https://github.com/ubasellini/LocationScale>, or in the Supplementary Material of the published article (<https://doi.org/10.1007/s10680-018-9497-x>). The available codes readily allow one to estimate the location and scale parameters for any HMD population as well as for the eleven models belonging to the LS family.

Modelling and forecasting mortality with the Segmented Transformation Age-at-death Distributions (STAD) model introduced in Chapter 3 can be readily performed using the codes provided at <https://github.com/ubasellini/ForecastingDistributions>, or in the Supplementary Material of the published article (<https://www.tandfonline.com/doi/full/10.1080/00324728.2018.1545918>).

The generalisation of the STAD model to the entire age range, i.e. the Three-Component STAD (3C-STAD) model presented in Chapter 4, can be implemented with the routines avail-

able at <https://github.com/ubasellini/3C-STADmodel>. Furthermore, this GitHub repository contains all the results presented in the final accepted version of the book chapter.

Moreover, routines and results derived from the Cohort STAD (C-STAD) model of Chapter 5 are placed in the GitHub repository <https://github.com/ubasellini/C-STAD>, which is currently private (i.e. accessible only to collaborators) to comply with the guidelines of the journal where the manuscript is under review. Nonetheless, the repository will be made publicly available upon eventual acceptance of the manuscript.

Finally, routines for smoothing, decomposing and forecasting mortality with the Three-Component smooth Lee-Carter (3C-sLC) model of Chapter 6 are available at https://osf.io/rgf8y/?view_only=1608f0f8d1e9466cb33733bb03057b80. This is a freely accessible open science framework link, which we have made anonymous for peer review of the manuscript.

Chapter 2

Location–Scale Models in Demography: A Useful Re-parameterization of Mortality Models

Ugofilippo Basellini

Vladimir Canudas-Romo

Adam Lenart

European Journal of Population, **35**(4), 645–673 (2019).

DOI: 10.1007/s10680-018-9497-x

Location-Scale Models in Demography: A Useful Re-parameterization of Mortality Models

Ugo Filippo Basellini^{1,2,†}, Vladimir Canudas-Romo³, and Adam Lenart²

¹*Institut national d'études démographiques (INED), Paris, France*

²*Interdisciplinary Centre on Population Dynamics (CPop) and Department of Public Health,
University of Southern Denmark, Odense, Denmark*

³*School of Demography, Australian National University, Canberra, Australia*

† Corresponding author: ugofilippo.basellini@ined.fr

This is a post-peer-review, pre-copyedit version of an article published in *European Journal of Population*. The final authenticated version is available online at: <https://link.springer.com/article/10.1007/s10680-018-9497-x>

Submitted: December 2017; Accepted: September 2018

Abstract

Several parametric mortality models have been proposed to describe the age pattern of mortality since Gompertz introduced his “law of mortality” almost two centuries ago. However, very few attempts have been made to reconcile most of these models within a single framework. In this article, we show that many mortality models used in the demographic and actuarial literature can be re-parameterized in terms of a general and flexible family of models, the family of location-scale (LS) models. These models are characterized by two parameters that have a direct demographic interpretation: the location and scale parameters, which capture the shifting and compression dynamics of mortality changes, respectively. Re-parameterizing a model in terms of the LS family has several advantages over its classic formulation. In addition to aiding parameter interpretability and comparability, the statistical estimation of the LS parameters is facilitated due to their significantly lower correlation. The latter, in turn, further improves parameter interpretability and reduces estimation bias. We show the advantages of the LS family over the typical parameterization of mortality models with two illustrations using the Human Mortality Database.

Keywords: Mortality Modelling · Law of Mortality · Shifting · Compression · Gamma-Gompertz · Extreme-Value

2.1 Introduction

2.1.1 Parametric mortality models

The search for a model of human mortality has a fairly long history: mortality modelling has indeed developed into an established research topic since the first half of the eighteenth century ([Tabeau, 2001](#)). In particular, several efforts have been directed towards parametric models, which assume a parametric distribution to describe the age pattern of mortality.

Parametric models have long been used by actuaries, demographers and medical scientists for smoothing data, eliminating and/or reducing errors, constructing life tables, aiding inferences from incomplete data, facilitating comparisons of mortality, and forecasting ([Keyfitz, 1982](#)). The popularity of these models can be attributed to at least six advantages of representing mortality schedules with a single curve governing all data points ([Congdon, 1993](#)):

- (a) *Smoothness*: there are no uneven age-variations of mortality rates due to random statistical fluctuations. This is particularly advantageous when looking at very old-age mortality, where the number of deaths and exposure-to-risk are much smaller than at other ages.
- (b) *Parsimony*: mortality schedules of several points corresponding to many different ages can be represented only by a few parameters.
- (c) *Interpolation*: mortality rates for any specific age can be analytically derived. In practice, this is very important when only five-year age-group rates are known, or when the mortality schedule is incomplete.
- (d) *Comparison*: mortality data always refer to a specific population, which could be as broad as the national population of a country, or as narrow as the policyholders of an insurance company. Populations differ by their composition of gender, age, marital and social status, lifestyle, health and regional geography (e.g. postcode). The comparison of several mortality patterns can be readily and more easily performed by estimating the model's parameters for each population studied.
- (e) *Trends and forecasting*: the assessment of trends over time and forecasting into the future are facilitated.
- (f) *Analytic manipulation*: the properties of the model employed are generally known, and these can be used in more complex demographic settings. For example, one common procedure is to specify the uncertainty associated with the model projections.

More recently, parametric models have been used to study the shifting and compression dynamics of mortality changes. The remarkable mortality reductions that occurred in most developed countries during the twentieth century are generally divided into two different stages. In the first half of the century, fast mortality declines at younger ages produced a compression of the age-at-death distribution, with the majority of deaths concentrated in a smaller age-interval ([Fries, 1980](#); [Myers and Manton, 1984](#); [Rothenberg et al., 1991](#); [Kannisto, 2000](#); [Cheung et al., 2009](#)). Then, as rates of mortality improvements slowed down at younger ages and accelerated at older ages ([Kannisto et al., 1994](#); [Vaupel et al., 1998](#); [Wilmoth and Horiuchi, 1999](#)), the distribution of deaths started to shift to older ages, with a shape remaining practically constant ([Bongaarts, 2005](#); [Cheung et al., 2005](#); [Cheung and Robine, 2007](#); [Canudas-Romo, 2008](#)).

Bergeron-Boucher et al. (2015) introduced a decomposition method based on a re-parameterization of mortality models that allows to differentiate between the two dynamics and estimate their contribution to increases in life expectancy. In particular, the authors find that the shifting dynamic was responsible for more than 70% of the increase in average lifespan for Swedish females since the mid-1960s. Furthermore, de Beer and Janssen (2016) presented a new parametric model that formally captures the two dynamics, and they show that two-thirds of the increases in life expectancy for Japanese, French, American and Danish females and males between 1950 and 2010 were due to shifting mortality.

2.1.2 Short history of parametric mortality models

The research of a “law of mortality” has been an interesting topic since the development of the first life tables by Graunt (1662) and Halley (1693). One of the very first attempts to model mortality with a parametric distribution goes back to DeMoivre (1725). A century later, Gompertz (1825) made one of the most well-known early contributions to the field, as he theorized an increasing exponential effect of age on the force of mortality.

A few years later, Makeham (1860) suggested a modification of the Gompertz model to overcome the underestimation of mortality at young adult ages by adding a constant non-ageing-related risk term. Shortly afterwards, Thiele (1871) proposed a model to capture the non-monotonic shape of human mortality along the full age range. In particular, he suggested to decompose human mortality into three different groups that operate principally, or almost exclusively, upon childhood, middle and old ages, respectively. This assumption has been extensively used since then for modelling purposes (Siler, 1979; Heligman and Pollard, 1980).

The Logistic model, first discovered by Perks (1932), is a general model which includes the Makeham law as a special case. Recently, this model has been used to describe the force of mortality at very high ages, for example above age 80 (Wilmoth et al., 2000): the logistic function implies that death rates approach a fixed upper limit, as suggested both empirically and theoretically (Vaupel et al., 1998; Gampe, 2010).

Subsequently, the Gamma-Makeham model was introduced by Beard (1971), who showed that a logistic function could arise in a simple model of a heterogeneous population. Indeed, the aggregation of several individual Makeham hazards with different levels of initial mortality can result in a logistic curve (Thatcher et al., 1998). Afterwards, this finding was developed extensively as the *frailty* model (Vaupel et al., 1979).

Along with the Gompertz, the Weibull model is widely used as a parametric form of the hazard function to adequately capture the ageing process (Missov and Vaupel, 2015). Weibull (1951) introduced this model to represent the durability and failure due to wear and tear of technical systems, such as ball bearings, automobile components, and electrical insulation. An analogy to mortality can be made by considering death as the result of the failure of bodily organs or damage to cells (Thatcher et al., 1998). The model has been used extensively in biological and medical research, for example, in studies on the time of occurrence of tumours in human populations or laboratory animals (Lawless, 2011).

One limitation of all the models for adult mortality mentioned above is that their hazard function, be it increasing, decreasing or constant, must be monotonic, whatever the values of

the parameters. This may be inappropriate in some settings, for example when the course of a disease is such that mortality reaches a peak after some finite period, and then slowly declines. Among others, two models have been proposed to overcome this issue: the Log-Logistic and the Log-Normal models. Indeed, these models have non-monotonic hazard functions, which make them suitable in several situations, such as the modelling of some sets of cancer survival data (Bennett, 1983).

Finally, in recent years, it has been claimed that the Extreme-Value model could provide great future prospects for mortality analysis and forecasting (Willekens, 2001). This model is generally used to describe the failure times or lifetimes of systems that cease to function whenever the weakest component fails.

2.1.3 Aims

In this paper, we aim to show that several well-known mortality models can be re-parameterized in terms of two parameters that describe the shifting and compression dynamics of mortality changes. These models belong to a rather general family of parametric models, the family of location-scale (LS) models. The parameters of the LS family have a direct demographic interpretation, and as such, they are easier to understand than those of the classic formulation of mortality models.

In addition, re-parameterizing mortality models in terms of the LS family has an important statistical advantage: the estimation of the LS parameters is greatly facilitated, as the rather high correlation between estimators of the classic models is significantly reduced. The lower correlation, in turn, improves parameter interpretability and reduces estimation bias.

This article is organized as follows. In Section 2.2, we provide an overview of the mathematical methods that we use throughout this article. In particular, we first present the family of LS models, and we show that several parametric models belong to the family. In addition, we describe the data that we employ in this article, and the estimation procedure for the models' parameters. In Section 2.3, we present two illustrations that demonstrate the advantages of the LS family over the classic parameterization of mortality models. First, we assess the shifting and compression dynamics of mortality changes directly from the estimation of the LS parameters in four high-longevity countries during the years 1960–2016 by gender. Second, we show that the estimated parameters of the LS family have a significantly lower *between* and *within* country correlation than classic models for females and males in thirty-three countries from 1960 until the most recent available year. In Section 2.4, we discuss the results and we conclude.

2.2 Methods

2.2.1 Life-table functions

Let $l(x)$ be the life-table probability of surviving from birth to age x , and $\mu(x)$ be the force of mortality at age x . Then, we have that $l(x) = l(0)e^{-\int_0^x \mu(a)da}$, where $l(0)$ is the radix of the population. Moreover, let $f(x)$ be the probability density function describing the distribution of deaths in the life-table population at age x . Then, $\int_0^\omega f(a)da = l(0)$, where ω is the highest age attained in the population. For simplicity, we let $l(0)$ be equal to one. The relationship that

exists between the distribution of deaths, the force of mortality and the survival function for a given age x is $f(x) = \mu(x)l(x)$ (Preston et al., 2001).

2.2.2 The location-scale family of mortality models

Location-scale distributions are well-known in reliability theory and lifetime data analysis. As such, we define the location-scale (LS) family of mortality models according to the literature. Specifically, let X be a continuous random variable. We say that X belongs to the LS family if we can express the probability density function $f(x)$ as:

$$f(x) = \frac{1}{c} f_{LS}\left(\frac{x-u}{c}\right), \quad x \in \mathbb{R}, \quad (2.1)$$

where $u \in \mathbb{R}$ and $c > 0$ are the location and scale parameters respectively, and $f_{LS}(\cdot)$ is a continuous function that does not depend on any unknown parameters. In particular, $f_{LS}(\cdot)$ represents the standard form of the distribution, i.e. when $u = 0$ and $c = 1$.

Similarly, we say that X belongs to the log-location-scale (LLS) family of models if we can express the probability density function $f(x)$ as:

$$f(x) = \frac{1}{cx} f_{LLS}\left(\frac{\ln(x)-u}{c}\right), \quad x > 0, \quad (2.2)$$

where the parameters u , c and $f_{LLS}(\cdot)$ are defined as in equation (2.1) (Mukhopadhyay, 2000; Lawless, 2011; Meeker and Escobar, 2014).

Equivalently, the two families can be defined in terms of the force of mortality $\mu(x)$. Specifically, for the LS family the force of mortality satisfies:

$$\mu(x) = \frac{1}{c} \mu_{LS}\left(\frac{x-u}{c}\right), \quad x \in \mathbb{R}, \quad (2.3)$$

while for the LLS family $\mu(x)$ satisfies:

$$\mu(x) = \frac{1}{cx} \mu_{LLS}\left(\frac{\ln(x)-u}{c}\right), \quad x > 0, \quad (2.4)$$

where u and c are defined as above, and $\mu_{LS}(\cdot)$ and $\mu_{LLS}(\cdot)$ represent the standard form of the force of mortality (i.e. when $u = 0$ and $c = 1$) for the two families, respectively.

The role of the location and scale parameters can be easily understood with an illustration. Figure 2.1 shows the effects of changes in the two parameters on the density function of the LS family of mortality models.

The location parameter u shifts the density function $f(x)$ along the x -axis without altering its shape. In the left panel of Figure 2.1, an increase (decrease) in u shifts the initial density to the right (left): this effect can be interpreted as a postponement (anticipation) of mortality. As such, changes in the parameter u correspond to a pure *shifting* effect, that is a shift of the density to older (younger) ages, without any shape changes.

The scale parameter c affects the variability of the density function. In the right panel of Figure 2.1, a decrease (increase) in c results in a compression (expansion) of the initial density

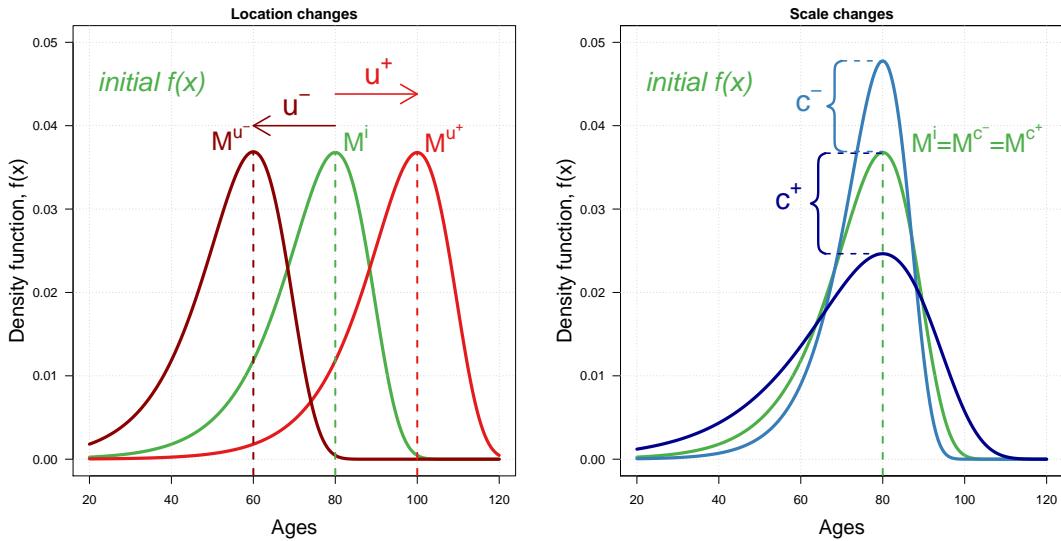


Figure 2.1: Illustration of changes in the location u (left panel) and scale c (right panel) parameters on the density function $f(x)$ of the location-scale family of mortality models. The density corresponds to the Gompertz model, and changes for u are from 80 to 60 (u^-) and 100 (u^+), and for c from 10 to 7.7 (c^-) and 15 (c^+). M denotes the old-age modal age at death.

Source (Figs. 2.1-2.2): Authors' own elaborations.

around the old-age modal age at death M (the adult age at which most of the deaths occur). As such, changes in the parameter c correspond to a pure *compression* effect.

Figure 2.2 shows the corresponding effects of these changes on the force of mortality $\mu(x)$. In the left panel, location changes correspond to shifts of the mortality pattern (which are parallel for the Gompertz case shown in the Figure); increases in u move $\mu(x)$ to the right (or down), while decreases in u result in left (or upwards) shifts. In the right panel, changes in the scale parameter modify the slope of the mortality pattern. This effect can be interpreted as a change in the “speed” of mortality: a decrease (increase) in c accelerates (slows down) the ageing process, so that mortality increases at a faster (lower) rate.

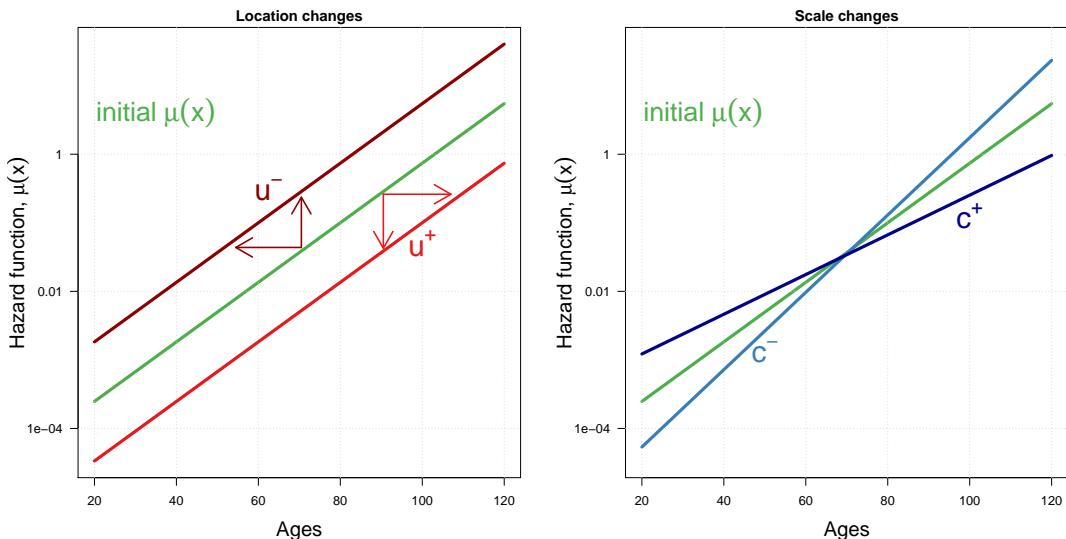


Figure 2.2: Illustration of changes in the location u (left panel) and scale c (right panel) parameters on the force of mortality $\mu(x)$ (in log scale) of the location-scale family of mortality models. The force of mortality corresponds to the Gompertz model, and changes for u and c are the same as in Fig. 2.1.

The location and scale parameters thus capture and disentangle the shifting and compression dynamics of mortality. While we have shown the two effects in isolation for illustrative purposes, mortality changes typically occur simultaneously, and the two parameters can vary at the same time.

We can now proceed to show that several mortality models used in the demographic and actuarial literature belong to the LS family of mortality models. We first show this with a detailed derivation for the Gompertz model in the next Section (and for the Weibull model in Appendix 2.5.1). Then, we present a summary Table of the mortality models that can be reconciled with the LS family in Section 2.2.4.

2.2.3 Gompertz and the location-scale family

A model that is often used in demographic and actuarial analysis is the Gompertz model. This is one of the earliest attempts to find a “law of mortality”: at the beginning of the nineteenth century, [Gompertz \(1825\)](#) discovered that for a large part of the age range (though not including infancy and youth or very old ages) the force of mortality increases with age at a steady exponential rate. The Gompertz model is generally expressed in the form:

$$\mu(x) = ae^{bx}, \quad x \geq 0, \quad (2.5)$$

where $a > 0$ denotes the level of the force of mortality at the starting age of the analysis, and b corresponds to the rate of ageing ([Thatcher et al., 1998](#)). From the life-table functions introduced in Section 2.2.1, we can derive the density function $f(x)$ of the Gompertz model:

$$f(x) = \mu(x)l(x) = \mu(x)e^{-\int_0^x \mu(a)da} = a \exp\left[bx - \frac{a}{b}(e^{bx} - 1) \right]. \quad (2.6)$$

It is possible to show that the Gompertz model is closely related to the location-scale (LS) family of mortality models. In particular, the LS-like parameterization of the Gompertz model for the force of mortality is:

$$\mu(x) = \frac{1}{c} e^{\frac{x-u}{c}}, \quad x \geq 0, \quad (2.7)$$

where $u \in \mathbb{R}$ and $c > 0$ are the location and scale parameters, respectively. The corresponding LS-like density function of the Gompertz model can be expressed as:

$$f(x) = \frac{1}{c} \exp\left[\frac{x-u}{c} - \exp\left(\frac{x-u}{c} \right) + \exp\left(-\frac{u}{c} \right) \right]. \quad (2.8)$$

Indeed, if we let the location and scale parameters be $u = \frac{1}{b} \ln\left(\frac{b}{a}\right)$ and $c = \frac{1}{b}$, and we substitute them in Equations (2.7) and (2.8), we obtain the classic Gompertz formulas in Equation (2.5) and (2.6).

The Gompertz model does not strictly belong to the LS family of mortality models: its force of mortality is only defined for $x \geq 0$, while the location-scale $\mu(x)$ can vary on the entire set \mathbb{R} (see Equation (2.3)). The truncation of the x -axis can be also observed in the functional form of $f(x)$ in Equation (2.8), which depends on the parameters u and c via the $\exp\left(e^{-\frac{u}{c}}\right)$ term.

Nevertheless, the Gompertz model is closely related to a location-scale distribution, the Gumbel (or type 1 extreme value distribution, [Johnson et al., 1995](#)) : “*the Gompertz distribution is a special case of the Gumbel distribution for the minima, i.e., when $x := -x$ and truncated at $x = 0$* ” (p. 2923, [Lenart and Missov, 2016](#)). As such, the Gompertz model can be considered a *truncated* location-scale distribution.

The location and scale parameters of the Gompertz model are reported in Table 2.1 of the following Section, together with the formulas of the classic and LS-like $\mu(x)$; the classic and LS-like functional forms of $f(x)$ are reported in Table 2.3 of Appendix 2.5.2.

Finally, it is interesting to observe that, if we let the location parameter be equal to the old-age modal age at death M , that is $u = M$, and we keep $c = \frac{1}{b}$, then Equation (2.7) becomes:

$$\mu(x) = be^{b(x-M)}, \quad (2.9)$$

which is the parameterization of the Gompertz model in terms of the modal age at death ([Horiuchi et al., 2013; Missov et al., 2015](#)).

2.2.4 Other parametric mortality models

Several models have been proposed to describe the age pattern of the force of mortality during the last two centuries (for a comprehensive review, see [Tabeau, 2001](#)), and many of them can be unified under the overarching family of location-scale models. Table 2.1 presents twelve well-known models of mortality that either belong to the location-scale (LS) and log-location-scale (LLS) families or that are closely related to them. The Table reports each model’s parameterization in terms of the classic and LS force of mortality $\mu(x)$, $\mu_{LS}(x)$ or $\mu_{LLS}(x)$, location u and scale c parameters. The classic and LS functional form $f(x)$, $f_{LS}(x)$ or $f_{LLS}(x)$ are reported in the Table 2.3 in Appendix 2.5.2.

A first interesting observation is that the Smallest Extreme-Value (Gumbel distribution for the minima) and the Gompertz model are characterized by the same hazard function. However, as already discussed in the previous Section, only the former strictly belongs to the LS family due to the truncation of the x -axis in the latter.

The Kannisto, Gamma-Gompertz, Minimal and Maximal Generalized Extreme-Value models do not strictly belong but are closely related to the LS family. First, the Kannisto model is specified by an unscaled logistic hazard function (which is equal to the Logistic model except for the $\frac{1}{c}$ term, see Table 2.1), and it has been extensively employed to smooth mortality at older ages (for example, [Wilmoth et al., 2007; Ševčíková et al., 2016](#)). Second, the Gamma-Gompertz is a three-parameter model that includes the Kannisto as a special case: its hazard function has a logistic shape whose asymptote can be different than one. Also this model has gained relevant prominence during the last decades ([Vaupel et al., 1979; Colchero et al., 2016; Missov and Lenart, 2013](#)). Finally, the Minimal and Maximal Generalized Extreme-Value models are characterized by three parameters, and for a fixed value of the third parameter (i.e. ξ), their force of mortality satisfies Equation (2.3).

The three models belonging to the LLS family (Weibull, Log-Logistic and Log-Normal) are generally used in the analysis of survival data rather than portraying the human mortality

Table 2.1: Mortality models belonging to the location-scale (LS) and log-location-scale (LLS) families, and models closely related to them, together with their parameterization in terms of the classic, LS and LLS force of mortality $\mu(x)$, $\mu_{LS}(x)$, $\mu_{LLS}(x)$, location u and scale c parameters.

Models belonging to the LS family	$\mu(x)$	$\frac{1}{c} \mu_{LS}\left(\frac{x-u}{c}\right)$	u	c
Logistic	$\frac{b \exp(a+bx)}{1+\exp(a+bx)}$	$\frac{1}{c} \frac{\exp\left(\frac{x-u}{c}\right)}{1+\exp\left(\frac{x-u}{c}\right)}$	$-\frac{a}{b}$	$\frac{1}{b}$
Normal	$\frac{1}{\sigma} \frac{\phi\left(\frac{x-\lambda}{\sigma}\right)}{1-\Phi\left(\frac{x-\lambda}{\sigma}\right)}$	$\frac{1}{c} \frac{\phi\left(\frac{x-u}{c}\right)}{1-\Phi\left(\frac{x-u}{c}\right)}$	λ	σ
Smallest Extreme-Value	$\frac{1}{\sigma} \exp\left(\frac{x-\lambda}{\sigma}\right)$	$\frac{1}{c} \exp\left(\frac{x-u}{c}\right)$	λ	σ
Largest Extreme-Value	$\frac{1}{\sigma} \frac{\exp(-\frac{x-\lambda}{\sigma})}{\exp[\exp(-\frac{x-\lambda}{\sigma})]-1}$	$\frac{1}{c} \frac{\exp(-\frac{x-u}{c})}{\exp[\exp(-\frac{x-u}{c})]-1}$	λ	σ
Models belonging to the LLS family	$\mu(x)$	$\frac{1}{cx} \mu_{LLS}\left(\frac{\ln(x)-u}{c}\right)$	u	c
Weibull	$ab(ax)^{b-1}$	$\frac{1}{cx} \exp\left(\frac{\ln(x)-u}{c}\right)$	$-\ln(a)$	$\frac{1}{b}$
Log-Logistic	$\frac{b}{x} \frac{\exp(a+b \ln(x))}{1+\exp(a+b \ln(x))}$	$\frac{1}{cx} \frac{\exp\left(\frac{\ln(x)-u}{c}\right)}{1+\exp\left(\frac{\ln(x)-u}{c}\right)}$	$-\frac{a}{b}$	$\frac{1}{b}$
Log-Normal	$\frac{1}{\sigma x} \frac{\phi\left(\frac{\ln(x)-\lambda}{\sigma}\right)}{1-\Phi\left(\frac{\ln(x)-\lambda}{\sigma}\right)}$	$\frac{1}{cx} \frac{\phi\left(\frac{\ln(x)-u}{c}\right)}{1-\Phi\left(\frac{\ln(x)-u}{c}\right)}$	λ	σ
Models related to the LS family	$\mu(x)$	LS-like $\mu(x)$	u	c
Gompertz	$a \exp(bx)$	$\frac{1}{c} \exp\left(\frac{x-u}{c}\right)$	$\frac{1}{b} \ln\left(\frac{b}{a}\right)$	$\frac{1}{b}$
Gamma-Gomp	$\frac{a \exp(bx)}{1+\frac{a}{b} \gamma[\exp(bx)-1]}$	$\frac{1}{c} \frac{\exp\left(\frac{x-u}{c}\right)}{1+\gamma[\exp\left(\frac{x-u}{c}\right)-\exp(-\frac{u}{c})]}$	$\frac{1}{b} \ln\left(\frac{b}{a}\right)$	$\frac{1}{b}$
Kannisto	$\frac{\exp(a+bx)}{1+\exp(a+bx)}$	$\frac{\exp\left(\frac{x-u}{c}\right)}{1+\exp\left(\frac{x-u}{c}\right)}$	$-\frac{a}{b}$	$\frac{1}{b}$
Minimal Generalized Extreme-Value	$\frac{1}{\sigma} \left[1 + \xi\left(-\frac{x-\lambda}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}$	$\mu(x)$	λ	σ
Maximal Generalized Extreme-Value	$\frac{1}{\sigma} \frac{\exp\left(-s^{-\frac{1}{\xi}}\right) s^{-\frac{1}{\xi}-1}}{1-\exp\left(-s^{-\frac{1}{\xi}}\right)}$ where $s = 1 + \xi\left(\frac{x-\lambda}{\sigma}\right)$	$\mu(x)$	λ	σ

Note: $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ and $\Phi(z) = \int_{-\infty}^z \phi(w) dw$ denote the probability density function and the cumulative distribution function of the standardized normal distribution, respectively.

pattern. Nevertheless, they are characterized by two important demographic properties related to the longevity and lifespan inequality of the population under their mortality pattern assumption (for additional details, see [Gigliarano et al., 2017](#)).

Finally, it is interesting to note that the three models of the LLS family have the property that the transformation $Y = \log(X)$ belongs to the location-scale class. Indeed, the Weibull, Log-Logistic and Log-Normal models for X correspond to the Smallest Extreme-Value, Logistic and Normal models for Y ([Lawless, 2011](#)).

2.2.5 Data and estimation procedure

For the illustrations shown in this article, we use available data from the [Human Mortality Database](#) (HMD, 2019). In particular, we employ death counts D_x and exposure-to-risk E_x in single years of age x for all the HMD countries that have available data starting from the year 1960 to keep the time range of the analysis sufficiently long.

This results in a subset of thirty-three countries (and thirty-nine populations): Australia, Austria, Belarus, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Estonia, Finland, France, East and West Germany, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Luxembourg, Netherlands, New Zealand (total, Maori and non-Maori population), Norway, Poland, Portugal, Russia, Slovakia, Spain, Sweden, Switzerland, UK (total population, England and Wales, Scotland and Northern Ireland), Ukraine and USA. The remaining countries of the HMD (Chile, Germany total population, Greece, Israel, Slovenia and Taiwan) are not considered due to shorter mortality series. In the illustration in Section 2.3.1, we focus only on a subset of four high-longevity countries, namely Denmark, Japan, Sweden and the USA.

It should be further noted that the data employed here are the un-smoothed deaths and person-years from the HMD. As such, model fitting occurs on actual rather than adjusted data, and goodness-of-fit depends uniquely on the ability of the model to capture observed un-adjusted mortality trends.

To estimate the parameters of a model, be it expressed in the classic or in the location-scale (LS) parameterization, we assume that death counts D_x at a given age x follow a Poisson distribution (Brillinger, 1986):

$$D_x \sim \mathcal{P}(E_x \mu_x(\boldsymbol{\theta})), \quad (2.10)$$

where $\mu_x(\boldsymbol{\theta})$ is the parametric hazard function of interest, $\boldsymbol{\theta}$ is the vector of the model's parameters (classic or LS), and E_x is the exposure-to-risk. Under this assumption, the parameters can be estimated by maximizing the following log-likelihood function:

$$\ln \mathcal{L}(\boldsymbol{\theta} | D_x, E_x) \propto \sum_{x=\alpha}^{\beta} \{D_x \ln(\mu_x(\boldsymbol{\theta})) - E_x \mu_x(\boldsymbol{\theta})\}, \quad (2.11)$$

where α and β denote the lowest and highest age groups in the analysis, respectively. The estimation of the model's parameters is carried out by maximizing Equation (2.11) in R ([R Development Core Team, 2019](#)) either with the standard general-purpose numerical optimizer `optim` or with the package `DEoptim` ([Mullen et al., 2011](#)). Routines for fitting the models of the LS family are available in the Supplementary material.

2.3 Illustrations

2.3.1 Parameter interpretation: shifting and compression

An important demographic question that can be directly studied with the location-scale (LS) family is the assessment of the shifting and compression dynamics of mortality changes. The estimation of the LS parameters over a specified time interval indeed allows to visually inspect the level and trend of each dynamic within the overall mortality development.

Here, we illustrate this property of the LS family by assessing the evolution of the two dynamics in four high-longevity countries, namely Denmark, Japan, Sweden and the USA. For each country, we fitted all the models of the LS family for adult females and males aged 30-110+ during 1960-2016, and we selected the best fitting model using the Bayesian Information Criterion (BIC, Schwarz, 1978, see Appendix 2.5.3 for additional details and computational procedure). The Minimal Generalized Extreme-Value (MinGEV) model is the best specification for both genders in the four countries (see Table 2.4 in Appendix 2.5.4 for the BIC and rankings of the different models). Figure 2.3 shows the estimated location u and scale c parameters, while the shape estimates are reported in Figure 2.9 in Appendix 2.5.4.

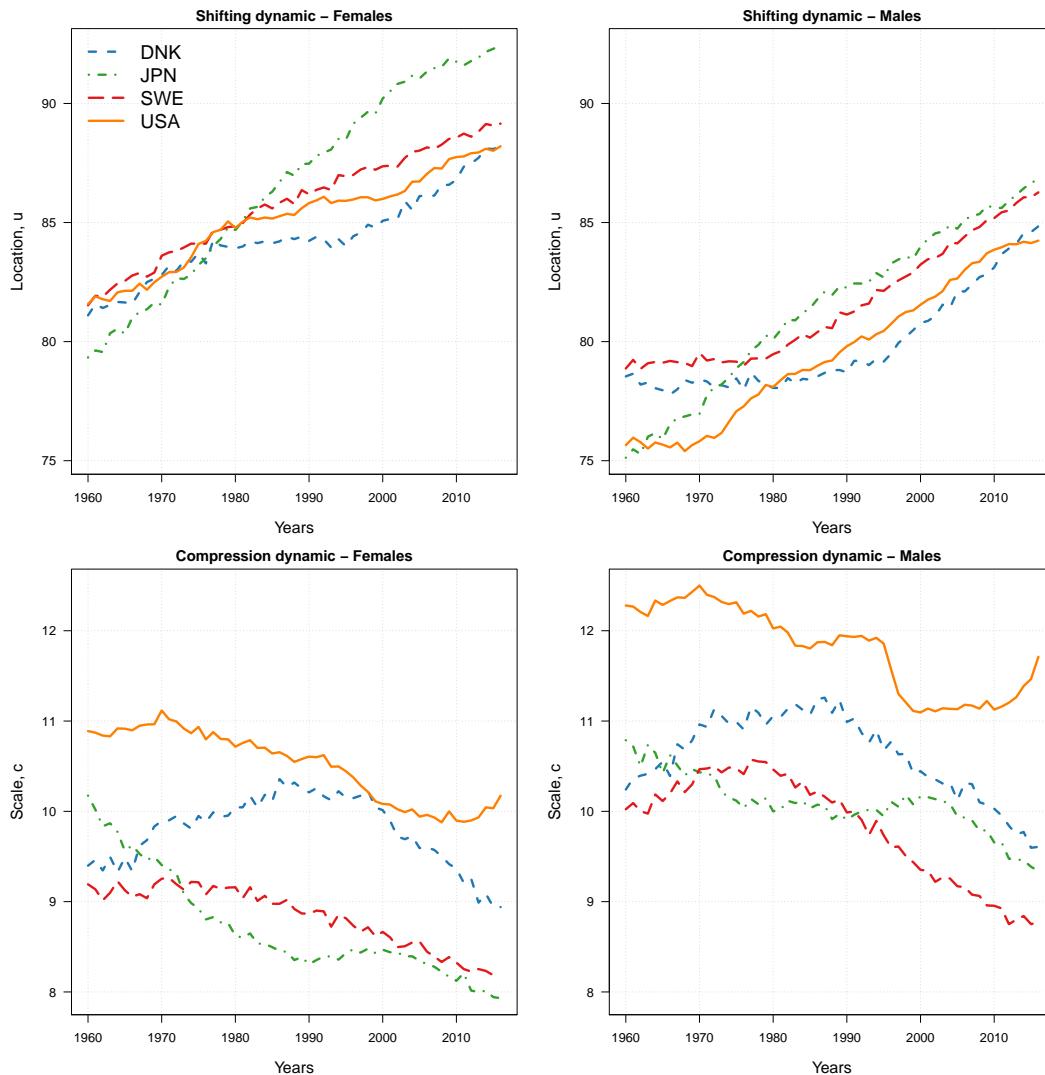


Figure 2.3: Estimated location u and scale c parameters of the Minimal Generalized Extreme-Value model for female (left) and male (right panels) adults aged 30-110+ in four high-longevity countries during 1960-2016.

Source (Figs. 2.3-2.11): Authors' calculations based on data from the [Human Mortality Database \(2019\)](#).

From the top panels of Figure 2.3, it is possible to determine the beginning of the shifting dynamic in the different countries by gender. For females, mortality started to shift before the 1960: indeed, the location parameter has been linearly increasing during all years in each country, albeit at different rates. Japanese females experienced the fastest rate of mortality

postponement, which shows some signs of deceleration in most recent years; in addition, a lack of shifting mortality for Danish females clearly emerges during 1980-1995, a period coinciding with their stagnation of life expectancy (Christensen et al., 2010; Lindahl-Jacobsen et al., 2016a,b). For males, the postponement of mortality started at different points in time in the four countries: while Japan was already experiencing this dynamic during the 1960s, the postponement started in the early 1970s in the USA, in the early 1980s in Sweden and in the early 1990s in Denmark.

From the bottom panels of Figure 2.3, the compression dynamic of mortality can be assessed (decreasing values of c correspond to compression). For females, a linear decreasing trend can be observed in Sweden and the USA, starting around the 1970s; in Japan, the decrease is very rapid until the 1990s, after which a period of stagnation occurred; and in Denmark, mortality first expanded for about twenty years at the beginning of the period (1965-1985), and then it started to compress again. For males, very different trends of compression, stagnation and expansion can be observed at different points in time.

On top of assessing changes in the two dynamics, the LS parameters readily allow a cross-country comparison on the level of the two dynamics. A first interesting observation is that both females and males in the USA are lagging behind other countries in terms of both dynamics. The higher scale parameters in the USA correspond to a greater variability of the age-at-death distribution, which translates into a higher number of “premature” deaths (i.e., deaths occurring at young adult ages). Figure 2.10 in Appendix 2.5.4 reports the estimated MinGEV age-at-death distributions in 2016 for the four countries: the share of premature deaths for USA females and males is indeed higher than for the other three countries.

Another interesting observation is that location parameters for Denmark are below those for Sweden, while the scale parameters are higher. A possible interpretation is that the different smoking behaviour of Danish and Swedish has been a major reason for this difference. Denmark is one of the few high-income countries that experienced stagnation in life expectancy in recent decades (Lindahl-Jacobsen et al., 2016b), which was paralleled by stagnation in lifespan inequality. Smoking has been indicated as the single most important factor in explaining the lower life expectancy (Sundhedsministeriet, 1994; Lindahl-Jacobsen et al., 2016a; Jacobsen et al., 2004, 2006) and the higher lifespan disparity in Denmark compared to Sweden (Aburto et al., 2018).

Furthermore, Japanese women experienced the greatest shift of mortality as well as more compression compared to the other countries. The very fast improvements of mortality since the end of the Second World War have brought them to be the “best-practice” or most longevous population worldwide since the 1980s (see Oeppen and Vaupel, 2002, Fig. 2). The Japanese age-profile of mortality has been shaped from over 35 years of best practice population, and is unique even among the closely competing countries such as South Korea and Hong Kong (Vallin and Meslé, 2016; Kontis et al., 2017).

Two important points are worth being mentioned here. First, the assessment of the shifting and compression dynamics would be the same if we had employed a different model of the LS family. Figure 2.11 in Appendix 2.5.4 shows the location u and scale c rescaled estimates for six models of the LS family fitted on Swedish adult females and males. The Figure shows that the location and scale estimates are very consistent across models: the same patterns of shifting and

compression dynamics emerge from the different LS models. Second, time trends of the location and scale parameters for the USA and Japan are in line with the empirical findings of shift and compression illustrated by [Canudas-Romo \(2008\)](#) and [Ouellette and Bourbeau \(2011\)](#): trends of the modal age at death and of the variability around (in the former) and above (in the latter) the mode closely resemble our results. Moreover, the trends of u for Swedish and Danish females are remarkably similar to the trends of their temporary life expectancy between ages 50 and 85 (see [Lindahl-Jacobsen et al., 2016a](#), Fig. 1(a)).

The analyses above, and in particular the assessment of the shifting dynamic, could not be performed based on the classic formulation of parametric models shown in the first column of Table 2.1. Figure 2.4 shows the Gamma-Gompertz (GG) estimates of the parameters u (for the GG LS model) and a (for the classic GG model) for males in the same four countries during 1960-2016. The estimates of u for the GG LS model are very close to those of the MinGEV model shown in Figure 2.3 (mean absolute percentage difference of 0.3%). In the right panel of Figure 2.4, the time trend of the parameter a reflects changes in the level of the force of mortality at the starting age of the analysis. As such, it does not allow to disentangle mortality changes due to the shifting dynamic as captured by the u parameter in the left panel. The trends of the two parameters (after reverting u or a) are indeed not comparable: in the USA, for example, the trend of a is quite erratic, while the parameter u shows an unequivocal beginning of the shifting dynamic from the 1970s.

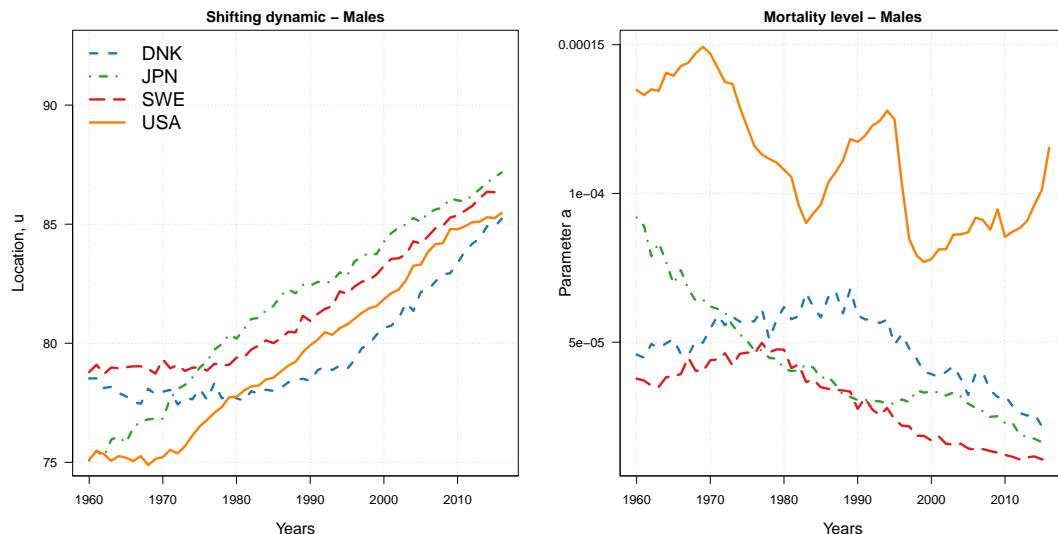


Figure 2.4: Estimated Gamma-Gompertz parameters u (LS model, left) and a (classic model, right) for male adults aged 30-110+ in four high-longevity countries during 1960-2016. Note: the y-axis of the left panel is identical to the one in the upper-right panel of Fig. 2.3 for comparison purposes.

Finally, we present a decomposition of mortality changes into shift and compression effects for the four countries. Our methodology is based and extends the decomposition method introduced by [Bergeron-Boucher et al. \(2015\)](#) to the LS-like parameterization of the Gompertz model. Appendix 2.5.5 reports the formulas and computational procedures that we have used for this analysis. Figure 2.5 shows the decomposition of changes in life expectancy at age 30 by location (shifting) and scale (compression) contributions in the Gompertz model for females in the four countries from 1960 until 2015 in 5-year intervals.

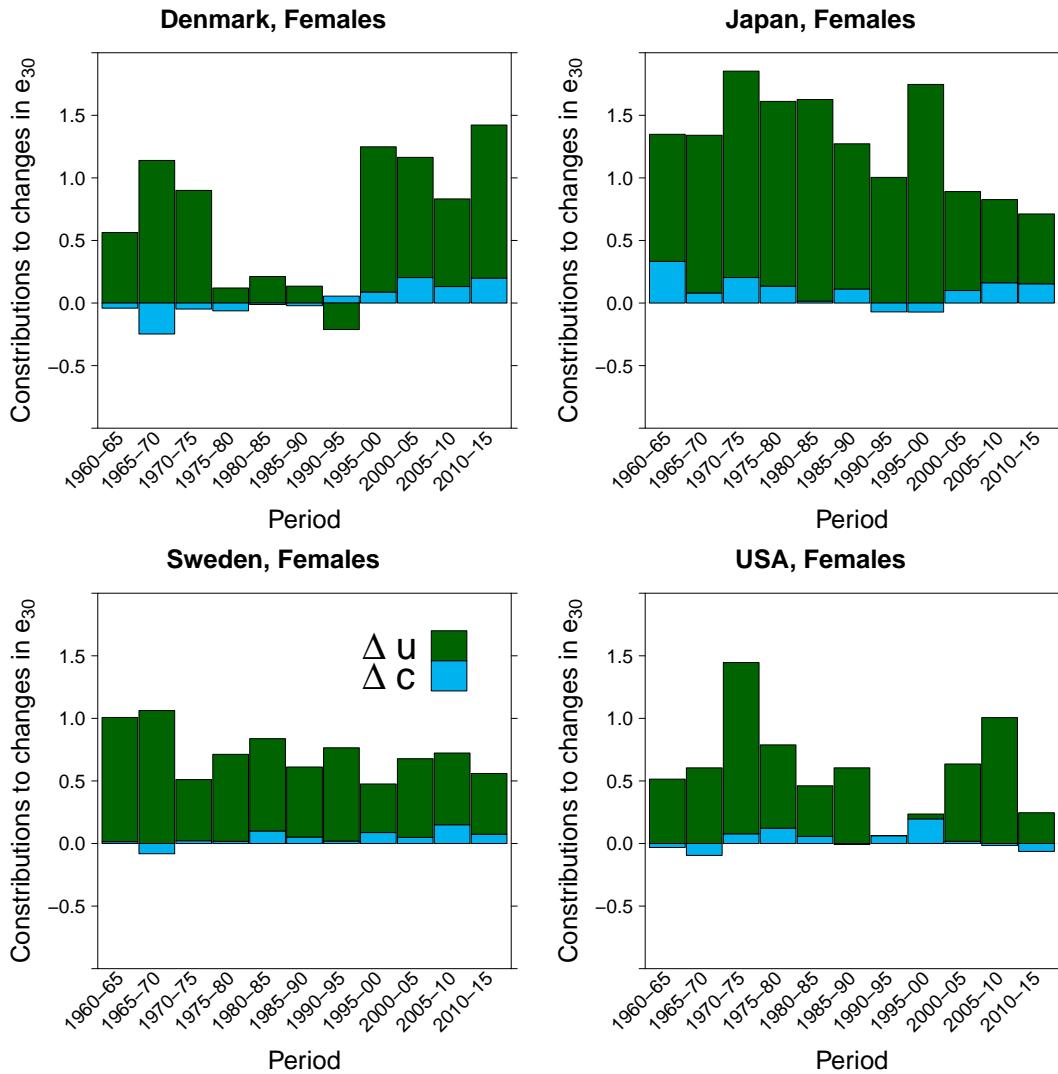


Figure 2.5: Trends over time of the contributions of the location u (green) and scale c (light blue) Gompertz parameters to changes in female life expectancy at age 30 for females in Denmark, Japan, Sweden and the USA, 1960-2015.

The Figure shows that changes in the location parameter, hence shifting mortality, are the main components driving changes in life expectancy during the period 1960-2015. Conversely, the contributions of changes in the scale parameter are limited, and even negatives in some periods. With respect to our previous discussions, the stagnation of the location parameter for Danish females between 1977 and 1995 is reflected in the small changes in life expectancy during this period. Moreover, the fast increase of the location parameter experienced by Japanese females has been the main contributor to the remarkable gains in life expectancy. These results are in line with those reported by [Bergeron-Boucher et al. \(2015, Fig. 7, Appendix C\)](#).

2.3.2 Correlation and bias of parameter estimates

It is well known that a high correlation between the estimators of model parameters is an undesirable property, as “*it shows a kind of inaccuracy of the estimators*” ([Gupta and Székely, 1994](#), p. 588). Reducing the correlation of the maximum likelihood estimators (MLE) of parametric

models is therefore a beneficial task, as it improves parameter interpretability and reduces estimation bias. In particular, [Gupta and Székely \(1994\)](#) have shown that location-scale distributions can generally be re-parameterized so that the maximum likelihood estimators are asymptotically independent.

[Missov et al. \(2015\)](#) have shown that the MLE of the parameters a and b of the classic Gompertz model in Equation (2.5) are highly (negatively) correlated. This is in general true for the parametric models in Table 2.1 that are expressed in their classic notation $\mu(x)$. Figure 2.6 shows the MLE of the typical a and b parameters of the Gamma-Gompertz, Gompertz and Kannisto mortality models for the thirty-three Human Mortality Database (HMD) countries listed above from 1960 until the most recent year by gender. Each point on the graph corresponds to a set of estimated parameters for a single country in a given year.

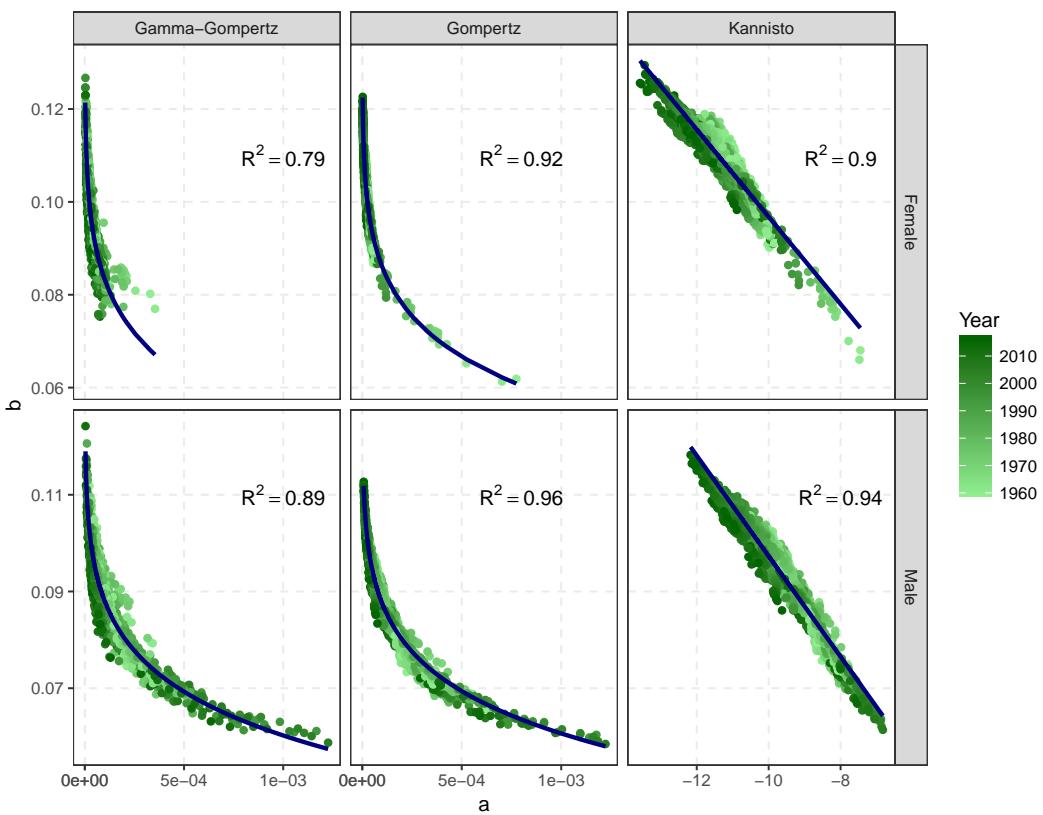


Figure 2.6: Maximum likelihood estimates and R^2 of a linear model for the classic parameters a and b for the Gamma-Gompertz, Gompertz and Kannisto model for thirty-three HMD countries from 1960 until the most recent year by gender. The a parameter was transformed in $a^{0.1}$ to linearise the relationship for the two Gompertz cases.

The estimated parameter b tends to decrease as a increases for all models, following a power function for the two Gompertz's specifications and a linear function for the Kannisto model. The variability around the two estimated parameters is very small: indeed, the R^2 of a linear model between a and b (after a suitable power transformation of the parameter a for the two Gompertz cases) is always very high (average R^2 of 0.90).

This rather high correlation between the estimators of classic parametric models can be reduced by re-parameterizing them in terms of the LS family. Figure 2.7 shows the MLE of the LS parameters u and c for the same models, countries and years by gender. From the graphs, it

is clear that the relationship between the LS estimates is much weaker than for the classic ones. The variability around the LS parameters is higher, and the R^2 of the linear regression between u and c is significantly reduced (average R^2 of 0.31). Furthermore, the relationship is not always negative as in the previous case: for the Kannisto model, increases in u tend to correspond to increases in c .

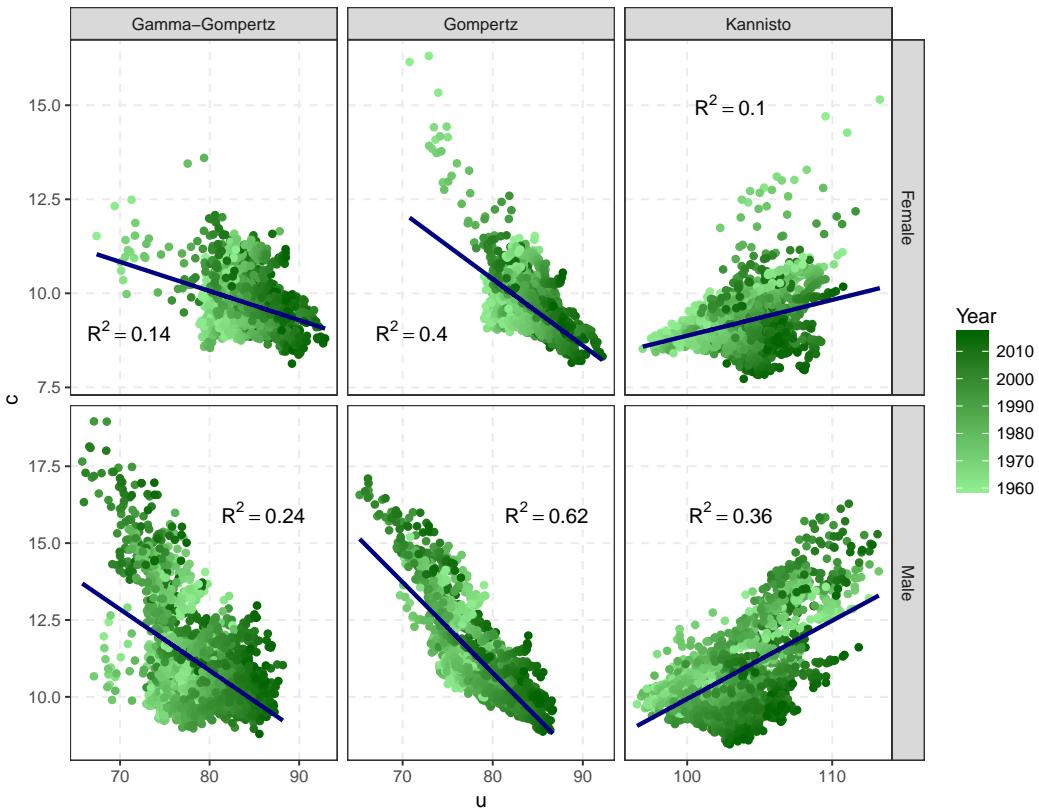


Figure 2.7: Maximum likelihood estimates and R^2 of a linear model for the LS parameters u and c for the Gamma-Gompertz, Gompertz and Kannisto model for thirty-three HMD countries from 1960 until the most recent year by gender.

The weaker relationship of the LS estimates *between* countries is generated by a lower correlation of u and c with respect to a and b *within* countries. Figure 2.8 shows the within-country absolute correlation of the two set of estimated parameters for the three-models: each point on the graph corresponds to the absolute correlation of the classic and LS parameters within one country during all the years considered in the analysis (1960 until the most recent available year) by gender. The great majority of points in Figure 2.8 fall below the diagonal, confirming that the absolute correlation between estimators of classic models' parameter can be significantly reduced by using the LS family.

Finally, we perform an experiment that demonstrates the reduction of estimation bias when using the LS family versus the classic models' parameterization. Specifically, we consider Swedish adult female mortality in 2000, and we estimate the Weibull, Logistic and Gompertz LS parameters u and c . We consider these estimates to be the “true” model parameters, and we derive the corresponding true a and b of the classic models via the formulas in Table 2.1. Then, we produce 100 simulations of age-specific Poisson death counts using the true hazard and the observed exposure times. For each simulation, we estimate the LS and classic model's parameters, and we

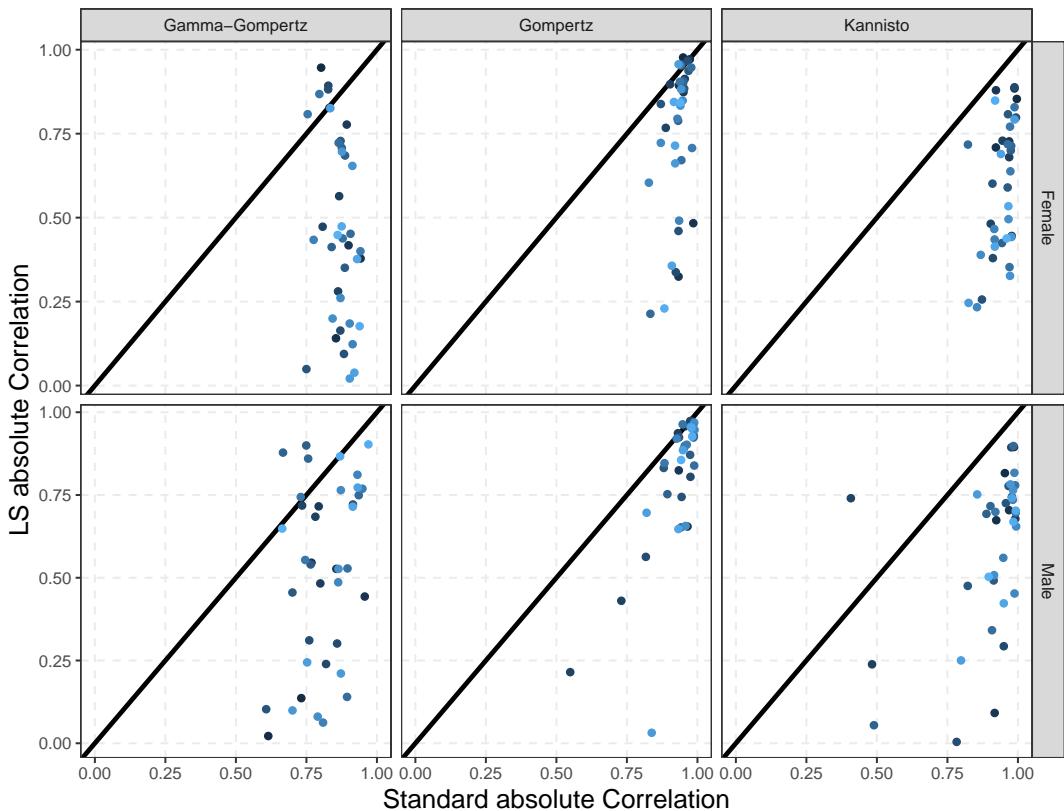


Figure 2.8: Within-country absolute correlation of the classic and LS parameterization of the Gamma-Gompertz, Gompertz and Kannisto model for thirty-three HMD countries from 1960 until the most recent year by gender.

compute the relative absolute bias (AB) for each parameter θ using the formula:

$$AB = \frac{|\hat{\theta} - \theta|}{\theta}, \quad (2.12)$$

where $\hat{\theta}$ is the estimated value of the true parameter θ (Pletcher, 1999). Table 2.2 reports the AB averaged over 100 simulations for the classic and LS parameters' estimates of the Weibull, Logistic and Gompertz models.

Table 2.2: Relative absolute bias (AB, in percentage terms) averaged over 100 simulations for the Weibull, Logistic and Gompertz classic parameters a and b , and the LS parameters u and c estimated on simulated Swedish adult female deaths in 2000.

Model	AB			
	a	b	u	c
Weibull	0.043	0.294	0.001	0.005
Logistic	0.344	0.352	0.044	0.351
Gompertz	3.048	0.327	0.037	0.326

The Table shows that the AB for the LS estimates is always lower than for the classic model parameterization: employing the LS formulas improves the precision of parameters' estimates. These results are in line with those of Missov et al., who performed a somewhat similar simu-

lation experiment for the Gompertz model and showed that “*a model misspecification leads to a relatively small bias in estimated M [i.e., u in the LS framework] in comparison to the bias in the estimated a* ” ([Missov et al., 2015](#), p. 1039).

2.4 Discussion and conclusion

In this article, we have shown that many mortality models used in the demographic and actuarial literature can be re-parameterized in terms of a general and flexible family of models, the family of location-scale (LS) models. The models belonging to this family are characterized by two parameters that have a direct demographic interpretation: the location and scale parameters capture the shifting and compression dynamics of mortality changes.

The study of compression and shifting mortality is an important issue in demographic research, as both dynamics translate differently into survival function, mortality density and hazard distributions ([Bergeron-Boucher et al., 2015](#)). Mortality compression is associated with a rectangularization of the survival curve and with decreasing variability in the distribution of ages at death ([Wilmoth and Horiuchi, 1999](#); [Cheung et al., 2005](#)). Mortality shifting is associated with a parallel shift of the force of mortality to lower ages ([Bongaarts, 2005](#)) and of the death distribution to higher ages ([Canudas-Romo, 2008](#)), while retaining their original shape. As we have shown in Section 2.3.1, the estimation of the parameters of the LS family allows an assessment of the level and trend of these two mortality dynamics.

As such, re-parameterizing classic models in their LS formulation is very useful in aiding parameter interpretability and comparability across different populations. We have shown as illustration in Section 2.3.1 the analysis of the shifting and compression dynamics of mortality changes in four high-longevity countries by gender during 1960-2016: the LS estimated parameters readily allow to examine and compare the two dynamics within and between countries.

In addition, the LS parameterization has an important statistical advantage over classic models: the rather high correlation between estimators of classic models’ parameterization is significantly reduced within the LS framework, and thus statistical estimation is facilitated. The lower correlation, in turn, further improves parameter interpretability and reduces estimation bias. As an illustration, we have shown in Section 2.3.2 that the correlation of the classic Gamma-Gompertz, Gompertz and Kannisto parameters for females and males aged 30-110+ in thirty-three Human Mortality Database countries is significantly higher than the LS parameterization, both within and between countries. Moreover, we performed simulation experiments to demonstrate the reduction of estimation bias for the Weibull, Logistic and Gompertz models when using the LS framework instead of the classic models’ formulas.

Another potentially important advantage deriving from the lower correlation is that the LS parameters could be forecasted more appropriately. The main advantages of using parameterization functions for forecasting mortality are the smoothness of forecast rates across age and the interpretability of the model’s parameters ([Bell, 1997](#); [Booth and Tickle, 2008](#)). Nevertheless, forecasting via well-known laws of mortality has been very limited due to the very high correlation between parameter’s estimators ([Booth and Tickle, 2008](#)). The lower correlation of the LS estimators could thus overcome this relevant issue. Parametric forecasts of mortality obtained

with classic and the LS models are beyond the scope of the current paper and will be explored and compared in future work.

Although we have shown that several mortality models belong to the LS family, it should be noted that the family does not embrace the full spectrum of parametric models. Additive hazard models developed for adult mortality (e.g. the Makeham and the Gamma-Gompertz-Makeham models) as well as for the entire age range, such as the [Siler \(1979\)](#), the [Heligman and Pollard \(1980\)](#) and the CoDe ([de Beer and Janssen, 2016](#)) models, do not belong to the LS family.

Nevertheless, the LS family reconciles several parametric mortality models under a unique framework, it readily allows the assessment of the shifting and compression dynamics of mortality changes, and it aids statistical estimation due to lower correlation between estimates of the model parameters. In turn, the latter could be further exploited for forecasting purposes. Re-parameterizing classic demographic models in terms of the LS family therefore offers great advantages, and the family should be considered for parametric mortality analysis.

2.5 Appendix

2.5.1 Weibull and the log–location–scale family

Here, we provide a similar derivation of Section 2.2.3 to show that the Weibull model belongs to the log–location–scale (LLS) family of mortality models.

Among its various parameterization, the Weibull model can be expressed in the form:

$$\mu(x) = ab(ax)^{b-1}, \quad x > 0, \quad (2.13)$$

where $a > 0$ and $b > 0$ are parameters ([Lawless, 2011](#)). From the life-table functions introduced in Section 2.2.1, we can then derive the density function $f(x)$ of the Weibull model:

$$f(x) = \mu(x)l(x) = \frac{b}{x} (ax)^b \exp[-(ax)^b]. \quad (2.14)$$

The Weibull model can be re-parameterized in terms of the LLS family. In particular, the LLS Weibull model for the force of mortality is:

$$\mu(x) = \frac{1}{cx} \mu_{LLS}\left(\frac{\ln(x) - u}{c}\right) = \frac{1}{cx} \exp\left(\frac{\ln(x) - u}{c}\right), \quad (2.15)$$

where $u \in \mathbb{R}$ and $c > 0$ are the location and scale parameters, respectively. The corresponding LLS density function of the Weibull model can be expressed as:

$$\begin{aligned} f(x) &= \frac{1}{cx} f_{LLS}\left(\frac{\ln(x) - u}{c}\right) \\ &= \frac{1}{cx} \exp\left[\frac{\ln(x) - u}{c} - \exp\left(\frac{\ln(x) - u}{c}\right)\right]. \end{aligned} \quad (2.16)$$

Indeed, if we let the location and scale parameters be $u = -\ln(a)$ and $c = \frac{1}{b}$, and we substitute them in Equations (2.15) and (2.16), we obtain the classic Weibull formulas in Equation (2.13)

and (2.14). As such, the Weibull model belongs to the LLS family of mortality models defined in Equations (2.2) and (2.4).

2.5.2 Location-scale functional form of twelve parametric models of mortality

Table 2.3 presents the classic, location-scale (LS) and log-location-scale (LLS) functional forms $f(x)$, $f_{LS}(\cdot)$ and $f_{LLS}(\cdot)$ of the mortality models presented in Section 2.2.4.

Table 2.3: Mortality models belonging to the location-scale (LS) and log-location-scale (LLS) families, and models closely related to them, together with their parameterization in terms of the classic, LS and LLS functional forms $f(x)$, $f_{LS}(x)$ and $f_{LLS}(x)$.

Models belonging to the LS family	$f(x)$	$\frac{1}{c} f_{LS} \left(\frac{x-u}{c} \right)$
Logistic	$\frac{b \exp(a+bx)}{[1+\exp(a+bx)]^2}$	$\frac{1}{c} \frac{\exp(\frac{x-u}{c})}{[1+\exp(\frac{x-u}{c})]^2}$
Normal	$\frac{1}{\sigma} \phi\left(\frac{x-\lambda}{\sigma}\right)$	$\frac{1}{c} \phi\left(\frac{x-u}{c}\right)$
Smallest Extreme-Value	$\frac{1}{\sigma} \exp\left[\frac{x-\lambda}{\sigma} - \exp\left(\frac{x-\lambda}{\sigma}\right)\right]$	$\frac{1}{c} \exp\left[\frac{x-u}{c} - \exp\left(\frac{x-u}{c}\right)\right]$
Largest Extreme-Value	$\frac{1}{\sigma} \exp\left[-\frac{x-\lambda}{\sigma} - \exp\left(-\frac{x-\lambda}{\sigma}\right)\right]$	$\frac{1}{c} \exp\left[-\frac{x-u}{c} - \exp\left(-\frac{x-u}{c}\right)\right]$
Models belonging to the LLS family	$f(x)$	$\frac{1}{cx} f_{LLS} \left(\frac{\ln(x)-u}{c} \right)$
Weibull	$ab(ax)^{b-1} \exp[-(ax)^b]$	$\frac{1}{cx} \exp\left[\frac{\ln(x)-u}{c} - \exp\left(\frac{\ln(x)-u}{c}\right)\right]$
Log-Logistic	$\frac{b}{x} \frac{\exp(a+b \ln(x))}{[1+\exp(a+b \ln(x))]^2}$	$\frac{1}{cx} \frac{\exp(\frac{\ln(x)-u}{c})}{[1+\exp(\frac{\ln(x)-u}{c})]^2}$
Log-Normal	$\frac{1}{\sigma x} \phi\left(\frac{\ln(x)-\lambda}{\sigma}\right)$	$\frac{1}{cx} \phi\left(\frac{\ln(x)-u}{c}\right)$
Models related to the LS family	$f(x)$	LS-like $f(x)$
Gompertz	$a \exp[bx - \frac{a}{b}(e^{bx} - 1)]$	$\frac{1}{c} \exp\left[\frac{x-u}{c} - \exp\left(\frac{x-u}{c}\right) + \exp\left(-\frac{u}{c}\right)\right]$
Gamma-Gomp	$\frac{a \exp(bx)}{\{1 + \frac{a}{b} \gamma[\exp(bx)-1]\}^{\frac{1}{\gamma}+1}}$	$\frac{1}{c} \frac{\exp(\frac{x-u}{c})}{\{1 + \gamma[\exp(\frac{x-u}{c}) - \exp(-\frac{u}{c})]\}^{\frac{1}{\gamma}+1}}$
Kannisto	$\frac{\exp(a+bx)}{[1+\exp(a+bx)]^{\frac{1}{b}+1}}$	$\frac{\exp(\frac{x-u}{c})}{[1+\exp(\frac{x-u}{c})]^{1+c}}$
Minimal Generalized Extreme-Value	$\frac{1}{\sigma} \frac{\exp\left\{-\left[1+\xi\left(-\frac{x-\lambda}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}}{\left[1+\xi\left(-\frac{x-\lambda}{\sigma}\right)\right]^{\frac{1}{\xi}+1}}$	$\frac{1}{c} \frac{\exp\left\{-\left[1+\xi\left(-\frac{x-u}{c}\right)\right]^{-\frac{1}{\xi}}\right\}}{\left[1+\xi\left(-\frac{x-u}{c}\right)\right]^{\frac{1}{\xi}+1}}$
Maximal Generalized Extreme-Value	$\frac{1}{\sigma} \frac{\exp\left\{-\left[1+\xi\left(\frac{x-\lambda}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}}{\left[1+\xi\left(\frac{x-\lambda}{\sigma}\right)\right]^{\frac{1}{\xi}+1}}$	$\frac{1}{c} \frac{\exp\left\{-\left[1+\xi\left(\frac{x-u}{c}\right)\right]^{-\frac{1}{\xi}}\right\}}{\left[1+\xi\left(\frac{x-u}{c}\right)\right]^{\frac{1}{\xi}+1}}$

Note: $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$ denotes the probability density function of the standard normal distribution.

2.5.3 Derivation of the best fitting parametric model

The selection of the best fitting parametric model can generally be made along different metrics and criterion. In this article, the estimation of a model's parameter is achieved by maximum likelihood (Section 2.2.5); within a Poisson framework, the Bayesian Information Criterion (BIC, Schwarz, 1978) is therefore a natural metric to compare different models, as it provides a good trade-off between model parsimony and accuracy.

Specifically, within a Poisson framework, the Deviance is often used as a measure of discrepancy between observed and fitted data, and it is defined as:

$$\text{Dev} = 2 \sum_y \sum_x \left[D_{x,y} \ln \left(\frac{D_{x,y}}{\hat{D}_{x,y}} \right) - (D_{x,y} - \hat{D}_{x,y}) \right], \quad (2.17)$$

where $D_{x,y}$ and $\hat{D}_{x,y}$ denote the observed and fitted number of deaths at age x and year y , respectively. This is a “badness-of-fit” measure, as higher values correspond to worse models in terms of goodness-of-fit.

In the two-dimensional age and time setting, the BIC can then be computed as:

$$\text{BIC} = \text{Dev} + \ln(mn) \text{ED} \quad (2.18)$$

where m and n are the dimensions (length) of age and time, respectively. ED denotes the Effective Dimension, or total number of parameters, of a model. Lower BIC values are associated with better models, and the trade-off between accuracy and parsimony is accounted for by the two components of the BIC.

2.5.4 Sec. 2.3.1: additional results

Here, we present some additional results corresponding to the analyses of Section 2.3.1.

Table 2.4 shows the BIC and rankings of the different LS models. From the Table, it emerges that the Minimal Generalized Extreme-Value (MinGEV) model is the best specification for both genders in the four countries.

Table 2.4: BIC values (divided by 100) of nine LS models for adult females and males in Denmark, Japan, Sweden and the USA, 1960-2016.

Model	DNK		JPN		SWE		USA	
	F	M	F	M	F	M	F	M
MinGEV	106	81	2172	799	203	105	1738	1331
Ga-Go	128	84	4607	1237	295	130	3693	2318
Gomp	142	92	5240	1433	309	137	4937	3989
Kann	236	111	7182	2529	476	232	8774	5829
Weib	775	479	22265	13026	1533	1110	36699	32631
MaxGEV	1457	872	43370	23773	3373	2234	61906	37915
Logis	1951	1507	32937	25744	3457	2934	76047	60910
Norm	2546	1824	57583	37984	5268	4027	102331	71574
LargEV	8116	6986	149526	118413	16008	13829	307781	243955

Note: the models are listed by ascending order of BIC for Danish females.

Figure 2.9 shows the estimated shape ξ parameters of the MinGEV model for the four countries by sex during 1960-2016.

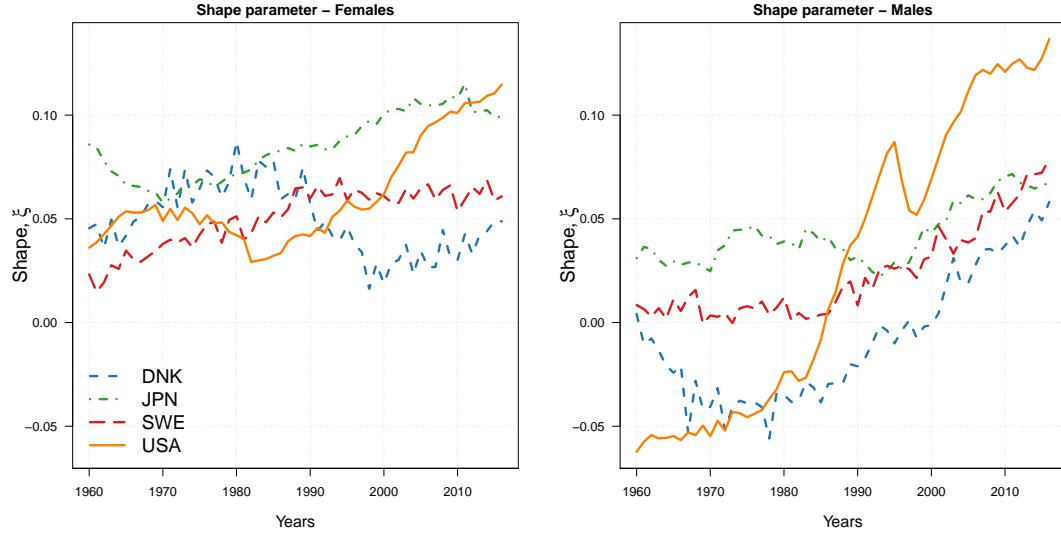


Figure 2.9: Estimated shape ξ parameters of the Minimal Generalized Extreme-Value model for female (left) and male (right panel) adults aged 30-110+ in four high-longevity countries during 1960-2016.

Figure 2.10 shows the estimated MinGEV age-at-death distributions for the four countries by sex in 2016. From the Figure, it is possible to observe that the share of premature deaths for USA females and males is higher than for the other three countries. In addition, the smaller compression of the USA distribution of deaths compared to the other countries clearly emerges from the two graphs.

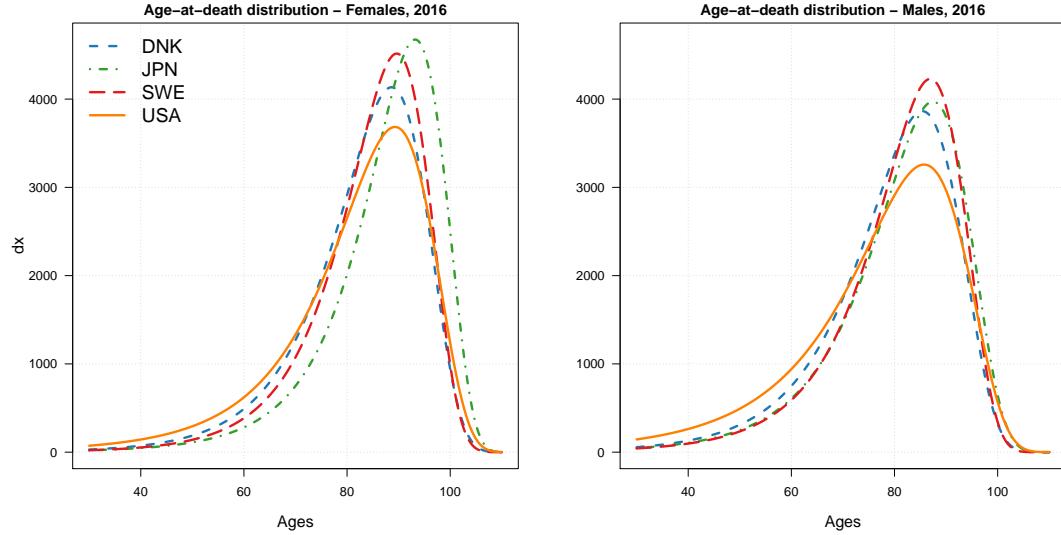


Figure 2.10: Age-at-death distributions in 2016 for female (left) and male (right panel) adults aged 30-110+ in four high-longevity countries corresponding to the Minimal Generalized Extreme-Value model estimates.

Figure 2.11 shows the location u and scale c estimates for six models of the LS family fitted to Swedish adult female and male mortality during 1960-2016. The parameters have been rescaled for comparability, and while here we focus on Sweden, the results are the same for the other countries.

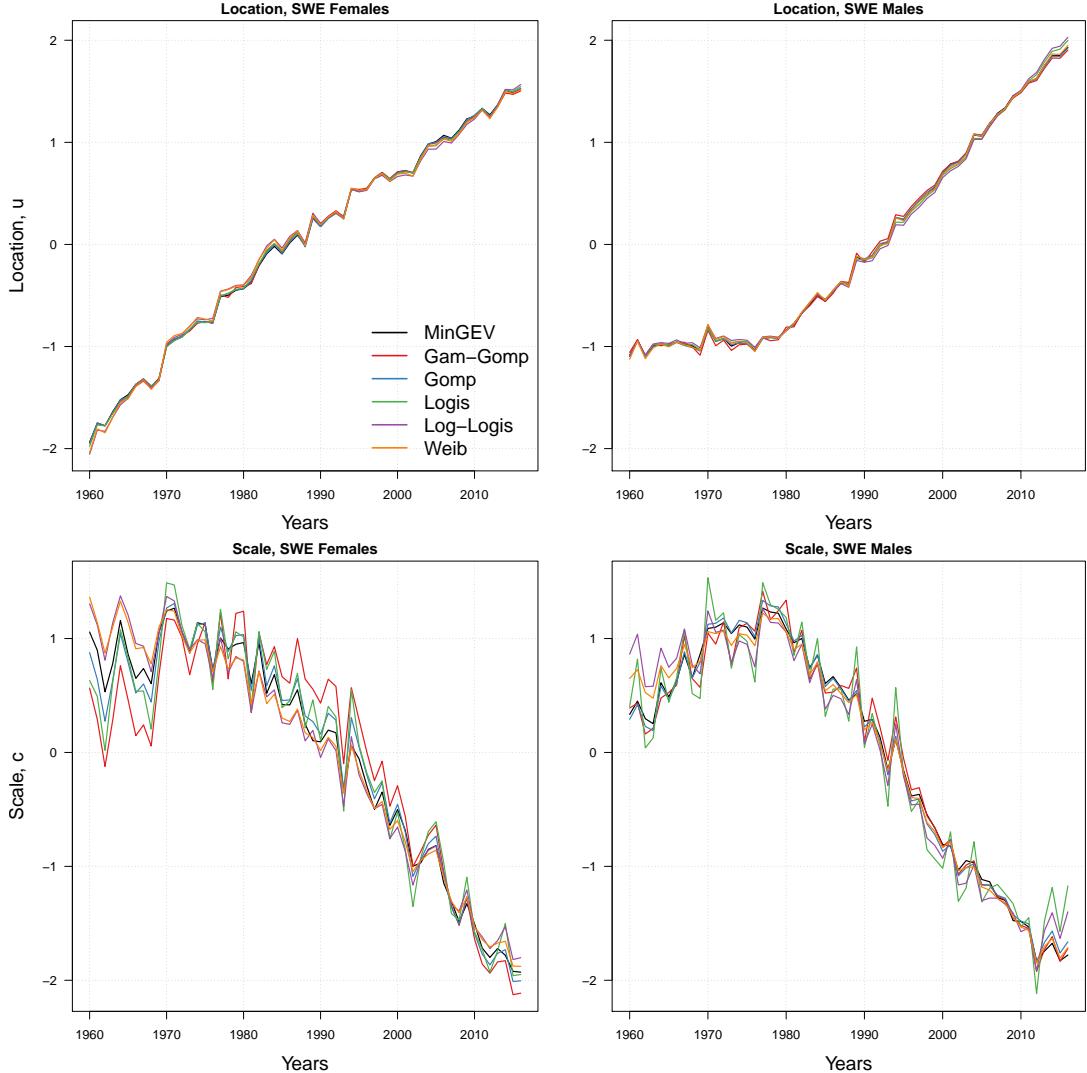


Figure 2.11: Location u and scale c rescaled estimates of six LS models for female and male adults aged 30-110+ in Sweden during 1960-2016.

The Figure shows that the location and scale estimates are very consistent across models: the former are always extremely close to each other, as well as the latter which are characterized by a slightly higher volatility. As such, the very same patterns of shifting and compression dynamics emerge from employing different LS models due to the similarity of the models' estimates.

2.5.5 Decomposition of mortality changes into location and scale effects

Here, we decompose changes in life expectancy at age 30 ($\dot{e}_{30,t}$) into two components:

$$\dot{e}_{30,t} = \Delta u + \Delta c, \quad (2.19)$$

where Δu and Δc are the gains in life expectancy resulting from the changes in the location (shift) and scale (compression) parameters, respectively.

Taking advantage of the findings reported in Fig. 2.11, namely the consistency and comparability of the location-scale parameters across different specification of the LS family, we focus on

the decomposition of the Gompertz model. Specifically, we extend the methodology presented by Bergeron-Boucher et al. (2015) to the LS-like parameterization of the Gompertz model.

Equation (2.7) introduced the LS-like parameterization of the Gompertz model. Here, we make explicit the time dependency of the model by letting the location and scale parameters be a function of time t :

$$\mu_{x,t} = \frac{1}{c_t} e^{\frac{x-u_t}{c_t}}. \quad (2.20)$$

Let a dot on top of a variable denote its derivative with respect to time (Vaupel and Canudas-Romo, 2003). The change over time in the force of mortality ($\dot{\mu}_{x,t}$) can be decomposed into respective components of change for the location (\dot{u}_t) and scale (\dot{c}_t) parameters:

$$\begin{aligned} \dot{\mu}_{x,t} &= \dot{u}_t \left[-\frac{\mu_{x,t}}{c_t} \right] + \dot{c}_t \left[-\frac{\mu_{x,t}}{c_t} \left(1 + \frac{x-u_t}{c_t} \right) \right] \\ &= \dot{u}_t f_u(\mu_{x,t}) + \dot{c}_t f_c(\mu_{x,t}), \end{aligned} \quad (2.21)$$

where $f_u(\mu_{x,t})$ and $f_c(\mu_{x,t})$ are weighting function of the hazard rate for the location and scale parameters, respectively.

Similarly to the force of mortality, we can derive the time change of life expectancy. Specifically, life expectancy at age 30 can be expressed as:

$$e_{30,t} = \int_{30}^{\omega} l_{a,t} d_a, \quad (2.22)$$

where $l_{a,t}$ is the survival function at age a and time t . Changes in life expectancy at age 30 ($\dot{e}_{30,t}$) can thus be written as:

$$\dot{e}_{30,t} = \int_{30}^{\omega} \dot{l}_{a,t} d_a = - \int_{30}^{\omega} l_{a,t} \int_{30}^a \dot{\mu}_{x,t} d_x d_a, \quad (2.23)$$

where $\dot{l}_{a,t}$ is the time derivative of the survival function. If we substitute Eq. (2.21) into Eq. (2.23), we can decompose the changes in life expectancy at age 30 ($\dot{e}_{30,t}$) into changes due to the location and scale parameters as:

$$\dot{e}_{30,t} = \underbrace{\dot{u}_t \int_{30}^{\omega} l_{a,t} \int_{30}^a f_u(\mu_{x,t}) d_x d_a}_{\Delta u} + \underbrace{\dot{c}_t \int_{30}^{\omega} l_{a,t} \int_{30}^a f_c(\mu_{x,t}) d_x d_a}_{\Delta c}. \quad (2.24)$$

The first term in Eq. (2.24) represents the gain in life expectancy resulting from a change in location (Δu), corresponding to a shifting pattern, while the second term is the gain in life expectancy produced by a change in variability (Δc), indicating a compression pattern. These are the equivalent terms of Eq. (2.19) in the Gompertz model. Specifically, we employ discrete approximations to estimate derivatives such as those in Eq. (2.24) (see Bergeron-Boucher et al., 2015, Appendix B).

Chapter 3

Modelling and forecasting adult age-at-death distributions

Ugo Filippo Basellini

Carlo Giovanni Camarda

Population Studies, **73**(1), 119–138 (2019).

DOI: 10.1080/00324728.2018.1545918

Modelling and forecasting adult age-at-death distributions

Ugofilippo Basellini^{1,2,†} and Carlo Giovanni Camarda¹

¹*Institut national d'études démographiques (INED), Paris, France*

²*Interdisciplinary Centre on Population Dynamics (CPop) and Department of Public Health,
University of Southern Denmark, Odense, Denmark*

†Corresponding author: ugofilippo.basellini@ined.fr

This is a post-peer-review, pre-copyedit version of an article published in *Population Studies*.

The final authenticated version is available online at:

<https://doi.org/10.1080/00324728.2018.1545918>

Submitted: November 2017; Accepted: July 2018

Abstract

Age-at-death distributions provide an informative description of the mortality pattern of a population, but they have generally been neglected for modelling and forecasting mortality. In this article, we use the distribution of deaths to model and forecast adult mortality. In particular, we introduce a relational model that relates a fixed “standard” to a series of observed distributions by a transformation of the age axis. The proposed Segmented Transformation Age-at-death Distributions (STAD) model is parsimonious and efficient: using only three parameters, it captures and disentangles mortality developments in terms of shifting and compression dynamics of mortality. In addition, mortality forecasts can be derived from parameters’ extrapolation using time series models. We illustrate and compare our methodology with the Lee-Carter model and its variants for females in four high-longevity countries. We show that the STAD fits very well the observed mortality pattern, and that its forecasts are more accurate and optimistic than the Lee-Carter variants.

Keywords: Mortality forecasting · Mortality modelling · Relational models · Smoothing · Modal age at death · Lifespan variability · Lee-Carter variants

3.1 Introduction

The unprecedented rise in life expectancy over the last two centuries is considered to be one of the most remarkable achievements of modern times. Best-practice female life expectancy has been rising for 160 years at a steady pace of almost three months per year ([Oeppen and Vaupel, 2002](#)). In current low-mortality and economically developed countries, life expectancy at birth for both sexes increased from around 30 to 45 years in the mid-nineteenth century to about 80 years in recent periods ([Meslé and Vallin, 2011](#)).

These sharp mortality declines triggered a fast acceleration of population growth from the beginning of the nineteenth century ([Maddison, 2006](#)). Nowadays, an ageing process accompanies the rapid increase in the world's population: virtually every country in the world is experiencing growth in the number and proportion of older persons ([United Nations, 2015](#)). In the developed world, the large cohort of baby boomers is moving into post-retirement ages, and the baby-bust generations and younger cohorts are relatively small, thus creating funding problems for social security and elderly health care provision ([Booth, 2006](#)).

As the challenges linked to population growth and ageing become manifest, innovative models to accurately portray and forecast mortality are increasingly needed to support public policies, guide individual choices and further scientific research. Mortality modelling has developed into an established research topic since the first half of the eighteenth century ([Tabeau, 2001](#)). Actuaries have produced forecasts of mortality since 1924 at least, in response to the adverse financial effects of mortality improvements on life annuities and pensions ([Pollard, 1987](#)). However, it is only in the last thirty years that more sophisticated statistical methodologies to forecast mortality have been proposed and used ([Booth and Tickle, 2008](#)).

Three functions are generally used to study human mortality: the hazard, the survival and the probability density function. These functions are complementary: knowing any one of them allows to uniquely derive the other two ([Klein and Moeschberger, 2003](#)). Although they describe the same stochastic phenomenon, the vast majority of mortality analyses and forecasting models are based on age-specific mortality rates (the discrete-time equivalent of the hazard, for example [Lee and Carter, 1992](#); [Currie et al., 2004](#); [Li and Lee, 2005](#); [Cairns et al., 2006](#); [Hyndman and Ullah, 2007](#)), or on summary measures of the mortality pattern (such as life expectancy at birth, as in [Torri and Vaupel, 2012](#); [Raftery et al., 2013](#)).

One of the reasons why mortality rates have been predominantly used in modelling and forecasting is that they readily represent the change in the risk of death over age and time. The left panel of Figure 3.1 shows death rates in logarithmic scale for Japanese females in 1960 and 2010 (data obtained from the [Human Mortality Database, 2019](#)). Data have been smoothed for illustrative purposes. The three-component pattern of human mortality, theorized at least one and a half centuries ago ([Thiele, 1871](#)), clearly emerges from the graph of death rates: a decreasing hazard at juvenile ages, capturing infant and childhood mortality; a subsequent mortality hump at young adult ages, reflecting accidents and external-cause mortality; and an exponentially (linearly in logarithmic scale) increasing hazard at older ages capturing senescence. In addition, the downward shift of the mortality curve shows the general decrease in mortality at all ages over the fifty-year period.

However, the investigation of mortality rates does not provide an immediate answer to two key questions in mortality studies: (Q1) How long does a population live on average? (Q2) How variable are ages at death? On one hand, life expectancy at birth has been the most widely used measure of longevity, although in current low-mortality countries, the adult modal age at death is claimed to be a valid alternative indicator of longevity (Kannisto, 2001; Horiuchi et al., 2013). On the other hand, stimulated by the *compression-rectangularization* hypothesis proposed by Fries (1980), the variability in ages at death has received considerable attention in recent decades as a natural complement to the average duration of life (Myers and Manton, 1984; Rothenberg et al., 1991; Wilmoth and Horiuchi, 1999; Shkolnikov et al., 2003; Kannisto, 2000). This variability is a form of lifespan inequality intrinsic to a population, with higher variation reflecting greater uncertainty in the expected time of death from an individual standpoint and greater heterogeneity in health from a population perspective (Edwards and Tuljapurkar, 2005; Edwards, 2013; Sasson, 2016).

In addition to providing a very informative description of the mortality experience of a population, the age-at-death distribution yields readily available information on the “central longevity indicators” (mean, median and modal age at death, Cheung et al., 2005; Canudas-Romo, 2010) as well as on lifespan variability. The right panel of Figure 3.1 shows the life-table age-at-death distributions corresponding to the mortality rates in the left panel.

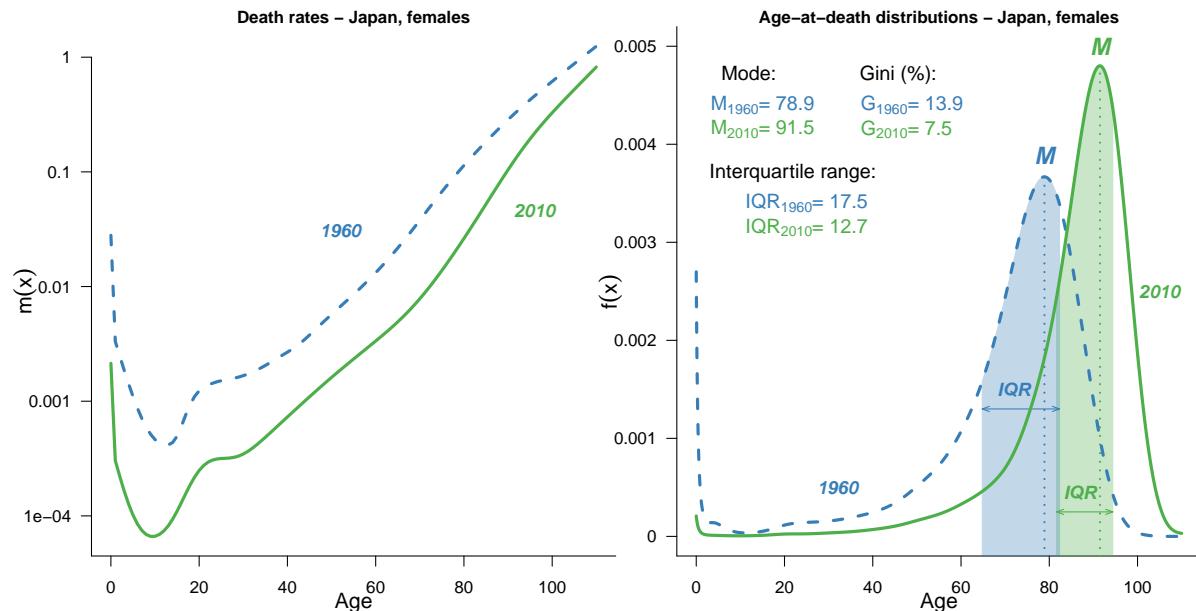


Figure 3.1: Death rates in logarithmic scale (left panel) and corresponding age-at-death distributions (right panel) for Japanese females in 1960 (blue line) and 2010 (green line). Data have been smoothed for illustrative purposes. Source (Figs. 3.1-3.8): Authors’ calculations based on data from the [Human Mortality Database \(2019\)](#).

The graph clearly shows the modal age at death (the age at which most of the deaths occur) and its significant shift to the right from 1960. Moreover, the spread of the distribution provides an indication of the lifespan variability. A decrease in variability over time can be observed directly (the distribution in 2010 is taller and more compressed than the one in 1960) and can be measured, for example with the interquartile range of life-table ages at death or the Gini coefficient (for comprehensive reviews, see Wilmoth and Horiuchi, 1999; van Raalte and Caswell,

2013). Age-at-death distributions thus provide key insights on longevity and lifespan variability that cannot be grasped directly from either mortality rates or the survival function.

Trends in longevity and lifespan variability across countries and time have been extensively investigated in the demographic literature. While almost universal agreement has been reached on the remarkable and sustained increase in longevity over the last two centuries in more economically developed countries, assessments of trends of lifespan variability depend, at least to some extent, on the measure being used (see, for example, Shkolnikov et al., 2003, pp. 314-318). A comprehensive review of the findings is beyond the scope of this paper; however, in broad terms, it has been shown that increases in longevity are associated with decreasing variability of deaths in long-lived industrialized human societies (Smits and Monden, 2009; Vaupel et al., 2011) and across the full range of human experience, for both males and females (Colchero et al., 2016).

Despite being well suited to portray mortality patterns and to study longevity and lifespan variability, age-at-death distributions have generally been neglected in modelling and forecasting. Only a few attempts have been made to explicitly use them to forecast mortality. Some efforts have been made using Compositional Data Analysis: Oeppen (2008) and Bergeron-Boucher et al. (2017) suggested a modified version of the Lee-Carter model (Lee and Carter, 1992), in which the life-table death distribution (instead of the logarithm of death rates) is forecast using Principal Component Analysis. Furthermore, Pascariu et al. (2019) introduced a vector-autoregressive time series model to forecast a number of statistical moments of the distribution of deaths.

In this paper, we propose a novel methodology for modelling and forecasting adult mortality that is based on age-at-death distributions. Our model captures mortality dynamics by transforming the age axis of a reference “standard” distribution that is fixed over time. The transformation function is characterized by three parameters that describe changes in longevity and in lifespan variability, and it takes the form of a segmented linear function. For this reason, we call our model the *Segmented Transformation Age-at-death Distributions*, or STAD, model. The total variability is divided into two components, the variability before and after the modal age at death; the model thus sheds light on the independent contribution of these two age ranges to the overall lifespan variability. The shifting and compression dynamics of mortality are captured and disentangled by the three parameters of the model, and mortality forecasts are derived from extrapolation of the parameters using time series models.

This paper is organized as follows. In Section 3.2, we review the methodology that we propose in this article. We start by introducing the STAD model in Section 3.2.1; then, we show how to choose and compute an appropriate standard distribution in Section 3.2.2. Section 3.2.3 presents the data and the smoothing procedure that we employ in our analyses, while the methods for estimating and forecasting the STAD parameters are presented in Section 3.2.4. In Section 3.3, we provide an illustration of our methodology by estimating and forecasting female mortality in four high-longevity countries, namely Sweden, Japan, France and Denmark. In Section 3.3.1, we evaluate the goodness-of-fit of the STAD model by comparing the estimated mortality pattern with the observed data, both graphically and formally. In Section 3.3.2, we assess the accuracy of the STAD forecasts by performing three out-of-sample validation exercises with forecast horizons of 10, 20 and 30 years, respectively. Finally, in Section 3.3.3, we show

the mortality forecasts of the four countries up to 2040. In each section, we use data retrieved from the [Human Mortality Database \(2019\)](#), and we compare our forecasts with those of the Lee-Carter model and its most well-known variants ([Shang et al., 2011](#)). In Section 3.4, we discuss the results and conclude.

3.2 Methods

3.2.1 The Segmented Transformation Age-at-death Distributions Model

For ease of presentation, we consider here only two adult age-at-death distributions, defined on the age range 30-110: a “standard” distribution $f(x)$, and an observed distribution $g(x)$, where x denotes age. The functional form of the distributions does not have any restrictions, as it can be parametric, nonparametric or an actual age-at-death distribution. Let $t(x; \boldsymbol{\theta})$ be a transformation function of the age axis and a set of parameters $\boldsymbol{\theta}$ such that the observed distribution conforms to the standard on the warped axis, that is:

$$g(x) = f[t(x; \boldsymbol{\theta})] \quad (3.1)$$

i.e. the distribution $g(x)$ is derived from altering the age axis of the standard distribution $f(x)$. Our aim is to find a transformation function that describes mortality developments in a parsimonious and rigorous manner. Given this purpose, we propose a transformation function $t(x; \boldsymbol{\theta})$ that depends only on (i) the changes in modal ages at death between $f(x)$ and $g(x)$; (ii) the differences in the variability of deaths of $f(x)$ and $g(x)$ before and (iii) after their modal ages.

Formally, let $s = M^g - M^f$ be the difference between the modal ages at death of $g(x)$ and $f(x)$. Then, the transformation function of the proposed *Segmented Transformation Age-at-death Distributions* (STAD) model can be expressed as follows:

$$t(x; s, b_L, b_U) = \begin{cases} M^f + b_L(x - s - M^f) & \text{if } x \leq M^g \\ M^f + b_U(x - s - M^f) & \text{if } x > M^g \end{cases} \quad (3.2)$$

where the non-negative coefficients b_L and b_U denote the change in the variability from $f(x)$ to $g(x)$ before and after their modes, respectively.

In words, the transformation function $t(x; s, b_L, b_U)$ takes the form of a segmented linear model that breaks at the value of M^g . The slopes of each linear part b_L and b_U capture the amount of expansion/reduction of variability of deaths needed before and after M^g , respectively, so that $f(x)$ fit the observed distribution $g(x)$. Changes in the mortality pattern of $g(x)$ with respect to $f(x)$ can thus be concisely described by the three parameters s , b_L and b_U . The three parameters have a very important demographic interpretation, as they directly capture the shifting and compression dynamics of mortality changes observed during the twentieth century ([Fries, 1980](#); [Wilmoth and Horiuchi, 1999](#); [Bongaarts, 2005](#); [Canudas-Romo, 2008](#)).

A schematic overview of the STAD model is useful to better explain the segmented linear transformation and the shifting/compression mortality dynamics. Figure 3.2 presents the effects of applying the segmented function $t(\cdot)$ to a given standard distribution $f(x)$ (black line of the graphs). When a simple shift is adopted (red line of the graphs), the standard distribution is

only moved to the right (as in the graph, or to the left), maintaining the same variability (as measured by the variance, if the age range is shifted too) before and after the modal age at death. This is a special case of Eq. (3.2) in which both b_L and b_U are equal to 1 and the transformation function becomes: $t(x; s) = x - s$. As such, the parameter s regulates the *shifting* dynamic of mortality.

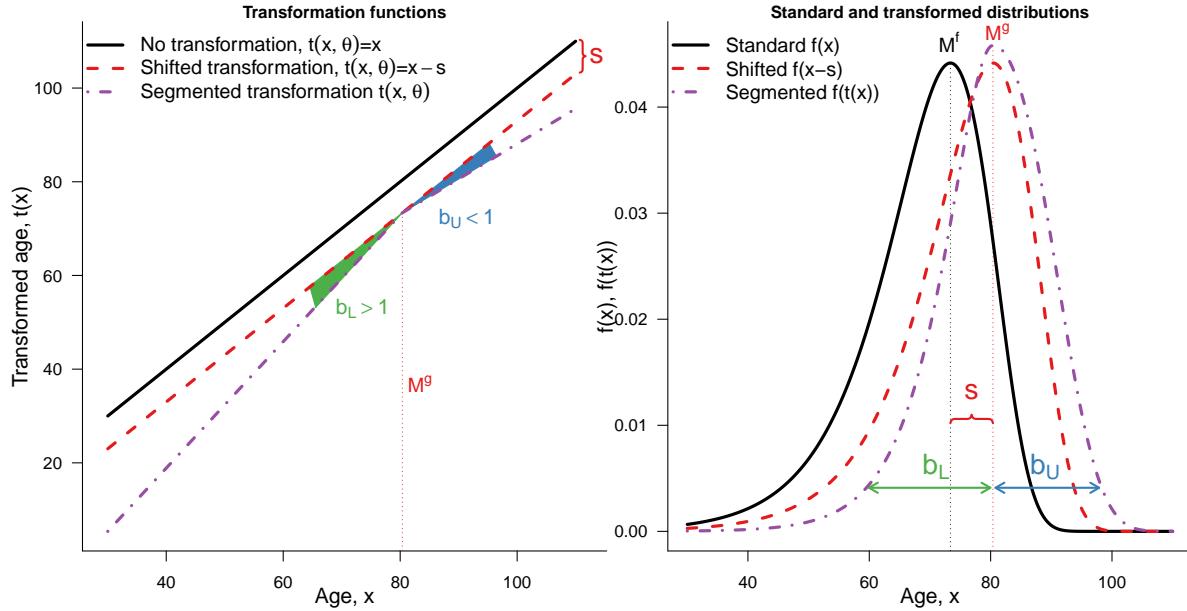


Figure 3.2: A schematic overview of the effects of transforming the age axis using the *Segmented Transformation Age-at-death Distributions* model.

More flexible scenarios can be achieved by modifying the values of b_L and b_U that act jointly with the shifting parameter s . When b_L and b_U are greater than 1, the ages before and after the mode of $g(x)$ are shrunk, and variability in ages at death is decreasing with respect to $f(x)$. On the other hand, when b_L and b_U are smaller than 1, the ages before and after M^g are expanded, so that age at death variability increases compared to the standard. These two parameters thus capture the *compression* dynamic of mortality in two age ranges, as they expand or reduce the variability of deaths before and after the mode. In the example presented in Figure 3.2, the ages before the mode are shrunk ($b_L > 1$) and the ages above are expanded ($b_U < 1$), so that the resulting variability is reduced before the mode and increased above the mode with respect to the standard (purple line of the graphs).

Interestingly, the STAD model is able to capture two limit scenarios of mortality patterns. On one hand, the complete rectangularization of the survival curve can be obtained by letting b_L and b_U go to infinity. In this case of maximum compression, the death distribution becomes a probability mass function with probability one at the modal age at death, and lifespan variability is zero as all individuals die at the same age. On the other hand, the case of maximum expansion is achieved when b_L and b_U are equal to zero. In this case, the death distribution becomes a uniform distribution, and lifespan variability is maximized.

Finally, it should be noted that despite introducing a break point at M^g in $t(\cdot)$, there is no disruption at the mode of the transformed distribution. This can be seen in the left panel of Figure 3.2: although b_L and b_U have different values, the segmented linear transformation is

continuous at the age M^g , which remains the modal age of the warped distribution depicted in purple.

3.2.2 The standard distribution

The standard distribution of the STAD model has a very important role: every observed distribution in the time period analysed can be derived from applying the transformation function $t(\cdot)$ with appropriate parameters s , b_L and b_U to the standard. As such, the standard is a “reference”, or background, distribution that should contain the representative features of the observed distributions. The STAD model can thus be thought as a relational model (Brass, 1971), and therefore choosing a suitable standard is both desirable and necessary for improving the fit of the model.

A first approach to select the standard could be to take a summary measure (such as the mean or the median) of a series of observed distributions. However, this would lead to a bias, as depicted in the left panel of Figure 3.3. Let us consider, for simplicity, only two observed distributions $g_1(x)$ and $g_2(x)$. Taking the simple mean of them results in a standard $f(x)$ with a fairly unrealistic shape that does represent neither mortality age-pattern in $g_1(x)$ nor in $g_2(x)$. The reason for this behavior is that $f(x)$ conflates age segments that are positioned before and after the mode in the two observed distributions: the decreasing part of $g_1(x)$ after its mode is averaged with the increasing part of $g_2(x)$ before its mode.

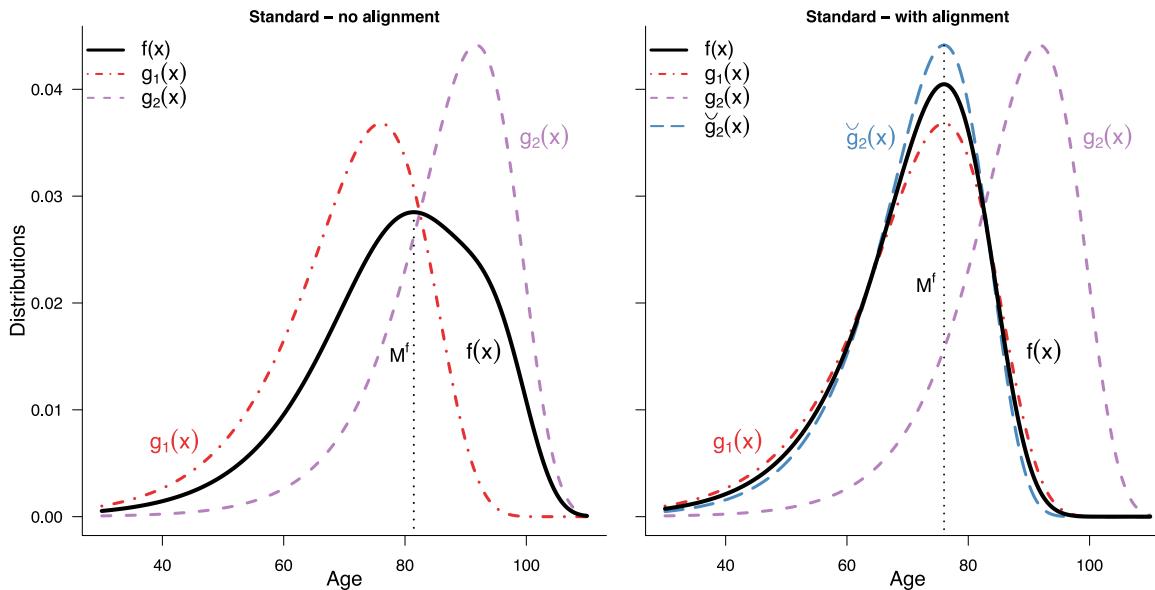


Figure 3.3: Standard distributions $f(x)$ (black line) computed as mean of the observed densities $g_1(x)$ and $g_2(x)$ (left panel) and as mean of the observed $g_1(x)$ and aligned $\check{g}_2(x)$ (right panel).

In order to avoid this problem, we first transform the observed densities to line up their qualitative features. Specifically, we shift the densities so that their modal ages at death are equal to the mode of the first observed density, maintaining their variability unchanged. This procedure is known as *landmark registration*, and is commonly used in Functional Data Analysis (Ramsay and Silverman, 2005) with the aim of “eliminating uninteresting differences in functions so that the remaining functional variation is (more) completely concerned with the differences

of interest” (Clarkson et al., 2005, p. 88). Here, by aligning the distributions we remove the differences in the mode (that are already captured by the parameter s), and we focus on the differences in their variability.

Having aligned the observed distributions, the standard distribution is then computed as the mean of the aligned distributions. In the right panel of Figure 3.3, $g_2(x)$ is shifted to the left such that its mode aligns with that of $g_1(x)$, and $f(x)$ is computed as the mean of the observed $g_1(x)$ and the aligned $\check{g}_2(x)$. This registered mean provides a better representation of the two densities: the shape of $f(x)$ is more realistic, its mode is equal to those of $g_1(x)$ and $\check{g}_2(x)$, and its variability is the average variability of the two distributions.

Finally, for capturing any type of shifting, we need to extend the age-support of the standard distribution, $f(x)$. For example, in the simple shift case in Figure 3.2, the value of the shifted distribution (red line) at age 30 is equal to the value of the expanded standard (black line) evaluated at $x - s$, i.e. beyond the original 30-110 age range. We can carry out this extension since we assume that the senescent (older adult) component of mortality is independent from the juvenile and younger adult ones. We achieve this task by means of nonparametric techniques: we extrapolate the mortality age-pattern of $f(x)$ without any rigid assumption on its form.

3.2.3 Data and estimation of the standard distribution

We apply the STAD model to the female population of four high-longevity countries, namely Sweden, Japan, France and Denmark. Whereas Sweden was selected as benchmark population for its high standard in data quality even at most advanced ages (Vaupel and Lundström, 1994; Wilmoth and Lundström, 1996), Japan and France have repeatedly been identified as countries where average lifespan is very high (Wilmoth and Lundström, 1996; Oeppen and Vaupel, 2002; Cheung and Robine, 2007; Canudas-Romo, 2008, 2010; Horiuchi et al., 2013). Finally, Denmark showed a rather peculiar mortality trajectory: among the highest longevity countries until the 1980s, this population was characterized by a stagnation of life expectancy in the last two decades of the twentieth century (Christensen et al., 2010; Lindahl-Jacobsen et al., 2016a,b). In order to test the fitting and forecasting performance of our STAD model, we apply it on these diverse mortality developments.

As already mentioned, our interest is restricted to the senescent (older adult) component of mortality. We thus start our analyses from age 30. Furthermore, we aim to forecast mortality up to 2040, hence we choose 1980-2014 as the fitting period. This allow us to study the most recent mortality developments (excluding improvements in the 1970s due to the cardiovascular revolution, Vallin and Meslé, 2004) while keeping the length of the fitting period long enough to forecast twenty-six years ahead.

Specifically, data are obtained from the Human Mortality Database (HMD, 2019) and they are observed death counts $D_{x,y}$ and exposure-to-risk $E_{x,y}$, classified by age at death, $x = 30, \dots, 110+$, and year of death, $y = 1980, \dots, 2014$. Moreover, in the following, we assume that the number of deaths at given age x and year y follows a Poisson distribution (Brillinger, 1986):

$$D_{x,y} \sim \mathcal{P}(E_{x,y} \mu_{x,y}) \quad (3.3)$$

where $\mu_{x,y}$ is the hazard function, whose estimation is the object of any mortality model.

As presented in Section 3.2.1, the STAD model is expressed in continuous notation. In order to use the model with observed discrete mortality data, a nonparametric smoothing approach is useful, so that mortality can be evaluated at any finer scale of the age axis, practically at a continuous level. In addition, the smoothing procedure avoids imposing rigid parametric mortality structures that could contribute to misleading outcomes.

The first step in the STAD model consists in extracting the standard distribution as mean of the aligned observed distributions with respect to the first modal age at death (1980 in our case). To perform this task, we need to compute age-at-death distributions and associated modes for each year. Instead of modelling directly the life-table deaths, we make use of the assumption in Eq. (3.3) and smooth actual death counts using a P -splines approach and exposures as offset (Eilers and Marx, 1996; Camarda, 2012). The advantage of this approach is twofold. First, we use an objective selection of the amount of smoothing based on the Bayesian Information Criterion (BIC, Schwarz, 1978, see Appendix 3.5.1). This approach accounts for the size of the population analysed due to the Poisson assumption of death counts (Brillinger, 1986). In contrast, smoothing life-table deaths is more subjective, because the optimal degree of smoothing depends on the life-table radix: given the same age-specific mortality pattern, higher starting radices are associated with a lower amount of smoothing. Second, we smooth observed rather than life-table deaths because the latter are constructed from adjusted death rates, as the HMD employs a logistic function for ages 80 and above (Wilmoth et al., 2007).

Having smoothed mortality, we can derive the corresponding smooth age-at-death distributions: $g(x) = \mu(x)l(x)$, where the survival function $l(x)$ can be computed numerically from the smooth death rates, evaluated at extremely fine grid. As shown in Ouellette and Bourbeau (2011), this procedure further allows us to directly extract modal ages at death for each year, M^y . Given the series of M^y , we already have the estimate of the shifting parameter for each year, i.e. the difference in modal ages at death between a given year and the first year, $s_y = M^y - M^{1980}$.

Using estimates from the P -splines approach, we can express the smooth observed distribution for year y as follows:

$$g_y(x) = \exp [\mathbf{B}(x)\boldsymbol{\beta}_y] \quad (3.4)$$

where $\mathbf{B}(x)$ is a basis of equally spaced B -splines over ages x and $\boldsymbol{\beta}_y$ are coefficients specific to the distribution $g_y(x)$. The estimated series of s_y allow us to align all distributions by simply re-evaluating the B -splines basis in Eq. (3.4) on the shifted age axes:

$$\breve{g}_y(x) = \exp [\mathbf{B}(x - s_y)\boldsymbol{\beta}_y]. \quad (3.5)$$

The standard distribution $f(x)$ can then be obtained by averaging the aligned observed distributions. Again, the standard distribution can be expressed as a linear combination of B -splines over x and new coefficients $\boldsymbol{\beta}_f$ specific to $f(x)$:

$$f(x) = \exp [\mathbf{B}(x)\boldsymbol{\beta}_f]. \quad (3.6)$$

It is noteworthy that the coefficients $\boldsymbol{\beta}_f$ contain all the representative age-specific mortality features that we will use for describing mortality developments over time, except shifting and compression/expansion that are captured by the parameters s , b_L and b_U .

3.2.4 Estimation and forecast of the STAD parameters

Given the estimated standard distribution $f(x)$, the model parameters can then be estimated. For each year y , we seek to find a set of three parameters $[s_y, b_{L,y}, b_{U,y}]$ that produces a transformation of the age axis defined in Eq. (3.2) such that the mortality pattern of the transformed standard $f(t(x; s_y, b_{L,y}, b_{U,y}))$ fits the observed data.

Whereas the shifting parameters \hat{s} are readily estimated within the alignment procedure, estimation of the compression/expansion parameters is achieved by maximum likelihood. Based on the Poisson assumption in Eq. (3.3), the parameters \hat{b}_L and \hat{b}_U are obtained by maximizing the following log-likelihood:

$$\ln \mathcal{L}(b_{L,y}, b_{U,y} | D_{x,y}, E_{x,y}, \hat{s}_y, \beta_f) \propto \sum_x \{D_{x,y} \ln(\mu_{x,y}^{\text{STAD}}) - E_{x,y} \mu_{x,y}^{\text{STAD}}\} \quad (3.7)$$

for $y = 1980, \dots, 2014$, where $\mu_{x,y}^{\text{STAD}}$ denotes the hazard function corresponding to the estimated segmented distribution of the STAD model.

On one hand, data and log-likelihood are expressed in terms of deaths, exposures and hazard. On the other hand, the STAD model is described in terms of distributions. This is not an issue because, as we already introduced in Section 3.1, hazard and distributions are complementary: knowing one of them allows to directly and uniquely derive the other. As such, the estimated \hat{b}_L and \hat{b}_U are the parameters that produce a transformed segmented distribution $f[\hat{t}(\cdot)]$ whose corresponding hazard μ^{STAD} maximizes Eq. (3.7).

A general-purpose numerical optimizer works well for maximizing Eq. (3.7) and estimate \hat{b}_L and \hat{b}_U . Routines for estimating the standard distribution as well as the transformation function for the age axis based on maximum likelihood were implemented in **R** ([R Development Core Team, 2019](#)) and are available in the supplementary material.

Having estimated the parameters over years, it is possible to model their time series: mortality forecasts are then obtained by extrapolation of the STAD parameters and the time-fixed standard distribution. In other words, we assume that mortality developments can be well approximated by a standard age-pattern and time-indexes capturing mortality dynamics in shifting and compression/expansion.

Among the numerous approaches available for time series modelling and forecasting, we combine univariate and multivariate models to describe and forecast the time evolution of the three parameters. Since the shifting and compression/expansion dynamics act independently in the STAD model, i.e. changes in s do not affect b_L and b_U and vice versa (as shown in Figure 3.2), we treat s separately from the two b parameters. For s , we choose the standard class of univariate autoregressive integrated moving average models, ARIMA(p, d, q). The three parameters of the ARIMA are chosen among different combinations, using an automatic step-wise algorithm based on the minimization of the corrected Akaike Information Criterion (the package default, see [Hyndman and Khandakar, 2008](#)).

Changes between the two b parameters could instead be correlated: for example, if a mortality change would maintain the same overall lifespan variability but compress the ages before the mode, then b_U would necessarily need to expand (or vice versa). As such, we favour a multivariate approach to model b_L and b_U , as this takes into account the interrelationship *within* and

between the two series. Consequently, for the b parameters we employ a vector autoregressive (VAR) model, which is the multivariate generalization of univariate autoregressive time series models (Hamilton, 1994). We perform model selection, estimation and diagnostic tests using the R package `vars` (Pfaff, 2008a,b). Forecasts of all three parameters are then computed with the `forecast` package (Hyndman et al., 2018a).

3.3 Results

3.3.1 Observed versus fitted data

We fit the STAD model to the female population of Sweden, Japan, France and Denmark using data described in Section 3.2.3. The estimated standard distribution and parameters \hat{s} , \hat{b}_L and \hat{b}_U allow us to derive the mortality pattern over time in each country. To assess the goodness-of-fit of the STAD, we investigate two summary measures of the mortality pattern that answer the questions outlined in Section 3.1: remaining life expectancy at age 30 (e_{30} , Q1) and the Gini coefficient at age 30 (G_{30} , Q2). The former measure is computed using standard demographic methods (Preston et al., 2001), while the latter is derived from the formula in Shkolnikov et al. (2003).

Figure 3.4 shows the observed and fitted e_{30} and G_{30} for females in the four countries over the years 1980-2014. The graphs show that the STAD model performs very well in terms of goodness-of-fit. In the top graphs, the actual and fitted e_{30} are always very close to each other, with a small loss of goodness-of-fit in some years for Danish females, probably due to their peculiar trend (Christensen et al., 2010; Lindahl-Jacobsen et al., 2016a,b). Similarly, the goodness-of-fit for G_{30} is satisfactory, with small discrepancies in the first years for Denmark and in the last years for Japan and France.

In order to perform a more formal statistical assessment of the goodness-of-fit of the STAD model beyond this graphical investigation, we compare the fit of the STAD with the standard Lee and Carter (LC, 1992) model and its variants, estimated on the same ages and years. We employ the Bayesian Information Criterion (BIC, Schwarz, 1978) to measure model differences in terms of trade-off between model parsimony and accuracy (see Appendix 3.5.1 for additional details and computational procedure).

In the following, we consider all the variants of the LC model that can be readily estimated with available packages in R. In particular, these models are: Lee and Miller (LM, 2001), Booth et al. (BMS, 2002), Brouhns et al. (BDV, 2002), Hyndman and Ullah (HU, 2007), robust Hyndman and Ullah (HUrob, 2007) and weighted Hyndman and Ullah (Huw, 2007) (for a concise review of these models, see Shang et al., 2011). The LC variants have been fitted using the R packages `demography` and `StMoMo` (Hyndman et al., 2018b; Villegas et al., 2017). For the HU, HUrob and Huw model, we chose the `demography` package default value of six principal components, as suggested by Hyndman and Booth (2008) and employed by Shang et al. (2011).

Table 3.1 shows the BIC values of the STAD and LC variants in the four countries (values of the Deviance and Effective Dimension are reported in Appendix 3.5.2). The STAD is chosen over the other models in Sweden and Denmark. For Japan, the standard Hyndman and Ullah

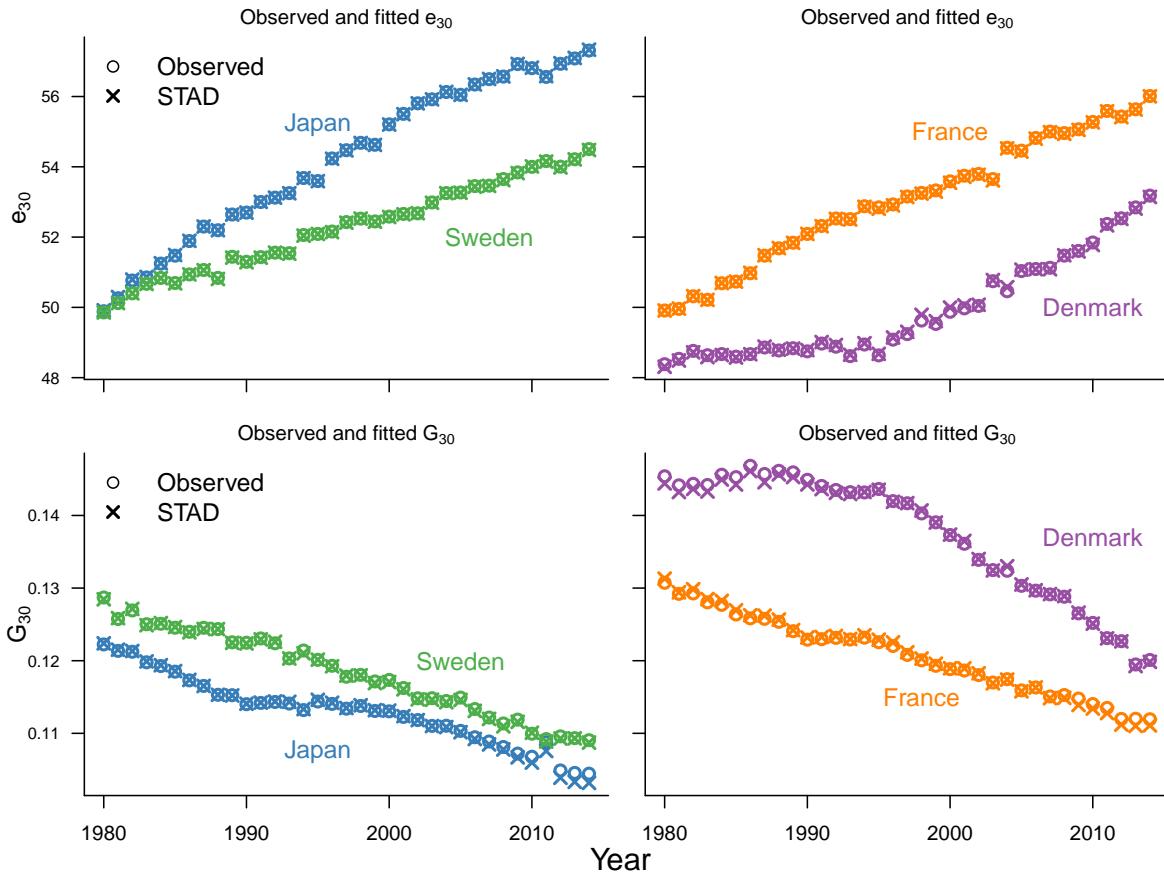


Figure 3.4: Comparison of observed (points) and STAD fitted (crosses) remaining life expectancy at age 30 (e_{30} , top graphs) and Gini coefficient at age 30 (G_{30} , bottom graphs) for females in Sweden, Japan, Denmark and France during 1980-2014.

method is the best fitting model, and the Poisson variant of the LC model (BVD) is preferred for France.

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	4276	4699	4703	4674	4611	8871	9040	9018
Japan	11053	11475	11818	11387	10790	10057	10694	10981
France	9141	7822	7951	7773	7663	9402	10132	9858
Denmark	5332	5613	5604	5587	5485	9020	9397	9890

Table 3.1: BIC values of the STAD and LC model and variants for females aged 30-110+ in four countries during 1980-2014. Lower values of the BIC (in bold) correspond to a better model.

3.3.2 Out-of-sample validation

Before projecting mortality into the future, we first assess the accuracy of the STAD forecasts by out-of-sample validation. In particular, we compare the observed versus the point and interval forecasts of the remaining life expectancy at age 30 (e_{30}), the Gini coefficient at age 30 (G_{30}), and the logged age-specific death rates over all ages and years ($\ln(m_{x,t})$). Lifespan disparity

measures such as the Gini index are in fact important indicators to evaluate mortality forecasts ([Bohk-Ewald et al., 2017](#)).

The procedure of dividing the observed dataset into two parts, the first for fitting the model, and the second for checking predictions, is very common in forecasting ([Chatfield, 2000](#)). Here, we perform three out-of-sample validation exercises: we fit the STAD and the [Lee and Carter](#) variants to three periods of equal length: 1970-2004, 1960-94 and 1950-84. Then, we forecast mortality until 2014 in all scenarios, corresponding to three forecast horizons of 10, 20 and 30 years, respectively. We then compare the point and interval forecasts of e_{30} , G_{30} and $\ln(m_{x,t})$ at each forecast year with the observed data.

For the point forecast analysis, we employ the mean absolute error (MAE) as measure of forecast accuracy, following standard time series practice ([Chatfield, 2000](#)). In addition, we only compare the forecast accuracy of the STAD model with the first four LC variants: we do not consider the functional data models (HU, HUrob and HUw) because their better performance is partially due to the much higher number of parameters employed (about four and five times more parameters than the LC variants and the STAD, respectively. See Effective Dimension in Appendix 3.5.2).

[Table 3.2](#) shows the point forecast accuracy as measured by the MAE of e_{30} , G_{30} and $\ln(m_{x,t})$ for the STAD and four LC variants in the three out-of-sample scenarios as well as for the four countries. The results of this analysis show that the STAD forecasts are overall more accurate than the LC ones. Out of 36 indicators, the STAD model is the most accurate almost half of the times (16 indicators). The LM is the second best performer (9 indicators), followed by the BMS (6 indicators), the BDV (4 indicators), and the original LC model (1 indicator).

In addition to analysing point forecasts, we further evaluate the accuracy of interval forecasts for e_{30} , G_{30} and $\ln(m_{x,t})$. Prediction intervals are a valuable tool for assessing the probabilistic uncertainty of point forecasts ([Shang et al., 2011](#)), and interval forecasts are very important as they allow, among others, a more thorough comparison of different forecasting methodologies ([Chatfield, 2000](#)).

For the LC variants, we derived the 80% prediction intervals from the uncertainty of the innovations in the random walk with drift model for the time index κ_t ; in addition, we considered the uncertainty of the drift parameter for the BMS method. For the STAD model, we derived the 80% prediction intervals from a bootstrapping procedure ([Efron and Tibshirani, 1994](#)): we computed 1000 simulations of the time series of the three parameters, and for each simulation we calculated the mortality pattern in each forecast year (the age-specific death rates and associated summary measures). From these simulations, we took the median as central forecast, and the lower and upper 10% quantiles to construct the 80% prediction intervals.

Specifically, we computed one-step-ahead 80% prediction intervals for each year (and age for $\ln(m_{x,t})$) in the forecasting period of the three out-of-sample exercises. Following [Shang et al. \(2011\)](#), we calculated the *coverage probability deviance* as the absolute difference between 0.8 (the nominal coverage probability) and the empirical coverage probability. The latter measure is given by the actual proportion of out-of-sample data that falls within the calculated prediction intervals. The lower the coverage probability, the more accurate the prediction intervals of a

Country	Test period	Measure	Model				
			STAD	LC	LM	BMS	BDV
Sweden	10y	e_{30}	0.11	0.11	0.09	0.08	0.08
		G_{30}	0.17	0.06	0.06	0.06	0.07
		$\ln(m_{x,t})$	0.10	0.12	0.12	0.12	0.11
	20y	e_{30}	0.25	0.41	0.38	0.41	0.38
		G_{30}	0.41	0.15	0.21	0.15	0.18
		$\ln(m_{x,t})$	0.14	0.16	0.18	0.16	0.16
	30y	e_{30}	0.87	0.81	0.80	0.80	0.78
		G_{30}	0.21	0.16	0.07	0.16	0.12
		$\ln(m_{x,t})$	0.15	0.19	0.18	0.19	0.18
Japan	10y	e_{30}	0.94	1.00	0.86	0.78	0.98
		G_{30}	0.11	0.55	0.14	0.07	0.40
		$\ln(m_{x,t})$	0.12	0.21	0.12	0.11	0.15
	20y	e_{30}	0.86	0.35	0.35	0.31	0.33
		G_{30}	0.74	1.10	0.78	1.07	1.04
		$\ln(m_{x,t})$	0.17	0.25	0.20	0.25	0.22
	30y	e_{30}	1.18	0.28	0.23	0.45	0.32
		G_{30}	1.06	1.68	1.25	1.56	1.40
		$\ln(m_{x,t})$	0.22	0.46	0.31	0.44	0.40
France	10y	e_{30}	0.13	0.44	0.30	0.35	0.36
		G_{30}	0.08	0.42	0.10	0.19	0.35
		$\ln(m_{x,t})$	0.07	0.12	0.07	0.10	0.11
	20y	e_{30}	0.70	0.55	0.33	0.47	0.47
		G_{30}	0.37	0.58	0.17	0.54	0.51
		$\ln(m_{x,t})$	0.13	0.16	0.13	0.16	0.12
	30y	e_{30}	0.21	0.38	0.36	0.38	0.38
		G_{30}	0.34	0.45	0.36	0.45	0.45
		$\ln(m_{x,t})$	0.12	0.16	0.15	0.16	0.15
Denmark	10y	e_{30}	0.75	1.26	1.01	1.57	1.45
		G_{30}	0.75	1.01	0.55	1.12	1.15
		$\ln(m_{x,t})$	0.16	0.20	0.18	0.22	0.21
	20y	e_{30}	1.58	1.06	1.01	1.10	1.12
		G_{30}	1.92	1.39	1.28	1.39	1.43
		$\ln(m_{x,t})$	0.29	0.26	0.23	0.25	0.25
	30y	e_{30}	0.66	0.72	0.64	0.74	0.73
		G_{30}	1.21	0.72	0.82	0.72	0.71
		$\ln(m_{x,t})$	0.27	0.26	0.28	0.24	0.25

Table 3.2: Mean absolute error of the STAD and Lee-Carter variants forecasts of e_{30} , G_{30} and $\ln(m_{x,t})$ for females in four countries and in three out-of-sample validation exercises: forecast horizon of 10 years (fitting period 1970-2004), 20 years (1960-94) and 30 years (1950-84). Lower values correspond to greater point forecast accuracy.

Note: in case of equal values, the ranking was decided from the third decimal place.

model; furthermore, with a 0.8 nominal coverage probability, the deviance can vary from 0 to 0.8.

Table 3.3 shows the coverage probability deviances of e_{30} , G_{30} and $\ln(m_{x,t})$ for the STAD and four LC variants in the three out-of-sample scenarios as well as for the four countries. Here, the STAD model is the second best performer. The LM outperforms the other models, as it produces more accurate prediction intervals for 10 indicators; the STAD, LC and BMS models

follow with 5 indicators, while the BDV model is most precise for only 3 of them. For the remaining 8 indicators there was a perfect draw among two or more models.

Country	Test period	Measure	Model				
			STAD	LC	LM	BMS	BDV
Sweden	10y	e_{30}	0.20	0.20	0.20	0.20	0.20
		G_{30}	0.70	0.10	0.10	0.10	0.20
		$\ln(m_{x,t})$	0.43	0.28	0.41	0.31	0.33
	20y	e_{30}	0.20	0.20	0.20	0.20	0.20
		G_{30}	0.75	0.20	0.10	0.20	0.15
		$\ln(m_{x,t})$	0.41	0.32	0.43	0.30	0.37
	30y	e_{30}	0.73	0.03	0.07	0.13	0.03
		G_{30}	0.43	0.03	0.20	0.03	0.00
		$\ln(m_{x,t})$	0.57	0.41	0.45	0.35	0.42
Japan	10y	e_{30}	0.80	0.80	0.50	0.40	0.80
		G_{30}	0.30	0.80	0.10	0.10	0.80
		$\ln(m_{x,t})$	0.66	0.67	0.54	0.40	0.68
	20y	e_{30}	0.50	0.10	0.00	0.15	0.10
		G_{30}	0.80	0.80	0.80	0.80	0.80
		$\ln(m_{x,t})$	0.61	0.61	0.51	0.60	0.60
	30y	e_{30}	0.07	0.20	0.20	0.20	0.20
		G_{30}	0.67	0.80	0.60	0.80	0.80
		$\ln(m_{x,t})$	0.48	0.65	0.53	0.65	0.65
France	10y	e_{30}	0.10	0.10	0.10	0.20	0.10
		G_{30}	0.40	0.80	0.20	0.10	0.70
		$\ln(m_{x,t})$	0.44	0.35	0.32	0.33	0.36
	20y	e_{30}	0.60	0.15	0.20	0.20	0.20
		G_{30}	0.80	0.70	0.20	0.30	0.60
		$\ln(m_{x,t})$	0.42	0.29	0.20	0.23	0.25
	30y	e_{30}	0.00	0.20	0.20	0.20	0.20
		G_{30}	0.60	0.20	0.20	0.20	0.20
		$\ln(m_{x,t})$	0.51	0.03	0.02	0.07	0.01
Denmark	10y	e_{30}	0.20	0.70	0.50	0.80	0.80
		G_{30}	0.80	0.80	0.80	0.80	0.80
		$\ln(m_{x,t})$	0.42	0.35	0.38	0.43	0.48
	20y	e_{30}	0.45	0.40	0.30	0.40	0.40
		G_{30}	0.80	0.80	0.75	0.80	0.80
		$\ln(m_{x,t})$	0.43	0.42	0.39	0.43	0.43
	30y	e_{30}	0.23	0.03	0.13	0.07	0.07
		G_{30}	0.57	0.73	0.70	0.73	0.73
		$\ln(m_{x,t})$	0.63	0.43	0.46	0.38	0.45

Table 3.3: Coverage probability deviances of the STAD and Lee-Carter variants forecasts of e_{30} , G_{30} and $\ln(m_{x,t})$ for females in four countries and in three out-of-sample validation exercises: forecast horizon of 10 years (fitting period 1970-2004), 20 years (1960-94) and 30 years (1950-84). Lower values correspond to greater interval forecast accuracy.

Note: in case of equal values, the ranking was decided from the third decimal place (when possible).

3.3.3 Forecasts to 2040

We now present the mortality forecasts of the STAD model from 2015 to 2040 using the fitting period of Section 3.3.1, and we compare the forecasts with two Lee and Carter (LC, 1992)

variants that best performed in the previous Sections, namely the [Hyndman and Ullah \(HU, 2007\)](#) method (lowest BIC for Japan) and the [Lee and Miller \(LM, 2001\)](#) model (most accurate forecasts after the STAD).

Figure 3.5 shows the estimated and forecast STAD parameters with 80% prediction intervals for females in Sweden, Japan, France and Denmark during the years 1980-2040. In all countries, the observed shifting dynamic is forecast to continue, with the modal age at death increasing by about 1.3 years per decade in Sweden, 2.1 in Japan, 1.7 in France and 1.1 in Denmark. The lifespan variability before and after the modal age at death is forecast to decrease in all countries (increasing values of b correspond to greater compression), with b_U increasing at a faster pace than b_L in all countries except Denmark.

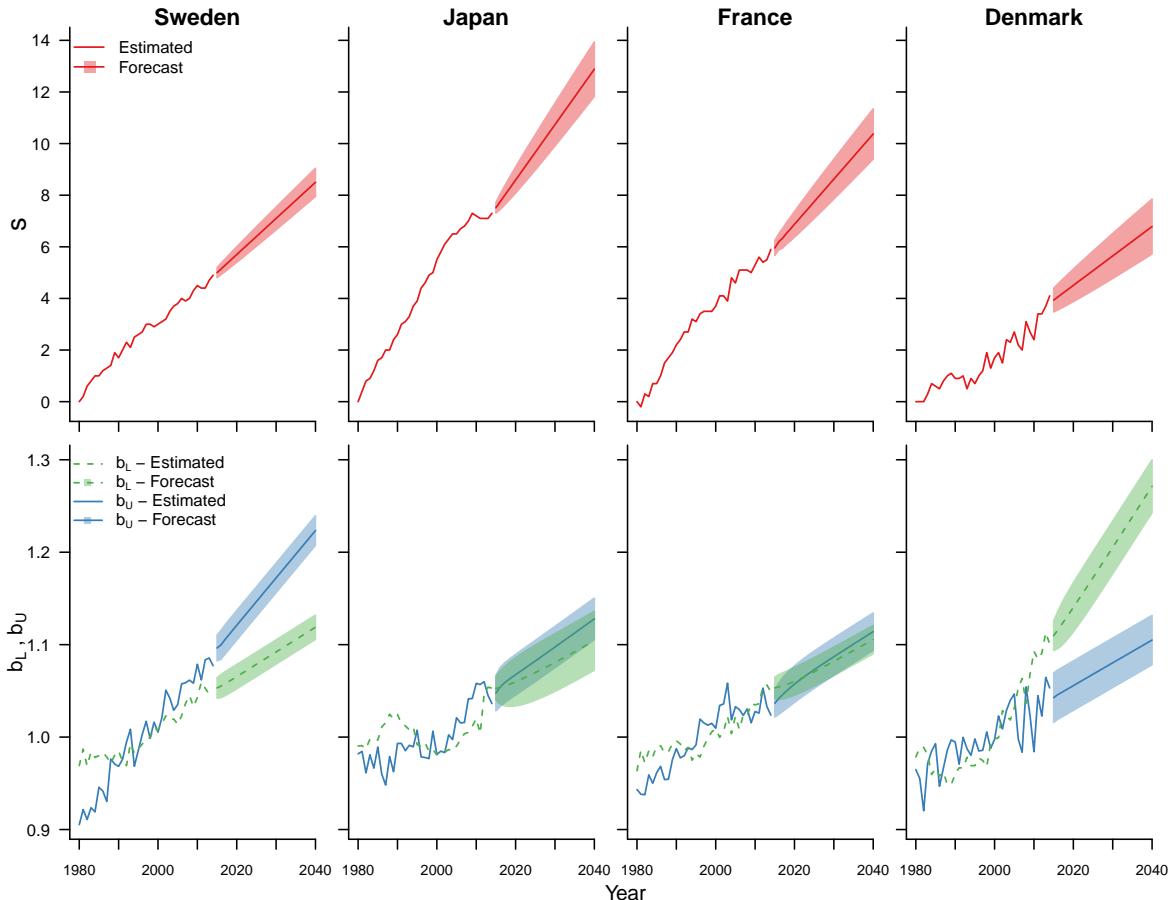


Figure 3.5: Estimated and forecast STAD parameters s (red line, top panels), b_L and b_U (green and blue line, respectively, bottom panels) with 80% prediction intervals for females in Sweden, Japan, France and Denmark during the years 1980-2040.

Figure 3.6 shows the observed and forecast remaining life expectancies at age 30 (e_{30}) and Gini coefficients at age 30 (G_{30}) with 80% prediction intervals in the four countries for the years 1980-2040, together with the forecasts of the LM and HU models. For the STAD, the 80% prediction intervals around the forecasts are derived from a bootstrapping procedure ([Efron and Tibshirani, 1994](#)).

The graphs of Figure 3.6 show that the STAD forecasts of e_{30} reflect well the past linear increase over the years 1980-2014, and that they are more optimistic than the LM and HU forecasts. In all countries, the slope of the future increase in e_{30} of the STAD is larger than

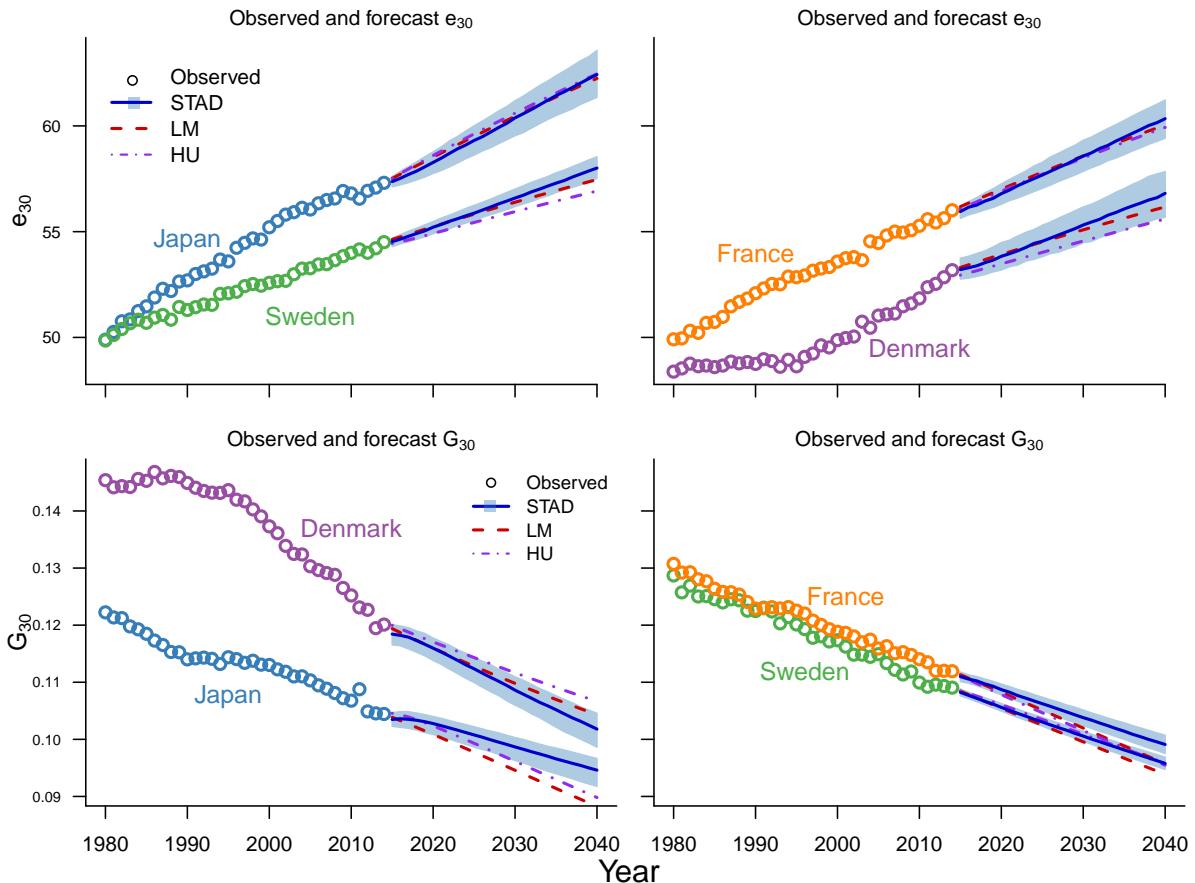


Figure 3.6: Observed and forecast remaining life expectancies at age 30 (e_{30} , top panels) and Gini coefficients at age 30 (G_{30} , bottom panels) for Swedish, Japanese, French and Danish females in 1980–2040 for the STAD (blue solid line and shading for the 80% prediction intervals), LM (red dashed line) and HU (purple dashed line with dots) models.

those of the LM and HU model (with the exception of the HU forecast in Denmark). In Japan and France, the initial forecasts of the STAD are lower than the LC ones, but the steeper rate of increase of the STAD results in higher forecasts by 2040. With respect to G_{30} , the STAD forecasts generally show a slower compression of mortality than the LM and HU, most noticeably in Japan.

In Figure 3.7, we compare the STAD, LM and HU age-specific mortality rates forecasts in 2040 for the four countries. Several differences emerge between the models from this age-pattern analysis. Mortality rates of the STAD model are smoother than the LM ones, as they are derived from a smooth standard distribution. This could be particularly advantageous for long-term projections. Discontinuities or jaggedness in the forecast age profile of mortality will be absent in the STAD forecasts, unlike in the LM model (Li et al., 2013). Importantly, the STAD forecasts do not display the fairly unrealistic S-shape that characterizes the LM and HU forecasts for Sweden, Japan and France. In addition, the HU forecast for Denmark display an even more undulating pattern.

Finally, Figure 3.8 shows the actual life-table deaths in 1980 and 2014, along with the age-at-death distributions in 2040 as computed from the STAD, LM and HU forecasts. The shifting mortality dynamic is more important and visible in the STAD model: in fact, the age-at-death

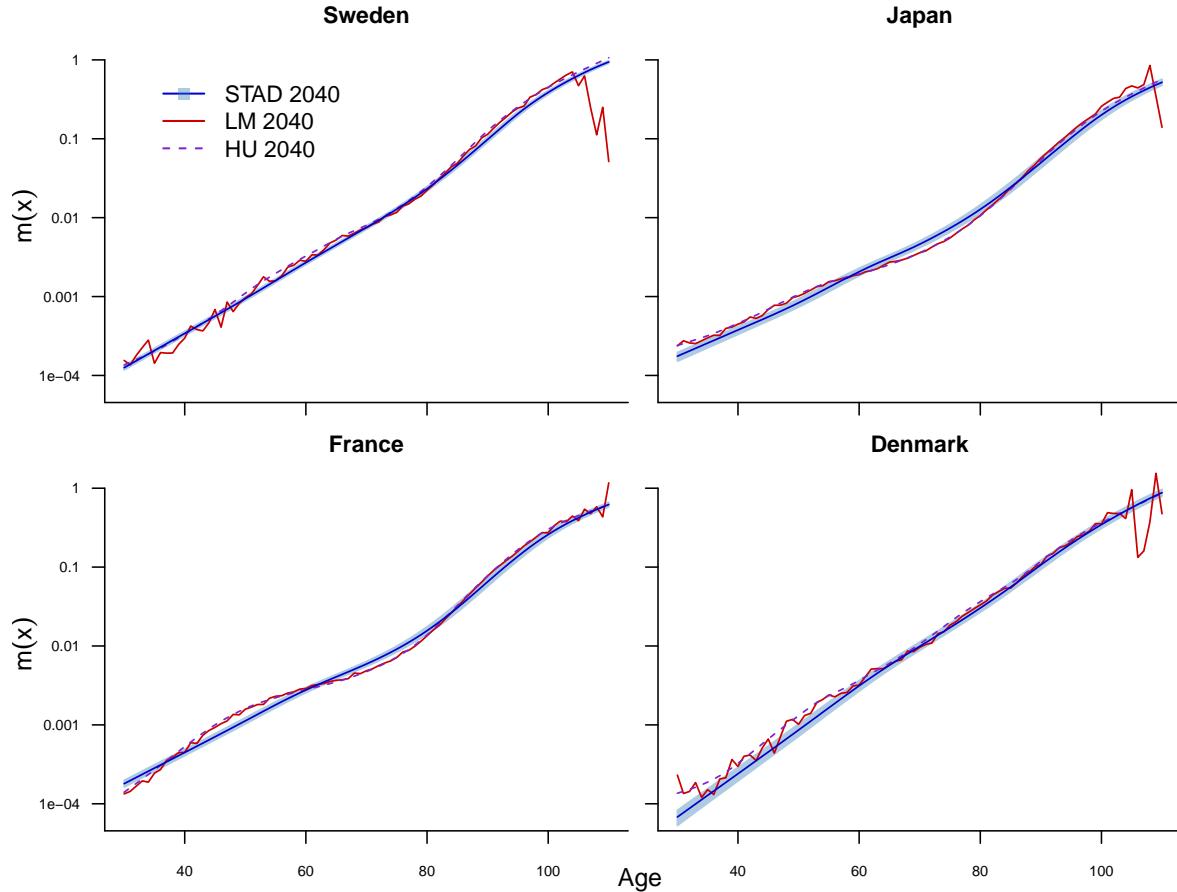


Figure 3.7: Forecast mortality rates in 2040 with 80% prediction intervals (in logarithmic scale) of the STAD (blue line and shading), Lee and Miller (LM, red solid line) and Hyndman and Ullah (HU, purple dashed line) models for Swedish, Japanese, French and Danish females.

distribution forecast is more shifted and less compressed than the LM and HU ones in all countries except Denmark (where the forecasts are similar).

3.4 Discussion and conclusion

Age-at-death distributions provide a very informative description of the mortality pattern of a population, and they are well suited to study two important indicators of population health, namely longevity and lifespan variation. Despite these advantages, they have generally been neglected in modelling and forecasting human mortality, as most techniques are based on the complementary age-specific mortality rates.

In this paper, we propose a novel methodology for modelling and forecasting adult mortality based on age-at-death distributions. To our knowledge, this is one of the very first attempts in this direction. In particular, we introduce a segmented linear transformation model that captures mortality dynamics over time from changes in modal ages at death and variability of the distributions with respect to a fixed standard. This approach is both parsimonious and revealing: using only three parameters, it enables us to portray and forecast mortality developments, capture the compression and shifting dynamics of mortality, and construct life-table functions. Given

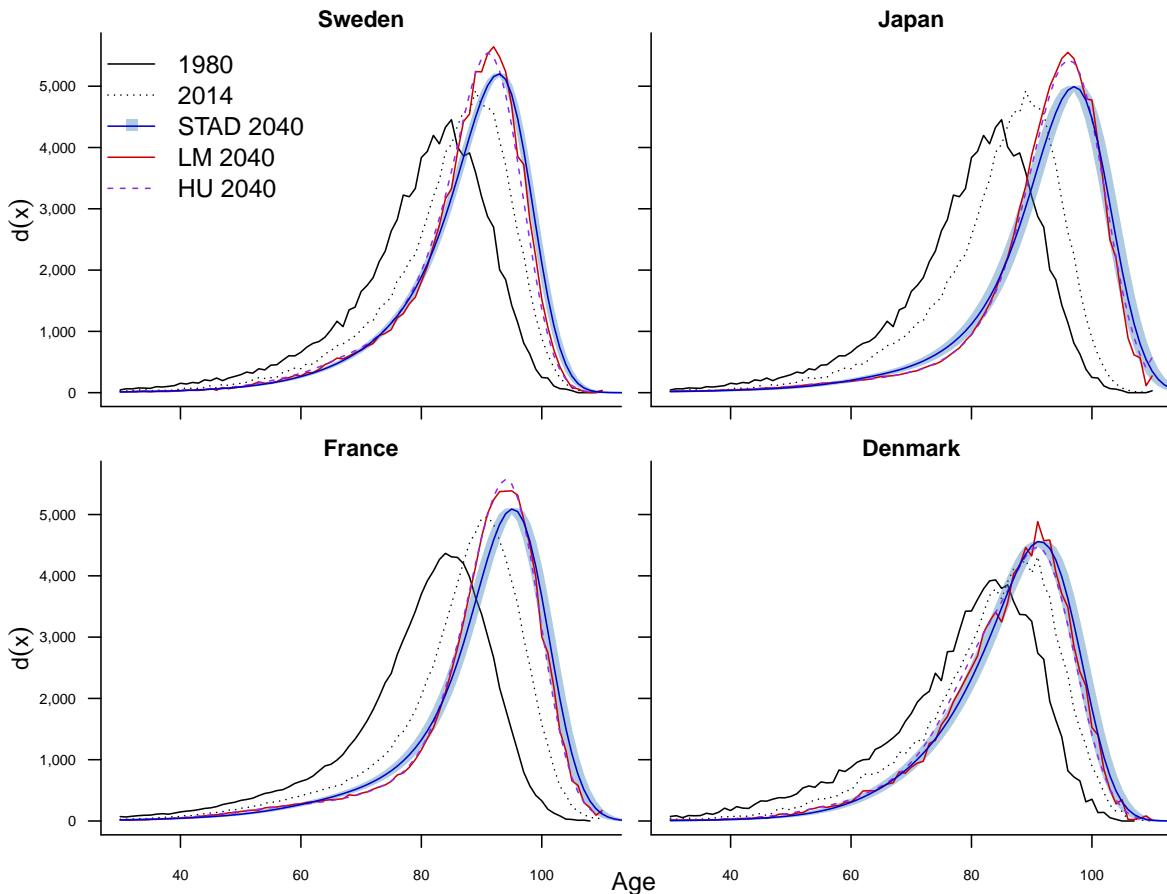


Figure 3.8: Observed life-table deaths in 1980 and 2014 and age-at-death distributions forecast in 2040 of the STAD (blue line and shading for the 80% prediction intervals), Lee and Miller (LM, red solid line) and Hyndman and Ullah (HU, purple dashed line) models for females in Sweden, Japan, France and Denmark.

its features, we call our suggested model *Segmented Transformation Age-at-death Distributions* (STAD).

The results indicate that the STAD model performs very well in terms of goodness-of-fit: the estimated remaining life expectancies and Gini coefficient (e_{30} and G_{30}) for females in four high-longevity countries (Sweden, Japan, France and Denmark) are very close to the observed historical values. Moreover, the STAD forecasts of e_{30} , G_{30} and of the logged age-specific death rates are more accurate than the Lee and Carter (LC, 1992) model and its variants in three out-of-sample validation exercises with forecast horizons of 10, 20 and 30 years. The STAD forecasts of e_{30} for the four countries up to 2040 reflect well the past linear increase observed between 1980 and 2014, and they are more optimistic than the LC variants forecasts, which have often under-predicted future gains in life expectancy (Lee and Miller, 2001). Finally, the STAD forecasts are characterized by greater shifting and smaller compression than the LC ones.

Theoretically, the STAD model can be viewed as a relational model (Brass, 1971). Here, transformations of a standard age-at-death distribution $f(x)$ describe mortality developments *over time* rather than across countries. From this perspective, an analogy can be made between the STAD and the LC model. The series of α_x in the LC can be viewed as a standard mortality age-profile, whose development over time is captured by the index κ_t modulated by a fixed rate

of mortality improvement at age x , β_x . As such, the LC model can also be interpreted as a relational model.

Our methodology is also inspired by the model proposed by [Camarda et al. \(2008\)](#) to analyse mortality developments from age-at-death distributions by transforming the age axis with a smooth non-linear warping function. Furthermore, [Cadena and Denuit \(2016\)](#) recently introduced an accelerated hazard relational model based on a smooth transformation of the age scale free of monotonic constraints, and they used it to estimate and project mortality in Belgium. The STAD model has elements that are common to both approaches, but it is also significantly different. The first model has been developed for mortality analyses, and the smooth function is not well suited for forecasting. The second model is based on the hazard function, and is subject to the same characteristics as models based on mortality rates discussed in the Introduction.

As in most relational models, the choice of the standard is important: all the features contained in $f(x)$ are in fact carried in the transformation function $t(\cdot)$. In order to maximize the representativeness of the standard with respect to the observed densities, we choose $f(x)$ as the mean of the aligned distributions, employing a landmark registration approach inspired from Functional Data Analysis ([Ramsay and Silverman, 2005](#)). With this technique, the ages occurring before and after the modal age are no longer conflated: differences in the mode (that are already accounted for by the parameter s) are removed, and the focus is only on differences in variability of the distributions. The use of functional data techniques in modelling and forecasting mortality already has other precedents in the literature ([Hyndman and Ullah, 2007](#); [Hyndman and Booth, 2008](#); [Hyndman et al., 2013](#)).

The shifting and compression dynamics of mortality changes have been investigated extensively in recent decades (for example, [Fries, 1980](#); [Kannisto, 2000](#); [Bongaarts and Feeney, 2002](#); [Bongaarts, 2005](#); [Canudas-Romo, 2008](#); [Thatcher et al., 2010](#); [Bergeron-Boucher et al., 2015](#); [de Beer and Janssen, 2016](#)). In most developed countries, a compression of mortality was observed during the first half of the twentieth century. In the second half of the century, the compression dynamic was mainly replaced by the shifting of the mortality schedule, with lifespan variability remaining practically constant. In addition to capturing longevity and lifespan variability changes, the three parameters of the STAD model provide information on these two dynamics. The parameter s regulates the shifting dynamic, as it measures the change in the modal age at death over time. The parameters b_L and b_U regulate the compression dynamic, as they expand or reduce the variability of deaths before and after the mode with respect to the standard distribution. These two parameters disentangle the contribution to the compression dynamic of two groups of people: those younger and those older than the modal age. This decomposition brings new evidence on the compression dynamic, adding a novel perspective to this stream of research ([Fries, 1980](#); [Myers and Manton, 1984](#); [Wilmoth, 1997](#); [Wilmoth and Horiuchi, 1999](#); [Lynch and Brown, 2001](#)).

The STAD model can thus be used to study the developments of the shifting and compression dynamics over a specified time interval, and this analysis can in turn usefully inform mortality projections. Appropriately accounting for these different dynamics in projections is a desirable property for forecasting mortality. For example, our analyses on females in four high-longevity countries between 1980 and 2014 suggest that the shifting dynamic has played a very important

role in mortality patterns (as it has been observed elsewhere, for example [Canudas-Romo, 2008](#); [Bergeron-Boucher et al., 2015](#); [de Beer and Janssen, 2016](#)). Compared to the LC model, the age-specific pattern of the STAD forecasts is characterized by a greater shift in all countries, as reflected in the age-at-death distribution forecast. Similar results have been shown for the Netherlands, using a new parametric model that considers both shifting and compression ([Janssen and de Beer, 2016](#)).

From a survival analysis perspective, the STAD model can be linked to the class of *accelerated failure time* (AFT) models. In survival analysis, AFT models have been introduced as an alternative to proportional hazard regression models to describe the effects of covariates in accelerating or decelerating the ageing process ([Wei, 1992](#); [Kalbfleisch and Prentice, 2002](#)). The STAD model can be interpreted as a particular type of AFT model, in which the ageing process is not uniformly modified with respect to the reference (the standard distribution), but it is rather modified in different ways before (b_L) and after (b_U) the modal age at death. For example, the ageing process could be accelerated before and decelerated after the mode with respect to the standard, or vice-versa, or accelerated/decelerated at different rates in the two age segments.

From this perspective, one limitation of the STAD model lies in the difficulty of interpreting its result from the individual standpoint. It is quite unrealistic to hypothesize that all individuals of a population age at two different rates, one before and one after the overall modal age at death. However, our aim here is to introduce a novel approach for modelling and forecasting mortality at the population level. The assumption of different rates of ageing can be more easily justified from this aggregate perspective, as the two parameters capture the *average* ageing rate in the two well-characterized age segments of the population. Nevertheless, while no experiments have been performed with medical and epidemiological datasets, the STAD might be unsuitable as a regression model for the analysis of survival data, and a completely smooth transformation function, as in [Camarda et al. \(2008\)](#), might be preferable. Furthermore, an additional limitation of the STAD model, as of all other models employed here, is that cohort effects are neither considered nor modelled. Analysis of cohort data is currently foreseen as future work.

Our interest in this article is limited to the senescent (older adult) mortality pattern, and our analyses start from age 30. Applying the STAD model to the entire age range reduces fitting accuracy. The reason for this loss of accuracy lies in the shape of the human mortality pattern. Since [Thiele \(1871\)](#), demographers use to decompose the mortality age-profile into three different groups operating principally, or almost exclusively, upon juvenile, younger adult and older adult ages. Applying the model from age 30 produces satisfactory results for the female populations analysed in this article because the first two components are negligible with respect to the senescent component. For males, the STAD model produces satisfactory results too (analyses not shown here), although the goodness-of-fit is slightly reduced compared to females. This is because the assumption that the younger adult mortality component (the accident hump) is negligible at age 30 is probably too strong for males. More accurate results for males could be obtained by extending this work to model and forecast the entire mortality pattern by: (i) disentangling and estimating the three independent age-specific mortality components and (ii) applying the STAD model, or a modified version of it, to each component. The Sum of Smooth

Exponentials model (Camarda et al., 2016) can be thought as a starting solution for the former issue.

We conclude with one last remark on our STAD model. We have chosen the modal age at death as the break-point of the segmented linear transformation because several demographic studies aforementioned investigated the variability of deaths after the modal age at death. In this respect, we add to this literature by studying the variability before and after the mode *simultaneously*. However, any other age of the distribution can be chosen as break-point, for example one of the other two central tendency measures (the mean and the median age). In practice, this might be more appropriate for studying mortality in countries with data deficiencies or for the analysis of causes of death, and this approach will be explored in future work.

3.5 Appendix

3.5.1 Deviance, Effective Dimension and BIC

The Deviance is often used as a measure of discrepancy between observed and fitted data. Within a Poisson framework, it is defined as:

$$\text{Dev} = 2 \sum_y \sum_x \left[D_{x,y} \ln \left(\frac{D_{x,y}}{\hat{D}_{x,y}} \right) - (D_{x,y} - \hat{D}_{x,y}) \right] \quad (3.8)$$

where $D_{x,y}$ and $\hat{D}_{x,y}$ denote the observed and fitted number of deaths at age x and year y , respectively. Higher values of the Deviance correspond to worse models in terms of goodness-of-fit.

The Bayesian Information Criterion (BIC, Schwarz, 1978) is frequently employed to assess model differences in terms of trade-off between model parsimony and accuracy. In a two-dimensional age and time setting, the BIC can be computed as:

$$\text{BIC} = \text{Dev} + \ln(mn) \text{ED} \quad (3.9)$$

where m and n are the dimensions (length) of age and time, respectively. ED denotes the Effective Dimension, or total number of parameters, of a model. Lower BIC values are associated with better models, and the trade-off between accuracy and parsimony is accounted for by the two components of the BIC.

3.5.2 Sec. 3.3.1: additional results

Here, we present the Deviance and Effective Dimension (ED) of the STAD model and LC variants corresponding to the analysis of Section 3.3.1. Table 3.4 shows the Deviance values of the STAD and LC variants in the four countries, and the ED of each model is reported in Table 3.5.

In terms of the Deviance, the Hyndman and Ullah method (HU) is the best fitting model, as its Deviance is always smaller than those of other models. Among others, one explanation for the improved fit of the model with respect to the STAD and other LC variants is that the HU model employs a very high number of parameters: almost four times more than the LC (765 versus 195) and five times more than the STAD (765 versus 138). This high parameterisation is

penalized by the BIC measure, so that the HU model is the optimal choice in only one instance (Japan, see Table 3.1). In terms of parsimony, the STAD model is the best performer, as it employs less parameters than the other models.

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	3179	3149	3153	3123	3060	2789	2958	2936
Japan	9962	9925	10268	9837	9240	3975	4612	4900
France	8046	6272	6400	6223	6113	3320	4050	3776
Denmark	4235	4063	4054	4036	3934	2938	3316	3808

Table 3.4: Deviance values of the STAD and LC model and variants for females aged 30-110+ in four countries during 1980-2014. Lower values of the Deviance (in bold) correspond to a better fit to the data.

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	138	195	195	195	195	765	765	765
Japan	138	195	195	195	195	765	765	765
France	138	195	195	195	195	765	765	765
Denmark	138	195	195	195	195	765	765	765

Table 3.5: Effective Dimension (ED, total number of parameters) of the STAD and LC model and variants for females aged 30-110+ in four countries during 1980-2014. Lower ED (in bold) corresponds to a more parsimonious model.

Chapter 4

A Three-component Approach to Model and Forecast Age-at-death Distributions

Ugofilippo Basellini
Carlo Giovanni Camarda

In Mazzuco, S., and Keilman, N. (eds.), *Developments in Demographic Forecasting*, Springer. Forthcoming.

A Three-component Approach to Model and Forecast Age-at-death Distributions

Ugofilippo Basellini^{1,2,†} and Carlo Giovanni Camarda¹

¹*Institut national d'études démographiques (INED), Paris, France*

²*Interdisciplinary Centre on Population Dynamics (CPop) and Department of Public Health,
University of Southern Denmark, Odense, Denmark*

[†]*Corresponding author: ugofilippo.basellini@ined.fr*

This is a post-peer-review, pre-copyedit version of a chapter that will be published in the Springer book *Developments in Demographic Forecasting* (eds. Mazzuco, S., and Keilman, N.) at the beginning of 2020. The final authenticated version will be open-access and available online.

Submitted: March 2019; Accepted: July 2019

Abstract

Mortality forecasting has recently received growing interest, as accurate projections of future lifespans are needed to ensure the solvency of insurance and pension providers. Several innovative stochastic methodologies have been proposed in most recent decades, the majority of them being based on age-specific mortality rates or on summary measures of the life table. The age-at-death distribution is an informative life-table function that provides readily available information on the mortality pattern of a population, yet it has been mostly overlooked for mortality projections. In this chapter, we propose to analyse and forecast mortality developments over age and time by introducing a novel methodology based on age-at-death distributions. Our approach starts from a nonparametric decomposition of the mortality pattern into three independent components corresponding to Childhood, Early-Adulthood and Senescence, respectively. We then model the evolution of each component-specific death density with a relational model that associates a time-invariant standard to a series of observed distributions by means of a transformation of the age axis. Our approach allows us to capture mortality developments over age and time, and forecasts can be derived from parameters' extrapolation using standard time series models. We illustrate our methods by estimating and forecasting the mortality pattern of females and males in two high-longevity countries using data of the Human Mortality Database. We compare the forecast accuracy of our model and its projections until 2050 with three other forecasting methodologies.

Keywords: Mortality forecasting · Mortality modelling · Relational models · Smoothing · Mortality decomposition · Life expectancy · Lifespan variability

4.1 Introduction

Population projections and mortality forecasts have been studied since the beginning of the twentieth century. The seminal works of [Whelpton \(1928, 1936\)](#) and [Lotka \(1939\)](#) on the cohort component method and the stable population contributed significantly to the development and application of population projections. Mortality forecasts go back at least to the beginning of the twentieth century, as actuaries were concerned about the financial effects of mortality improvements on life annuities and pensions ([Pollard, 1987](#)). It is however in the last three decades that mortality forecasting flourished, owing to the introduction and development of stochastic methodologies to project mortality.

Three functions can be used to analyse human mortality and its developments over age and time: the hazard, the survival and the probability density function ([Klein and Moeschberger, 2003](#)). These functions describe the same stochastic phenomenon and are uniquely related between each other: one can derive any two of them by knowing the third one, without the need of additional information.

Despite the complementarity of the mortality functions, the majority of forecasting techniques is based on age-specific mortality rates or death probabilities (for comprehensive reviews, see [Booth and Tickle, 2008](#); [Cairns et al., 2009](#); [Shang et al., 2011](#); [Stoeldraijer et al., 2013](#)). Most of these models take advantage of the regularities typically found in age- and time-patterns, such as the predominantly downward trend in age-specific mortality observed in many developed countries during the last 60 years, and they extrapolate the trends in the future using statistical methods ([Haberman and Renshaw, 2011](#)).

Nevertheless, the inspection of the other two functions can provide additional insights on mortality developments that one might not directly discern from a rate-based analysis. It is well known that the remarkable mortality improvements observed in these countries during the twentieth century are generally divided into two stages of mortality changes: compression and shifting dynamics (see, for example, [Fries, 1980](#); [Wilmoth and Horiuchi, 1999](#); [Kannisto, 2000](#); [Bongaarts, 2005](#); [Canudas-Romo, 2008](#)). Broadly speaking, the first stage took place in the first part of the century, as significant reductions in infant and childhood mortality resulted in greater equality in lengths of life. In the second part of the century, mortality improvements at older ages became more prominent, resulting in higher average lifespans with stagnating equality.

The age-at-death distribution is an excellent function to inspect these dynamics of mortality changes. Mortality compression can be detected from the reduction in the variability of the distribution, while shifting corresponds to a translation of the distribution to higher ages without relevant changes in its shape. In addition, the distribution provides immediate information on key questions in mortality studies, such as the longevity of the population, and the inequality in ages at death.

Figure 4.1 shows changes in the age-at-death distribution of Swiss males between 1950 and 2016. The graphical inspection of the death distribution readily provides information on the population's longevity, which is typically measured by life expectancy at birth or, in low mortality countries, by the modal age at death ([Kannisto, 2001](#); [Horiuchi et al., 2013](#)). Additionally, the variability of lifespans within the population can be directly assessed from the spread of the distribution or its interquartile range. The increase in longevity as well as the reduction of

lifespan variability for Swiss males during this period clearly emerge from Figure 4.1. Moreover, changes in the distribution over time highlight the two dynamics of mortality: for example, it is evident that the shifting dynamic of mortality started around the 1970-1980s, becoming more prominent in most recent decades, while the compression dynamic had been strongest in the decades 1950-1970 and 1990-2010.

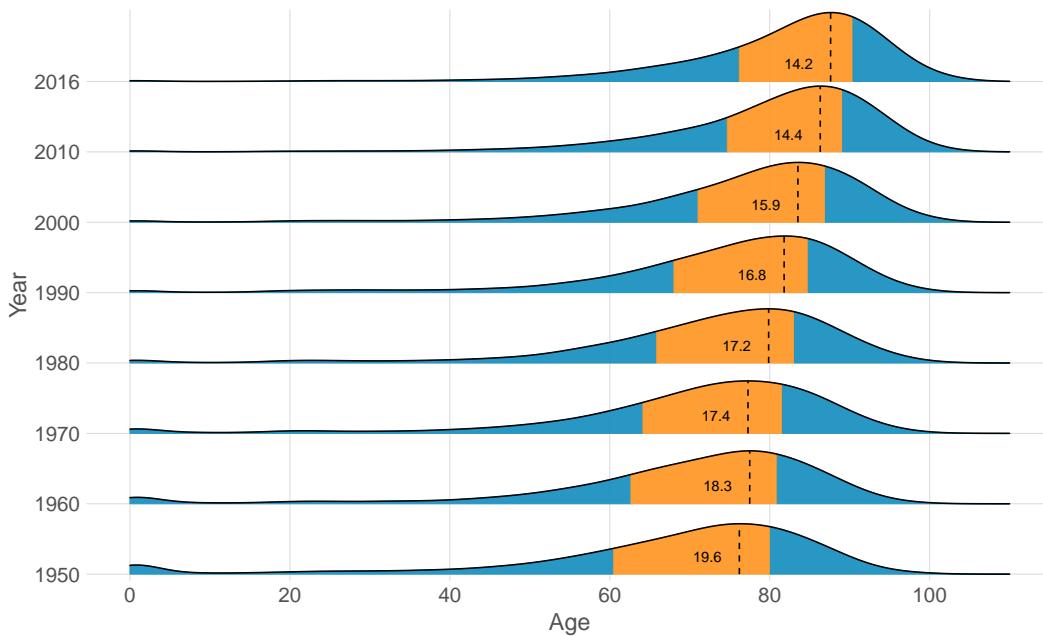


Figure 4.1: Changes in the age-at-death distribution for Swiss males at selected years between 1950 and 2016. The orange area corresponds to the interquartile range of the distribution, whose value is reported in print. The dashed line depicts the modal age at death. Data have been smoothed for illustrative purposes.

Source: Authors' own elaborations on data retrieved from the [Human Mortality Database \(2019\)](#).

Despite providing direct information on mortality patterns and trends over time, surprisingly few methods have been proposed to forecast mortality from age-at-death distributions. Among the firsts to abandon the conventional approach of using mortality rates, [Oeppen \(2008\)](#) and [Oeppen and Camarda \(2013\)](#) proposed to forecast the density of single and multiple-decrement life tables, using methodologies borrowed from compositional data analysis. [Bergeron-Boucher et al. \(2017\)](#) expanded on this work, suggesting a coherent model based on life-table deaths of fifteen Western European countries. Furthermore, [Basellini and Camarda \(2019b\)](#) proposed a relational model to forecast adult mortality from age-at-death distributions. Finally, [Pascariu et al. \(2019\)](#) suggested a vector autoregressive model to forecast the statistical moments of the death distribution.

In this chapter, we contribute to the growing literature of forecasting the age-pattern of mortality from age-at-death distributions. Specifically, we extend the Segmented Transformation Age-at-death Distributions (STAD) model proposed by [Basellini and Camarda \(2019b\)](#), which focuses on adult mortality only, to obtain mortality forecasts for the entire age range. While retaining the underlying methodology of the STAD model, here we introduce significant novelties to achieve our goal. In particular, our approach is based on two steps. First, we decompose the observed death counts into three additive mortality components, namely Childhood, Early-Adulthood and Senescent mortality. We perform this decomposition via the nonparametric

approach proposed by [Camarda et al. \(2016\)](#). Secondly, we model and forecast each component-specific age-at-death distribution employing specialized versions of the STAD model. As such, the Three-Component STAD (3C-STAD) model allows us to capture mortality developments over the entire age range, and forecasts are obtained from the extrapolation of the model's parameters using standard time-series techniques.

This chapter is organized as follows. In Section 4.2, we overview the methods that we introduce as well as the data that we employ. In Section 4.3, we provide two illustrations of our methodology by forecasting female and male mortality in two high-longevity countries. In particular, we first assess the accuracy of point and interval forecasts of the 3C-STAD model by performing three out-of-sample validation exercises. We then present the 3C-STAD forecasts until the year 2050. In both cases, we compare the 3C-STAD with three other well-known forecasting methodologies. Finally, in Section 4.4 we summarize and discuss our results.

4.2 Methods

4.2.1 Mortality functions

Human mortality can be analysed by any one of three complementary functions: the hazard, the survival and the probability density function ([Klein and Moeschberger, 2003](#)). In demography, for a given calendar year t , these functions are generally known as the force of mortality $\mu(x, t)$ at age x , the probability of surviving from birth to age x , $\ell(x, t)$, and the age-at-death distribution $f(x, t)$.

The three mortality functions are uniquely related between each other, and knowing one of them allows to determine the other two. In the following, without loss of generality, let $\ell(0, t)$, commonly labelled as the life-table radix, be equal to one, and let us drop the time index t to ease notation. The relationship that exists between the three functions at any age x is given by:

$$f(x) = \ell(x) \mu(x). \quad (4.1)$$

The probability of surviving $\ell(x)$ can be derived from the other two mortality functions:

$$\ell(x) = \exp\left(-\int_0^x \mu(a) da\right), \quad \ell(x) = \int_x^\omega f(x) dx, \quad (4.2)$$

where ω is the highest age attained in the population. Thus, combining (4.1) and (4.2) demonstrates the complementarity of the three mortality functions.

Since [Thiele \(1871\)](#), demographers and actuaries described human mortality into three different components that operates principally, or almost exclusively, upon childhood, middle and old ages, respectively. The attempt to decompose those three components stimulated numerous approaches (cf. Section 4.4). In a general setting, the hypothesis can be expressed as follows:

$$\mu(x) = \mu_c(x) + \mu_e(x) + \mu_s(x), \quad (4.3)$$

where the force of mortality $\mu(x)$ at age x is additively decomposed into three independent components, $\mu_c(x)$, $\mu_e(x)$, and $\mu_s(x)$. For ease of presentation, we labelled these mortality

component with Childhood, Early-Adulthood and Senescence, respectively. However, they theoretically operate over all ages x . Combining (4.1) and (4.3), the corresponding decomposition of the age-at-death distribution can be written as follows:

$$\begin{aligned} f(x) &= \ell(x) \mu_c(x) + \ell(x) \mu_e(x) + \ell(x) \mu_s(x) \\ &= f_c(x) + f_e(x) + f_s(x). \end{aligned} \quad (4.4)$$

Note that the overall age-at-death distribution $f(x)$ is a proper density function, i.e. $\int_0^\omega f(x) dx = 1$. Conversely, component-specific age-at-death distributions do not individually sum to one when integrated over the entire age range (cf. Equation (4.11) for the corresponding probability mass constraint in a discrete setting).

4.2.2 Data and mortality decomposition

Whereas risk of death acts continuously, mortality functions and models can be displayed only at particular ages and years. For modelling and forecasting mortality and for a specific sex and population, available data are thus observed death counts, $d_{x,t}$, and central exposures to the risk of death, $e_{x,t}$, with ages $x = 0, \dots, \omega$ and years t . In the following, we analyse the female and male populations of two high-longevity countries, Sweden and Switzerland, choosing a common time period (1950-2016) and with $\omega = 110+$. While Sweden was selected for the high standard in data quality, even at the oldest ages ([Vaupel and Lundström, 1994](#); [Wilmoth and Lundström, 1996](#)), Switzerland was chosen for its atypical mortality development, especially for males, related to the strong HIV epidemic during the 1980s ([Csete and Grob, 2012](#)). Data are taken from the [Human Mortality Database](#) (HMD, 2019).

We assume that the number of deaths at age x and year t is a random variable $D_{x,t}$ that follows a Poisson process ([Brillinger, 1986](#)):

$$D_{x,t} \sim \mathcal{P}(e_{x,t} \mu_{x,t}) \quad (4.5)$$

where the force of mortality $\mu_{x,t}$ is assumed to be constant over each year of age (i.e. from age x to $x+1$) and over each calendar year (i.e. from year t to $t+1$). This assumption implies that $\mu_{x,t}$ approximates the force of mortality at exact age $x + \frac{1}{2}$ and exact time $t + \frac{1}{2}$ ([Cairns et al., 2009](#)). Note that the notation $\mu_{x,t}$ is the discrete counterpart of the continuous notation $\mu(x,t)$ employed in Subsection 4.2.1. Moreover, death rates $m_{x,t} = d_{x,t}/e_{x,t}$ are the maximum likelihood estimators of the force of mortality $\mu_{x,t}$, if no structure is enforced over age and/or time.

The first step in the Three-Component Segmented Transformation Age-at-death Distributions (3C-STAD) model concerns the decomposition of the force of mortality into its three independent components $\mu_k(x)$, $k = c, e, s$. Instead of employing a parametric mortality model, we favour a non-parametric approach to avoid imposing a rigid structure and achieve a better fit to the observed data. For this purpose, we employ the Sum of Smooth Exponentials (SSE) model, which has been shown to provide insightful results for mortality analysis ([Camarda et al., 2016](#); [Remund et al., 2018](#)). In the following, we provide a short overview of the SSE model; for a more detailed description of the model, we refer the interested reader to the original paper of [Camarda et al. \(2016\)](#).

The SSE belongs to the class of multiple-component models (also known as competing hazard models, [Gage, 1993](#)), as it proposes an additive decomposition of the expected value of counts in multiple (smooth) components. In a given year t , let $\boldsymbol{\mu}$, \mathbf{d} and \mathbf{e} denote vectors over age of overall force of mortality, death counts and exposures, respectively. Within the SSE, we can model the force of mortality as the sum of three components $\boldsymbol{\gamma} = [\gamma_c : \gamma_e : \gamma_s]'$. The expected value of the Poisson process $\mathbf{d} \sim \mathcal{P}(\mathbf{e}\boldsymbol{\mu})$ is expressed as a composition of exposures and mortality components, i.e. $\mathbf{e}\boldsymbol{\mu} = \mathbf{C}\boldsymbol{\gamma}$, where the composition matrix $\mathbf{C} = [\mathbf{E} : \mathbf{E} : \mathbf{E}]$ is a block matrix that includes three times the diagonal matrix of population exposures $\mathbf{E} = \text{diag}(\mathbf{e})$ (one for each component of mortality). The composition matrix has the dual role of multiplying each component by the exposure times and of summing them to obtain the overall Poisson mean. The SSE model can be framed as a Composite Link Model ([Thompson and Baker, 1981](#)), and estimation of the model's parameters can be obtained by a modified version of the iterative reweighted least squares (IWLS) algorithm ([Eilers, 2007](#)).

The SSE model has several advantages over parametric decompositions of the force of mortality, which made it our favoured choice for the first step of the 3C-STAD. Although the SSE could accommodate parametric assumptions, it allows to model each component by assuming only smoothness over age (and eventually over time). We opted for this last more flexible setting. This can be achieved by expressing each component k as a linear combination of B -spline basis \mathbf{B}_k and associated coefficients $\boldsymbol{\alpha}_k$:

$$\gamma_k = \exp(\mathbf{B}_k \boldsymbol{\alpha}_k), \quad k = c, e, s. \quad (4.6)$$

Smoothness of γ_k is obtained by combining a large number of B -splines and a roughness penalty on the coefficients vector $\boldsymbol{\alpha}_k$ ([Eilers and Marx, 1996](#)). Note that the exponential in (4.6) guarantees positive component-specific force of mortality, as one would expect. Furthermore, component-specific shape constraints can be easily specified and included in the estimation procedure by additional asymmetric penalties. Here, we enforce monotonic decreasing and increasing constraints on the Childhood and Senescent components, respectively, and a log-concave shape for the Early-Adulthood component. These constraints further aid the identifiability of the model by ensuring that the three components are not interchangeable.

Another advantage of the SSE methodology is that it adequately blends the transitions between components, without imposing sharp delimitations where one stops and another one continues. Moreover, we employ the two-dimensional extension of the SSE model. In this way we both account for the significant age-time interactions and avoid abrupt changes over time in the interaction of the components. A detailed description of year-to-year mortality fluctuations is relevant in a forecasting perspective. In the SSE model, at the cost of overfitting, this flexibility is achieved by a large number of B -splines with a low smoothing parameter in the time dimension.

[Figure 4.2](#) shows an example of fitting the two-dimensional SSE model to Swiss males between 1950 and 2016: the three components of mortality clearly emerge, each one featuring the expected shape. Unlike the original SSE model, we start our analysis from age 0 which is treated in a specific manner. This particular age represents a clear discontinuity in the age-pattern of mortality, as mortality of newborns is sharply higher than death rates at later infant ages due to

malformations, pre-term births and birth-related complications (Chiang, 1984; Camarda et al., 2016). Hence, we incorporate the discontinuity in the first age of life by including, for the Childhood component, a specialized coefficient for this age, which is not penalized over age.

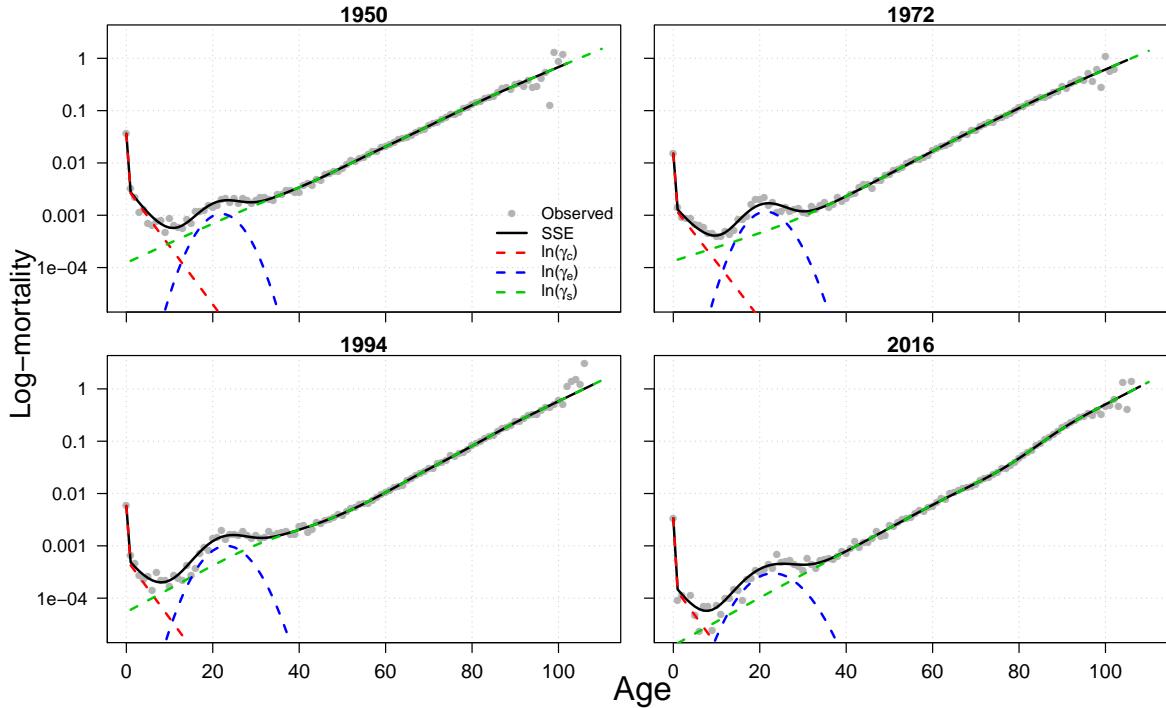


Figure 4.2: Observed and fitted mortality rates (in log scale) for Swiss males at selected years between 1950 and 2016. The force of mortality is decomposed into Childhood (γ_c), Early-Adulthood (γ_e) and Senescent (γ_s) components via the two-dimensional SSE model.

Source: As for Figure 4.1.

Outcomes from the SSE model allow us to obtain (i) the age-at-death distribution of each component over time (using standard life-table construction, Preston et al., 2001), and (ii) the expected number of deaths separated by component, $\hat{d}_k = e \hat{\gamma}_k$. This allows us to model and forecast age-at-death distributions independently for each component.

4.2.3 Modelling component-specific distributions

The second step of the 3C-STAD consists in modelling the component-specific age-at-death distributions. Since different features characterize the three components, we deal differently with each one of them.

Senescent mortality

We start by presenting the model employed for the Senescent component, originally proposed and described in greater details in Basellini and Camarda (2019b). The Segmented Transformation Age-at-death Distributions (STAD) is a relational model that relates a fixed time-invariant reference distribution, denoted standard, to a series of observed distributions via a segmented transformation of the age axis. In general, consider two age-at-death distributions $f(x)$ and $g(x)$, where the former is the standard, and the latter any observed distribution. The STAD model can be expressed as $g(x) = f[t(x; \theta)]$, where the transformation function $t(x; \theta)$ is characterized

by three parameters $\boldsymbol{\theta}$ that depend on: (i) the difference in modal ages at death between the two distributions, and (ii) the change in the variability of the two distributions *before* and *after* their modal ages.

Let $\nu_s = M_s^g - M_s^f$ denotes the difference between the mode of the Senescent distributions $g_s(x)$ and $f_s(x)$. The transformation function of the STAD model for the Senescent component, $t_s(\cdot)$, can then be written as:

$$t_s(x; \nu_s, b_s^\ell, b_s^u) = \begin{cases} M_s^f + b_s^\ell \tilde{x} & \text{if } x \leq M_s^g \\ M_s^f + b_s^u \tilde{x} & \text{if } x > M_s^g \end{cases} \quad (4.7)$$

where $\tilde{x} = x - \nu_s - M_s^f$, and b_s^ℓ and b_s^u denote the change in the variability of $g_s(x)$ with respect to $f_s(x)$ before and after the mode, respectively. Note that the superscript ℓ and u refer to the lower and upper segments of the age range (i.e. before and after the modal age at death).

The top panels in Figure 4.3 explain graphically the mechanisms underlying the STAD model for the Senescence component. Given a standard distribution (black lines in the graphs), let us consider the simpler case in which we vary the parameter ν_s but keep the variability parameters equal to 1, that is, $b_s^\ell = b_s^u = 1$. The transformation function in Equation (4.7) then simplifies to $t_s(x) = x - \nu_s$, and the resulting distribution is shifted along the x -axis by an amount equal to ν_s . This case corresponds to a shifting mortality scenario (blue lines in the graphs): the new distribution has the same shape and variability of the standard, but it is translated by the shifting parameter.

A more general development of mortality can be described by different values of the variability parameters, which act jointly with ν_s to modify the age-pattern of the standard distribution. When the two parameters are greater (lower) than 1, the variability of the segmented distribution is compressed (expanded) before and after the modal age at death with respect to the standard. In the top right panel of Figure 4.3, the segmented distribution has a lower variability ($b_s^\ell > 1$) before the mode and a higher variability ($b_s^u < 1$) above the mode as compared to the standard distribution. As such, increases in the two parameters capture the compression dynamic of mortality, distinguishing between changes that occur before and after the modal age at death.

Childhood mortality

The modal age at death for the Childhood component is invariably at age 0. The STAD is thus simplified and we drop from the transformation in (4.7) the part below the mode, i.e., we consider a left-truncated distribution with a constant mode at age 0. For the Childhood component, changes between the standard distribution, $f_c(x)$, and any observed distributions, $g_c(x)$, are modelled by varying the slope of the associated transformation of the age axis. In formulas, since $M_c^g = M_c^f = 0$, we can express the transformation of the age-axis as:

$$t_c(x; b_c^u) = b_c^u x. \quad (4.8)$$

The parameter b_c^u captures the change in the variability of the observed (left-truncated) distribution with respect to the standard distribution. The middle panels in Figure 4.3 present

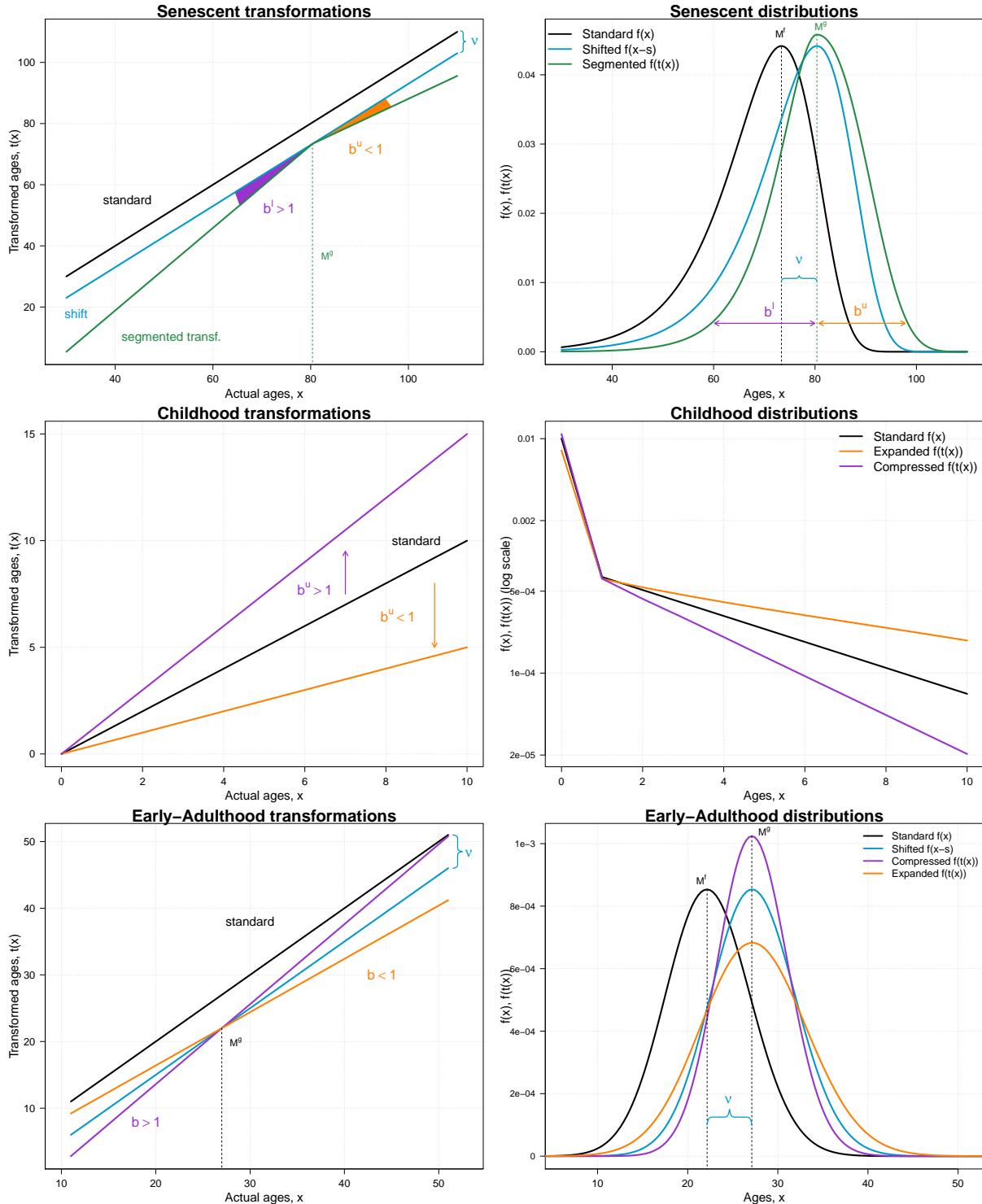


Figure 4.3: A graphical representation of the transformation functions (left panels) for the three components of the 3C-STAD model, and their effects on the corresponding component-specific age-at-death distributions (right panels).

Source: Authors' own elaborations.

this case. A parameter b_c^u larger than 1 will reduce the variability of the Childhood age-at-death distribution with respect to the standard one (purple lines). Vice versa, a slope smaller than 1 will lead to an increase of the variance of the associated distribution (orange lines).

Early-Adulthood mortality

The Early-Adulthood component of mortality is a typical and distinguishable feature of the human mortality pattern, which has been observed and modelled since the very first approaches to mortality decomposition (e.g. [Thiele, 1871](#); [Lexis, 1878](#); [Pearson, 1897](#)). Cause-of-death investigations of young excess mortality have often provided relevant policy recommendations ([Heuveline, 2002](#); [Remund et al., 2018](#)). As such, including this mortality component enhances the plausibility of fitted and forecast age-profiles, while improving the goodness-of-fit of the 3C-STAD model.

Transformations for the Early-Adulthood component account for changes in the component-specific modal age-at-death and for the variability of the observed distribution, $g_e(x)$, always with respect to the standard one, $f_e(x)$. Unlike the original STAD model, a linear transformation of the age axis without segmentation has been proven adequate for describing changes of the Early-Adulthood component over years. Therefore we do not differentiate between variability before and after the mode. This adaptation of the STAD can be thought as an Accelerated Failure Time model for age-at-death distributions, where the aging process is first shifted and then uniformly accelerated/decelerated with respect to the standard distribution.

Formally, we can write the transformation function for the Early-Adulthood component as:

$$t_e(x; \nu_e, b_e) = M_e^f + b_e \tilde{x} \quad (4.9)$$

where $\tilde{x} = x - \nu_e - M_e^f$, $\nu_e = M_e^g - M_e^f$ and the parameter b_e captures the change in the variability of the observed distribution $g_e(x)$ with respect to the standard $f_e(x)$. Bottom panels in Figure 4.3 illustrates the effect of $t_e(\cdot)$ on a theoretical standard distribution. A shifting mortality scenario for Early-Adulthood could be achieved by different values of the parameter ν_e , keeping $b_e = 1$ (blue lines). Alternatively, a b_e smaller than 1 leads to an increase of the variability of the distribution, simultaneously before and after the observed mode (orange lines). A shrinkage of the age axis is achieved by a b_e larger than 1, and it prompts a $g(x)$ with lower variability with respect to the standard $f_e(x)$ (purple lines).

4.2.4 Estimating and forecasting the 3C-STAD parameters

Being equipped with the component-specific transformation functions, we can move from the theoretical description of the 3C-STAD model to its actual application for modelling and forecasting a series of age-at-death distributions over time. The first step needed to achieve this goal is the choice of the standard distribution $f_k(x)$ for each component. For the Senescent component, we start by aligning the observed distributions to a common modal age at death, using a landmark registration approach frequently employed in Functional Data Analysis ([Ramsay and Silverman, 2005](#)). The alignment procedure corresponds to a plain shifting transformation of the observed densities, which preserve all their features except the modal value. The standard is then computed as the mean of the aligned distributions. This approach increases the representativeness of the standard, which does not conflate features of the distributions that occur at different distances with respect to the mode (for additional details and an explicative illustration, see [Basellini and Camarda, 2019b](#), pp. 122–124). For the Childhood and Early-Adulthood com-

ponents, we choose the standard as simple means of the observed distributions, as the alignment procedure is not required for the former, and it does not significantly improve the fit for the latter.

Table 4.1 summarizes all hypotheses made in the 3C-STAD model about each component, and the associated parameters that are necessary to estimate and forecast. Given the component-specific standard distributions, parameters of the transformation functions $t_k(\cdot)$ are estimated from the data by maximum likelihood. Here we make use of the outcomes of the SSE model (cf. Subsection 4.2.2), and expected number of deaths over age and time due to each component k , $d_{x,t}^k$ are modelled by the 3C-STAD. Given the actual exposures $e_{x,t}$ and assuming that component-specific expected deaths are Poisson distributed counts as in (4.5), we maximize the following log-likelihood function for each year t :

$$\ln \mathcal{L}(\boldsymbol{\theta}_{k,t} | d_{x,t}^k, e_{x,t}, \nu_{k,t}) \propto \sum_x [d_{x,t}^k \ln(\hat{\mu}_{x,t}^k) - e_{x,t} \hat{\mu}_{x,t}^k], \quad k = c, a, s \quad (4.10)$$

where $\hat{\mu}_{x,t}^k$ denotes the hazard of component k corresponding to the transformed distribution derived from $t_k(\cdot)$ applied in year t to the associated standard $f_k(x)$. In particular, the hazard $\hat{\mu}_{x,t}^k$ is derived from the age-at-death distribution $f_k(t_k(\cdot))$ using standard life-table formulas (Preston et al., 2001).¹ Note that the vector $\boldsymbol{\theta}_{k,t}$ contains only the variability parameter(s). For each year t , the shifting parameters ν_s and ν_e of the Senescent and Early-Adulthood components are computed as differences in the modal age at death between standard and observed distributions, as estimated by the SSE model.

Component, k	Transformation, $t_k(\cdot)$	Parameters		Standard, $f_k(x)$
		Shift	Variability	
Senescence	Segmented at the mode	ν_s	b_s^ℓ, b_s^u	Mean of aligned distr.
Childhood	Left-Truncated, no shift	—	b_c^u	Mean of distr.
Early-Adulthood	Linear, shift at the mode	ν_e	b_e	Mean of distr.

Table 4.1: Summary of the 3C-STAD model by component: type of transformation of the age axis, associated parameters and choice of the standard distribution.

Once the parameters have been estimated over all years t , we can model their trends using standard time-series methods. Mortality forecasts of the 3C-STAD model are then obtained by extrapolating the model's parameters and the time-fixed standard distributions. We combine univariate and multivariate approaches to achieve our goal. For the Senescent component, we employ the best fitting ARIMA(p,d,q) model for ν_s , and a VAR(1) model for b_s^ℓ and b_s^u (as in Basellini and Camarda, 2019b). For the Childhood component, the parameter b_c^u is modelled with the best fitting ARIMA(p,d,q) model, while for the Early-Adulthood parameters ν_e and b_e we employ a VAR(1) model.

The 3C-STAD acts directly on age-at-death distributions, therefore we must ensure that the sum over ages x of the three component-specific probability masses is equal to 1, that is:

$$\sum_x f_{x,t} = \sum_x (f_{x,t}^c + f_{x,t}^e + f_{x,t}^s) = 1 \quad (4.11)$$

¹One readily implemented approach to derive the hazard from age-at-death distribution in R is provided by the function `convertFx` in the `MortalityLaws` package (Pascariu, 2018).

for each year t . Consequently and in addition to the shifting/variability parameters, it is necessary to forecast the probability masses of the three components. Specifically, we recognize the compositional nature of a set of component-specific age-at-death distributions: we are dealing with three non-negative components that always sum to a constant. We thus employ a Compositional Data methodology to model and forecast the time series of component-specific probability masses (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011). Specifically, we transform the probability masses for each component obtained by the SSE model using an additive log-ratio transformation. This procedure produces two time-series that are unconstrained (i.e. they take values on the entire set of real numbers). The two transformed time-series are modelled and forecast with a VAR(1). We finally back-transform the results to obtain forecasts of the original time-series. For each forecast year, these back-transformed series sum up to 1 because they have been treated as compositional data. Note that this approach reduces the dimensionality of the forecasting problem for the probability masses by one dimension, i.e. from three to two time-series.

Finally, the complexity of our methodology requires a bootstrapping procedure to produce prediction intervals (PI, Efron and Tibshirani, 1994). We take into account the uncertainty of the 3C-STAD parameters by simulating 1000 new time-series of all parameters from randomly resampled residual values. For each simulation, we then forecast mortality patterns and associated summary measures. From the obtained distribution of forecast simulations, we took the median as central forecast, and the lowest and highest deciles to construct 80% PI. Residual bootstrap of this type has already been employed to construct PI in mortality models (Bergeron-Boucher et al., 2017; Basellini and Camarda, 2019b).

Routines for estimating and forecasting the parameters of the 3C-STAD model were implemented in R (R Development Core Team, 2019) and are available online². Our routines take advantage of the R packages `forecast`, `demography`, `MortalitySmooth`, `MortalityLaws` and `vars` (Pfaff, 2008a,b; Hyndman and Khandakar, 2008; Camarda, 2012; Hyndman et al., 2018a,b; Pascariu, 2018).

4.3 Results

4.3.1 Out-of-sample validation

Here, we assess the predictive performance of the 3C-STAD model using out-of-sample validation. Specifically, we employ data of the Human Mortality Database (2019) for the female and male populations of Sweden and Switzerland for the period 1950-2016. For each population, we perform three exercises, corresponding to validation periods of 10 years (training period 1950-2006), 20 years (training period 1950-1996) and 30 years (training period 1950-1986). The common starting year of analysis, 1950, was chosen in order to have training periods longer than validation horizons for each exercise.

To assess the performance of our forecasts, we employ the standard life-table functions: life expectancy at birth (e_0) as measure of population's longevity, and age-specific mortality rates (in

²R codes to replicate all results presented in this chapter are available at <https://github.com/ubasellini/3C-STADmodel>.

log scale, $\ln(m_{x,t})$), which measure the age-pattern and intensity of mortality. Additionally, we use the Gini coefficient (G_0), a measure of lifespan inequality, whose importance for evaluating mortality forecasts has been recently highlighted (Bohk-Ewald et al., 2017).

We compare the performance accuracy of the 3C-STAD model with three other forecasting methodologies. First, given its prominence and wide application, we employ the original Lee-Carter (LC) model (Lee and Carter, 1992). Second, since one limitation of the LC model is the lack of smoothness in the fitted and forecast mortality rates, we use the Hyndman-Ullah (HU) functional data model (Hyndman and Ullah, 2007), which overcomes this limitation by smoothing the starting data as a first step. Third, we choose the CODA model proposed by Oeppen (2008): this model is indeed closer in spirit to the 3C-STAD, as it models and forecasts the age-at-death distribution. The LC and HU models were estimated and forecast with the R packages `forecast` and `demography` (Hyndman et al., 2018a,b; Hyndman and Khandakar, 2008). The CODA model was fitted and forecast using the R codes provided in the Supplementary Material of Bergeron-Boucher et al. (2017).

Our evaluations of mortality forecasts are based on the accuracy of both point predictions and calibration of prediction intervals (PI), as both measures are relevant for the validation of probabilistic projections (Chatfield, 2000). Greater accuracy in point forecasts occurs when point predictions are closer to the observed data. To evaluate point forecasts, we employ the mean absolute error (MAE), which is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{t \in T} |\hat{y}_t - y_t|$$

where \hat{y}_t is the point forecast at time t for either life expectancy at birth, mortality rates or Gini coefficient. Associated out-of-sample observed values are denoted by y_t . The set of validation years is T , and N is the total number of data used for validation. Note that for mortality rates, mean is computed over ages, too.

Greater calibration of PI is achieved when the proportion of out-of-sample data that falls within the calculated PI is closer to the given nominal level (for example, 80% or 95%). To evaluate interval forecasts, we compute the empirical coverage probability (ECP) of the 80% PI for each model (as in, for example, Shang et al., 2011; Raftery et al., 2013). For the sake of consistency and fairness, we computed the PI for all models by the same bootstrapping procedure, i.e. residual bootstrap of the time-series of the model's parameters (cf. Subsection 4.2.4).

In addition to the MAE and ECP, scoring rules can be used to assess calibration and sharpness of probabilistic forecasts simultaneously (for a review, see Gneiting and Katzfuss, 2014). Scoring rules allow to jointly assess point and interval predictions by providing a summary measure of the predictive performance that forecasters aim to minimize. Here, we employ the Dawid-Sebastiani score (DSS) (Dawid and Sebastiani, 1999), which is given by:

$$\text{DSS}_t = \frac{(y_t - \mu_{F,t})^2}{\sigma_{F,t}^2} + 2 \ln \sigma_{F,t}, \quad t \in T$$

where $\mu_{F,t}$ and $\sigma_{F,t}^2$ are the first two central moments of the probabilistic forecast at time t , y_t is the associated out-of-sample observed value, and T is the set of validation years. We then compute the mean value of the DSS for all the data used for validation.

Table 4.2 reports the point, interval and probabilistic forecast accuracy of the four models in the three out-of-sample scenarios as well as for all the four populations analysed here. Bold values correspond to better performances. In terms of point forecast, the 3C-STAD is the most accurate model, as its forecasts are more or as precise as those of the other models. Out of 36 indicators, the 3C-STAD outperforms 20 times. The LC is the second most precise model with 9 indicators, followed by the HU and CODA models, each with 8 and 3 indicators, respectively. Note that the sum does not add up to the total number of indicators due to the draw of some models for some specific measures (for example, both the 3C-STAD and LC models are equally best performers for the indicator G_0 for Swedish females in the 30y exercise). In terms of interval forecast, the CODA outperforms all other models, being more accurate for 15 indicators over 36. The 3C-STAD, LC and HU follow, each with 12, 11 and 7 indicators, respectively. Finally, if we consider point and prediction accuracy simultaneously using the DSS measure, we find that the 3C-STAD model is the best performer, outperforming the others for 12 indicators. The LC, CODA and HU models follow with 9, 8 and 7 indicators, respectively.

4.3.2 Forecast to 2050

Having assessed and compared the forecast accuracy of the 3C-STAD model, we now present its mortality forecasts for the four populations analysed until 2050. As in the previous Subsection, we compare projections based on the 3C-STAD model with those of LC, HU and CODA models.

Figure 4.4 shows the observed and forecast life expectancy at birth (e_0) and Gini coefficient (G_0) in the four populations for the years 1950-2050. In terms of e_0 , the 3C-STAD forecasts are always more optimistic than those of the LC and HU model. With respect to CODA, the 3C-STAD is more optimistic for males and less optimistic for females. In terms of lifespan inequality, CODA forecasts are the most egalitarian in 2050 (lower values of G_0) for the female populations, while the 3C-STAD predicts more equality for males.

In Figure 4.5, we compare the age-specific mortality rates forecasts in 2050 for all populations. Several differences emerge between the models from this age-pattern analysis. Mortality rates of the 3C-STAD are smooth, lacking the jagged features visible in the LC and CODA forecasts. This is a great advantage for long-term mortality projections (Li et al., 2013). Additionally, the Swedish projections of the 3C-STAD do not display an unexpected S-shape displayed by other models in the age range 60-100.

Finally, Figure 4.6 shows the observed age-at-death distribution for the four populations in 2016, along with the 2050 forecasts of the four models. With respect to the other models, the 3C-STAD forecasts are characterized by greatest shift for all the populations. In addition to this, the 3C-STAD projections are also less compressed than those of other models, with the exception of Swedish males.

Country	Training period	Validation period	Measure	Point forecast (MAE)				Interval forecast (ECP for 80% PI)				Probabilistic forecast (mean DSS)							
				3C-STAD		3C-STAD		LC		LC		CODA		HU					
				e_0	G_0	0.11	0.29	0.40	0.16	100%	100%	90%	90%	-1.44	-1.03	-0.72	-2.15		
Switzerland Females	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.19	0.08	0.11	0.11	0.09	0.06	100%	80%	90%	80%	-4.12	-2.83	-2.89	3.98		
		20y	e_0 G_0 $\ln(m_{x,t})$	0.19	0.08	0.11	0.13	0.39	0.12	100%	35%	47%	37%	8.69	23.63	58.46	-4.98		
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.74	0.43	0.45	0.20	0.20	0.20	100%	100%	95%	95%	-0.54	-0.51	-0.51	-2.13		
		30y	e_0 G_0 $\ln(m_{x,t})$	0.21	0.40	0.30	0.13	0.30	0.23	51%	48%	54%	38%	-3.49	-2.73	-2.08	-3.13		
Switzerland Males	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.19	0.13	0.11	0.12	0.24	0.22	38%	35%	47%	37%	-0.54	-0.51	-0.51	-2.13		
		20y	e_0 G_0 $\ln(m_{x,t})$	0.19	0.13	0.11	0.13	0.23	0.20	100%	100%	100%	100%	-0.54	-0.51	-0.51	-2.13		
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.22	0.28	0.23	0.20	0.23	0.20	54%	50%	58%	43%	-0.59	32.49	17.09	-6.02		
		30y	e_0 G_0 $\ln(m_{x,t})$	0.22	0.28	0.23	0.20	0.23	0.20	54%	50%	58%	43%	-2.78	4.42	7.84	-7.59		
Sweden Females	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.17	0.13	0.19	0.41	0.34	0.78	100%	90%	100%	100%	0%	-1.44	-0.72	-0.32	3.96	
		20y	e_0 G_0 $\ln(m_{x,t})$	0.25	0.29	0.23	0.20	0.20	0.20	37%	32%	53%	31%	-2.18	22.42	38.49	-7.17		
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.31	0.31	0.27	0.31	0.31	0.31	22%	20%	25%	0%	2.82	3.14	2.65	43.43		
		30y	e_0 G_0 $\ln(m_{x,t})$	0.31	0.31	0.27	0.31	0.31	0.31	29%	29%	43%	17%	0.90	6.38	1.52	51.82		
Sweden Males	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.16	0.18	0.12	0.12	0.26	0.26	80%	70%	100%	100%	0%	2.39	4.04	31.51	3.31	
		20y	e_0 G_0 $\ln(m_{x,t})$	0.25	0.29	0.25	0.28	0.32	0.32	15%	15%	20%	20%	0%	5.49	168.12	56.82	6.37	
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.31	0.31	0.27	0.31	0.31	0.27	22%	22%	30%	30%	17%	3.10	1.84	9.40	39.06	
		30y	e_0 G_0 $\ln(m_{x,t})$	0.31	0.31	0.27	0.31	0.31	0.27	25%	35%	42%	21%	5.52	90.05	9.40	8.64		
Sweden Females	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.05	0.13	0.10	0.38	0.15	0.15	100%	100%	100%	100%	100%	-2.27	-1.56	-0.59	-2.27	
		20y	e_0 G_0 $\ln(m_{x,t})$	0.17	0.18	0.16	0.22	0.16	0.22	37%	39%	55%	44%	44%	-4.16	-3.67	-2.75	-2.33	
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.18	0.19	0.17	0.73	0.46	0.46	90%	90%	95%	95%	20%	-0.82	-0.68	0.28	-0.04	
		30y	e_0 G_0 $\ln(m_{x,t})$	0.18	0.19	0.19	0.22	0.20	0.20	95%	95%	54%	54%	28%	-3.92	-2.69	-2.20	-4.00	
Sweden Males	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.07	0.07	0.07	0.69	0.53	1.43	0.43	23%	23%	97%	7%	30%	0.92	-0.11	8.16	
		20y	e_0 G_0 $\ln(m_{x,t})$	0.07	0.07	0.07	0.74	0.54	1.43	0.43	23%	23%	87%	87%	20%	-2.31	-2.72	1.87	-3.59
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.18	0.21	0.18	0.27	0.19	0.19	39%	39%	39%	34%	29%	0.10	17.38	71.52	0.16	0.02
		30y	e_0 G_0 $\ln(m_{x,t})$	0.18	0.21	0.18	0.27	0.19	0.19	39%	39%	39%	34%	29%	0.10	17.38	71.52	0.16	0.02
Sweden Males	1950-2006	10y	e_0 G_0 $\ln(m_{x,t})$	0.03	0.10	0.11	0.38	0.41	0.81	100%	100%	90%	100%	100%	0%	-1.65	-0.95	-0.90	4.72
		20y	e_0 G_0 $\ln(m_{x,t})$	0.72	0.32	1.35	1.34	1.69	1.69	100%	100%	15%	15%	0%	49%	27%	-3.48	-3.04	-3.51
	1950-1996	30y	e_0 G_0 $\ln(m_{x,t})$	0.20	0.29	0.37	0.26	0.39	0.21	51%	10%	40%	32%	21%	0%	-0.17	-0.19	-0.90	1.50
		30y	e_0 G_0 $\ln(m_{x,t})$	0.29	0.38	2.65	2.43	2.59	3.14	7%	13%	23%	17%	18%	0%	18.49	7.64	72.16	149.89

Table 4.2: Mean absolute error (MAE), empirical coverage probability (ECP) for the 80% PI, and mean Dawid-Sebastiani score (DSS) of the 3C-STAD, LC, CODA and HU forecasts of e_0 , G_0 and $\ln(m_{x,t})$ for females and males in two countries and three out-of-sample exercises: validation periods of 10 years (training period 1950-2006), 20 years (1950-1996) and 30 years (1950-1986). Lower values of the MAE and of the DSS (in bold) correspond to greater forecast accuracy. Values of the ECP closer to the 80% nominal level (in bold) correspond to greater interval forecast accuracy. Source: As for Figure 4.1.

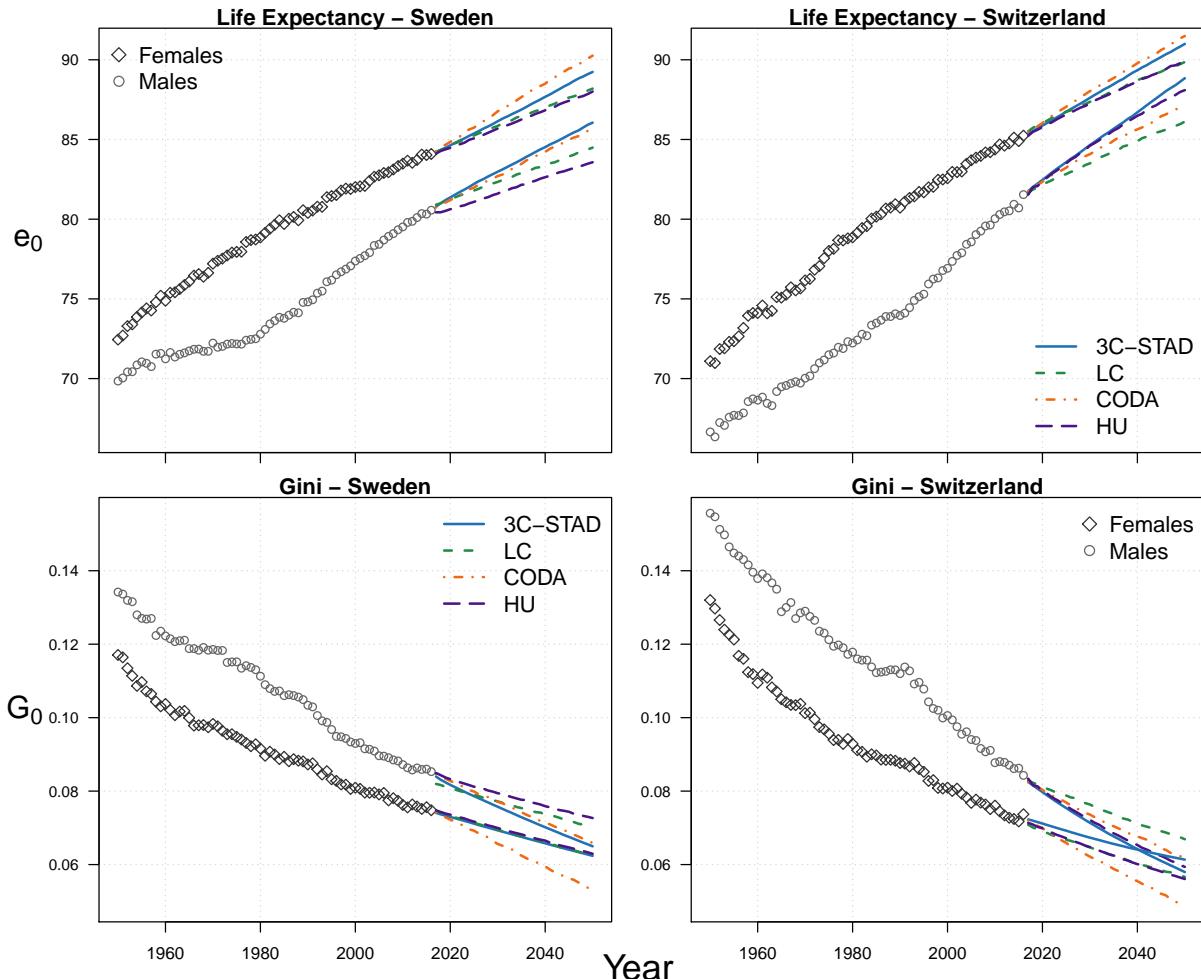


Figure 4.4: Observed and forecast life expectancy at birth (e_0 , top panels) and Gini coefficient (G_0 , bottom panels) females and males in Sweden and Switzerland, 1950–2050.

Source: As for Figure 4.1.

4.4 Discussion

Age-at-death distributions have generally been neglected for modelling and forecasting mortality, despite providing insightful information on mortality age-patterns and trends over time. In this chapter, we introduced a novel stochastic methodology to forecast mortality that is based on changes in age-at-death distributions. Our proposed Three-Component Segmented Transformation Age-at-death Distributions (3C-STAD) model captures and forecasts mortality developments over age and time by: (i) decomposing mortality into three independent components, namely Childhood, Early-Adulthood and Senescence, and (ii) modelling and forecasting changes in each component-specific age-at-death distributions.

The decomposition of the mortality age-pattern into multiple components has a long history in demographic analysis. In 1871, Thiele pioneered this decomposition by expressing the force of mortality as the sum of three independent components that operate principally, or almost exclusively, upon childhood, middle and old ages, respectively. Shortly afterwards, Lexis (1878) theorized a similar three-component decomposition, but he shifted the attention from the force

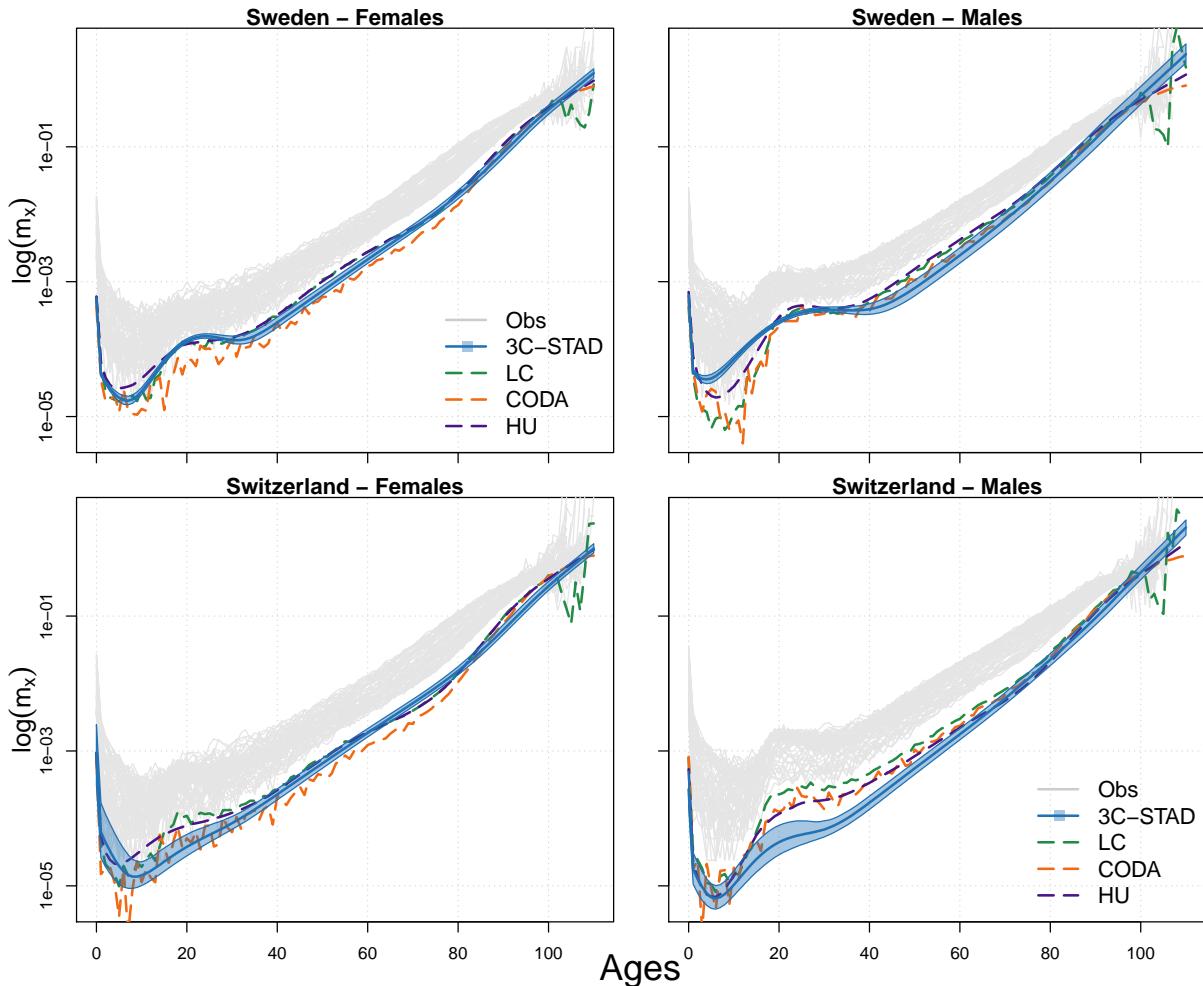


Figure 4.5: Observed age-specific mortality rates in 1950–2016 (grey lines) and forecast rates of four models in 2050 for females and males in Sweden and Switzerland. Shaded areas correspond to 80% PI for the 3C-STAD model.

Source: As for Figure 4.1.

of mortality to the age-at-death distribution. His ideas were followed upon and further elaborated by Pearson (1897), who divided the death density into five components, each one with its own distribution with different masses and degree of skewness. Finally, more recently, different parametric approaches have been proposed to decompose human mortality patterns (Siler, 1979; Heligman and Pollard, 1980; Kostaki, 1992; de Beer and Janssen, 2016; Mazzuco et al., 2018).

For our purposes, we performed a non-parametric decomposition using the Sum of Smooth Exponentials (SSE) model (Camarda et al., 2016). We favour this over other parametric approaches because it allows us to achieve a good fit to the observed data without imposing a rigid parametric structure, hence adapting the decomposition to a large and diverse range of mortality developments. Moreover, via the SSE model, we obtain smooth components with specific shape constraints, and a two-dimensional age-time perspective is incorporated into the mortality decomposition. Component-specific age-at-death distributions derived by the SSE model are then isolated to model and forecast their changes. To do so, we employ modified versions of the

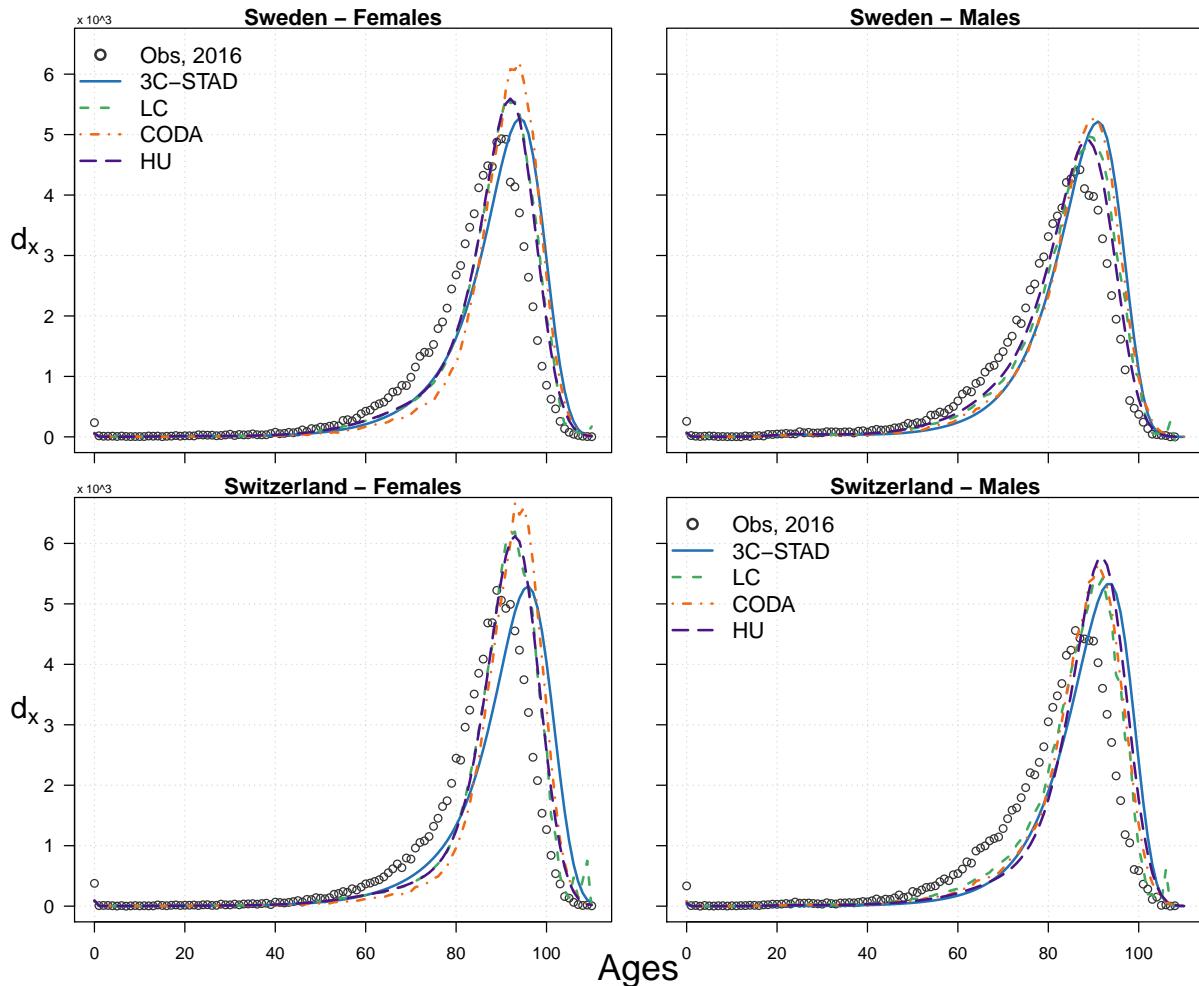


Figure 4.6: Observed age-at-death distribution in 2016 (grey points) and forecasts of four models in 2050 for females and males in Sweden and Switzerland.

Source: As for Figure 4.1.

relational model proposed by [Basellini and Camarda \(2019b\)](#), originally designed for forecasting only adult distributions of deaths.

We have applied the 3C-STAD model to the female and male populations of Sweden and Switzerland using data retrieved from the [Human Mortality Database \(2019\)](#). First, we assessed the point and interval forecast accuracy of the model by performing three out-of sample validation exercises. We have then forecast mortality for each population until 2050. In both cases, we compared the 3C-STAD projections with those of three well-known and employed methodologies: the Lee-Carter (LC, [Lee and Carter, 1992](#)), the CODA ([Oeppen, 2008](#)) and the Hyndman-Ullah (HU, [Hyndman and Ullah, 2007](#)) models. We compare forecasts of summary measures, such as life expectancy at birth (e_0) and lifespan inequality (as measured by the Gini coefficient, G_0), as well as age-specific functions, such as death rates or age-at-death distributions.

The results of the out-of sample validation exercises show that the 3C-STAD produces accurate mortality forecasts, both in terms of point forecasts and prediction intervals (PI). In particular, the 3C-STAD was the most accurate model for point forecasts with respect to other models. Additionally, the 3C-STAD PI outperformed the other models for one indicator out of three (see Table 4.2).

Concerning interval forecasts, CODA was found relatively more accurate, a result that might be related to the fact that “the PI are wider with a CODA method than with an LC method” ([Bergeron-Boucher et al., 2017](#), p. 546). However, when we considered point and interval forecasts simultaneously using a scoring rule, the wide PI of the CODA were penalized, and the 3C-STAD and LC models were preferred to the CODA. Within 3C-STAD framework, a possibility to improve estimation of PI would be to include the uncertainty related to the SSE decomposition. However, preliminary analyses showed that this approach raises computational burden without a significant widening of the forecast variability. It is likely that the reason is due to our usage of the SSE model. In the decomposition procedure, we aim to follow mortality data as close as possible, consequently the SSE model presents extremely small uncertainty. Nonetheless, we envisage alternative procedures to further improve estimation of the interval accuracy of the 3C-STAD model.

Mortality forecasts until 2050 for the four populations highlighted additional differences between models. The 3C-STAD and CODA forecasts of e_0 are generally more optimistic than those of the LC and HU models. Forecasting age-at-death distributions instead of mortality rates here translates into more optimistic forecast of life expectancy, a preliminary finding already observed elsewhere ([Bergeron-Boucher et al., 2019](#)). This could be an advantage, given that the LC forecasts have often under-predicted future gains in life expectancy ([Lee and Miller, 2001](#)). Significant differences further emerge from an age-specific analysis of the different projections. On one side, the 3C-STAD forecast rates are inherently smooth, which is a desirable property especially for long-term projections ([Li et al., 2013](#)). On the other side, the 3C-STAD forecast age-at-death distributions are characterized by greater shifting and smaller compression than those of other models. These projections seem more plausible, given that the shifting mortality dynamic has replaced the compression one in high-longevity countries in the most recent decades ([Canudas-Romo, 2008](#); [Bergeron-Boucher et al., 2015](#); [Janssen and de Beer, 2019](#)).

In general, we regard three characteristics as desirable for any forecasting methodology. First, the model should be able to capture and forecast mortality trends that can move in different directions across ages. Second, the relevant dynamics of mortality changes observed during the last century, i.e. shift and compression, should be appropriately accounted for. Third, the forecast age-profile of mortality rates should be smooth, without implausible jaggedness where rates of adjacent age groups have very different and volatile values. Despite being one of the most employed forecasting methodology by public and private companies, the seminal LC model does not satisfy any of these properties. The single time index regulates the direction of change for mortality rates at all ages, i.e. mortality improvements occur in the same direction at all ages. Furthermore, the model cannot account for the two mortality dynamics, and forecasts age-pattern are very volatile and jagged (see Figures 4.5 and 4.6).

Conversely, the 3C-STAD model meets all these three requirements. On one hand, the mortality decomposition allows us to capture and forecast mortality improvements across ages without rigid assumptions. Smoothness in the fitted and forecast age-profiles is a by-product of the non-parametric decomposition that we have employed. On the other hand, the 3C-STAD parameters capture and disentangle the shifting and compression mortality dynamics. The re-

cently proposed model of [Bardoutsos et al. \(2018\)](#) is another example of projection methodology that satisfies these features.

Obviously, the 3C-STAD is not free of shortcomings, and neither we claim here that it outperforms all other forecasting methodologies. In addition to the width of the PI mentioned before, the computational time needed to produce mortality forecasts could be improved. The estimation of the two-dimensional SSE model in fact generally requires around thirty minutes, and speeding this step up will be required to shorten computational times. Future mortality values are obtained by forecasting eight time-series. Although this feature might pose issues in other situations, all of these series have clear demographic meanings and rather intelligible trends. Combination of univariate and multivariate time-series approaches has thus provided a reliable tool for overcoming this seemingly critical drawback of the 3C-STAD model. Different approaches in extrapolating the eight time-series will be pursued, also for assessing consequences of specific future demographic scenarios. Moreover, in line with recent literature ([Li and Lee, 2005](#); [Hyndman et al., 2013](#); [Janssen et al., 2013](#); [Bergeron-Boucher et al., 2017](#)), future research will be directed towards the inclusion of coherence as an additional factor to improve forecasts for a group of (sub)populations.

To conclude, we have shown that the proposed 3C-STAD model offers great prospects for modelling and forecasting human mortality. In light of the generally pessimistic forecasts of the widely employed LC model ([Li et al., 2013](#); [Seligman et al., 2016](#)), forecasting methodologies, such as the 3C-STAD, should be explored by pension and insurance providers to better assess their solvency needs, and by statistical bureaus to produce alternative population projections.

Chapter 5

An age-at-death distribution approach to forecast cohort mortality

Ugo Filippo Basellini

Søren Kjærgaard

Carlo Giovanni Camarda

Manuscript under review.

An age-at-death distribution approach to forecast cohort mortality

Ugo Filippo Basellini^{1,2,†}, Søren Kjærgaard² and Carlo Giovanni Camarda¹

¹*Institut national d'études démographiques (INED), Paris, France*

²*Interdisciplinary Centre on Population Dynamics (CPop) and Department of Public Health,
University of Southern Denmark, Odense, Denmark*

† Corresponding author: ugofilippo.basellini@ined.fr

Abstract

Mortality forecasting has received increasing interest during recent decades due to the negative financial effects of continuous longevity improvements on public and private institutions' liabilities. However, little attention has been paid to forecasting mortality from a cohort perspective. In this article, we introduce a novel methodology to forecast adult cohort mortality from age-at-death distributions. We propose a relational model that associates a time-invariant standard to a series of fully and partially observed distributions. Relation is achieved via a transformation of the age-axis. We show that cohort forecasts can improve our understanding of mortality developments by capturing distinct cohort effects, which might be overlooked by a conventional age-period perspective. Moreover, mortality experiences of partially observed cohorts are routinely completed. We illustrate our methodology on adult female mortality for cohorts born between 1835 and 1970 in two high-longevity countries using data from the Human Mortality Database.

Keywords: Mortality forecasting · Mortality modelling · Relational models · Cohort life table ·

Smoothing

5.1 Introduction

Continuous and widespread gains in life expectancy ([Riley, 2001](#); [Oeppen and Vaupel, 2002](#)) are increasingly challenging governments and insurance companies to provide adequate pension products and elderly health care in ageing societies. Mortality forecasting has thus gained relevant prominence during the last decades, as relatively small differences in the expected lifetimes of pensioners can have significant effects on financial institutions' liabilities.

A growing number of models have recently been proposed to forecast human mortality using stochastic methodologies that produce probabilistic assessments of the future. For comprehensive reviews, see [Booth \(2006\)](#) and [Shang et al. \(2011\)](#). The vast majority of these techniques are based on period mortality: financial institutions are typically interested in the mortality developments of groups of individuals that comprise different birth cohorts. In addition, cohort data can be outdated, unavailable or incomplete, hence period life tables have been developed to analyse a hypothetical cohort as if its age-specific death rates pertained throughout its life ([Preston et al., 2001](#)).

The completion of the mortality experience of non-extinct cohorts is nonetheless interesting in the actuarial domain. Insurance companies are indeed interested in the future longevity of groups of people born in specific cohorts. In such settings, cohort forecasts are typically obtained by first forecasting mortality in a period fashion, and then extracting cohort mortality patterns from the diagonals of the projected Lexis surface. Although widely used, this approach seems rather counter-intuitive and inefficient. In this article, we propose a more direct and alternative approach to cohort mortality forecasting that is solely based on cohort data.

More generally, analysis and forecasts of cohort mortality are interesting and worth exploring for two main reasons. First, survival in real birth cohorts is different from survival in the hypothetical situation of unchanged period mortality rates because of: (i) tempo effects, (ii) cohort effects and (iii) selection (for a full discussion, see [Borgan and Keilman, 2019](#), pp. 90–92). Second, cohort mortality developments are *actually* observed, and they may differ from those of the synthetic cohorts assumed in period life tables. In other words, cohort life tables record information about what happened to real birth cohorts, whereas period life tables are models of what would happen to hypothetical cohorts if their age-specific period death rates remained constant throughout the life of the cohorts ([Preston et al., 2001](#)).

Analyses of age-cohort data have indeed provided different insights into mortality developments than studies based on the age-period perspective. For example, [Shkolnikov et al. \(2011b\)](#) have shown that best-practice cohort life expectancy for women born between 1870 and 1920 increased almost twice as fast as best-practice period life expectancy since 1840. Moreover, [Goldstein and Wachter \(2006\)](#) showed that, in populations experiencing steady mortality declines, period life expectancy can be regarded as a lagged indicator of cohort life expectancy. Finally, [Borgan and Keilman \(2019\)](#) have shown that the differences in period life expectancy between Japanese and Italian women on one side, and Scandinavian ones on the other, disappear or even reverse when considering cohort data: as such, they contend that these differences are caused by the distortion that period life tables imply in times of changing mortality. Given these considerations, forecasting cohort mortality from age-cohort data seems a more reasonable approach than extracting cohort patterns from age-period projections.

Models to forecast cohort mortality are relatively few in the literature. Among the first to use a cohort perspective, the [CMI Committee \(2007\)](#) employed the two-dimensional P -spline model of [Currie et al. \(2004\)](#) to complete the mortality experience of cohorts in England & Wales. Furthermore, [Chiou and Müller \(2009\)](#) proposed a functional data analysis approach for forecasting cohort log-hazard functions using Swedish mortality data. More recently, the combination of an EM algorithm with an eight-parameter model for the age-at-death distribution was suggested by [Zanotto and Mazzuco \(2017\)](#) for estimating deaths of non-extinct generations. Finally, [Rizzi et al. \(2019\)](#) proposed completing partially observed cohort age-at-death distributions using a penalized composite link model ([Eilers, 2007](#)), assuming a smooth underlying distribution over age.

In this article, we introduce a novel methodology to forecast adult cohort mortality. Rather than modelling mortality rates (the standard approach in mortality forecasting, as in, for example, the [Lee and Carter](#) model and its variants), our model is based on the distribution of deaths. Age-at-death distributions have recently received increasing attention in mortality forecasting ([Oeppen, 2008](#); [Bergeron-Boucher et al., 2017](#); [Basellini and Camarda, 2019a,b](#); [Pascariu et al., 2019](#)), as they provide a different and rather unexplored perspective on mortality developments that can be leveraged by forecasters. For this reason, we extend a newly introduced methodology to model and forecast adult age-at-death distributions ([Basellini and Camarda, 2019b](#)) with the aim of analysing and forecasting mortality developments across cohorts.

This paper is organized as follows. In Section 5.2, we describe the structure of cohort mortality data, and we illustrate mortality developments from the age-cohort perspective for one of the populations analysed in this article. In Section 5.3, we review the methodology proposed in this article, and the data used for the analyses. Section 5.4 presents two applications of our model to Swedish and Danish female adult mortality for the cohorts 1835–1970. In Section 5.5, we discuss the results of our methods and conclude.

5.2 Background

The structure of cohort mortality data differs from the conventional structure of age-period mortality analysis. Before introducing our methodology, it is convenient to describe the structure of the data used in this article, and to analyse mortality developments from the age-cohort perspective considered here.

Let $d_{x,c}$ and $e_{x,c}$ denote observed death counts and central exposures to the risk of death at age x for the birth cohort c , respectively. Data are arranged into two matrices $\mathbf{D} = (d_{x,c})$ and $\mathbf{E} = (e_{x,c})$, each of dimensions $m \times n$, where rows are classified by m single ages at death $\mathbf{x}' = [1, \dots, m]$, and columns by n single cohorts $\mathbf{c}' = [1, \dots, n]$. Unlike the case of age-period data, where data are generally fully observed, here the matrices \mathbf{D} and \mathbf{E} contain missing data, corresponding to periods (i.e. calendar years) beyond the last available year of data collection.

Displaying the matrix \mathbf{D} provides a clear understanding of the structure of cohort mortality data. Let \check{x} and \check{c} be the last age and cohort with fully observed data, respectively. Then, death

counts are available as follows, where “na” denotes the unobserved data:

$$\mathbf{D} = (d_{x,c}) = \begin{bmatrix} d_{1,1} & \dots & d_{1,\check{c}} & d_{1,\check{c}+1} & d_{1,\check{c}+2} & \dots & d_{1,n} \\ d_{2,1} & \dots & d_{2,\check{c}} & d_{2,\check{c}+1} & d_{2,\check{c}+2} & \dots & d_{2,n} \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ d_{\check{x},1} & \dots & d_{\check{x},\check{c}} & d_{\check{x},\check{c}+1} & d_{\check{x},\check{c}+2} & \dots & d_{\check{x},n} \\ d_{\check{x}+1,1} & \dots & d_{\check{x}+1,\check{c}} & d_{\check{x}+1,\check{c}+1} & d_{\check{x}+1,\check{c}+2} & \dots & na \\ \vdots & \dots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{m-2,1} & \dots & d_{m-2,\check{c}} & d_{m-2,\check{c}+1} & d_{m-2,\check{c}+2} & \ddots & na \\ d_{m-1,1} & \dots & d_{m-1,\check{c}} & d_{m-1,\check{c}+1} & na & \dots & na \\ d_{m,1} & \dots & d_{m,\check{c}} & na & na & \dots & na \end{bmatrix}.$$

$\underbrace{\hspace{10em}}$ Fully observed $\underbrace{\hspace{10em}}$ Partially observed

The matrix clearly shows that, for the cohorts after \check{c} , data are increasingly unobserved from the last age downwards. Exposure data in \mathbf{E} are unobserved for the same elements of \mathbf{D} .

The analysis of mortality developments over ages and cohorts will therefore inevitably display this data structure. Figure 5.1 shows an example of such structure: observed adult age-at-death distributions for Swedish females are illustrated for selected cohorts between 1835–1970 (data retrieved from the [Human Mortality Database, 2019](#)). For this population, the last cohort with fully observed data is 1906; more recent cohorts are thus characterized by a decreasing amount of observed data, as shown by the graphs.

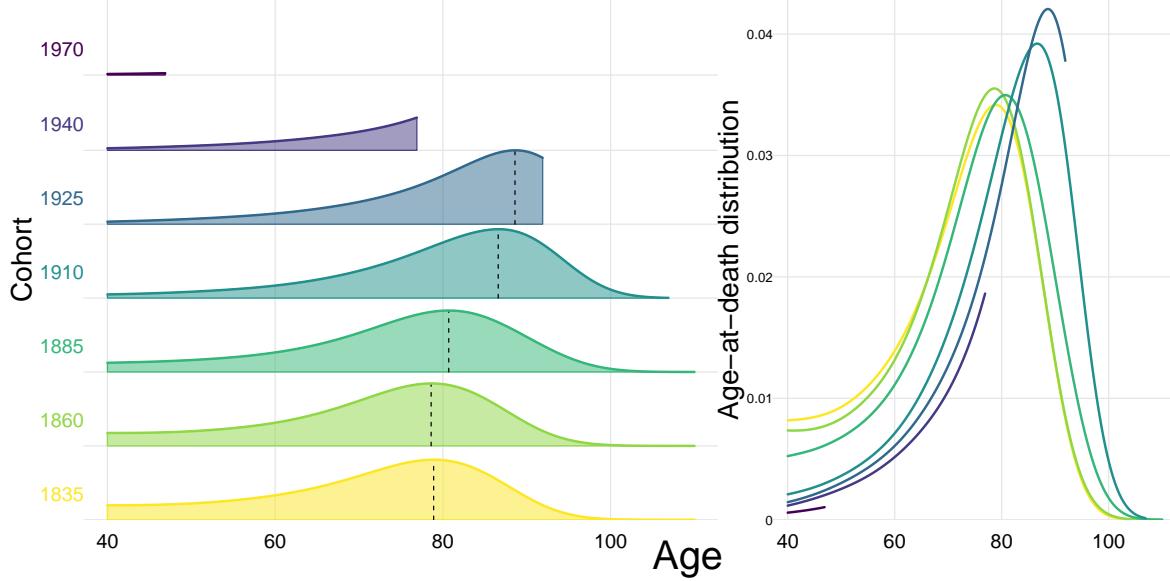


Figure 5.1: Observed age-at-death distributions for Swedish females aged 40–110+ for selected cohorts between 1835–1970. Dashed black lines correspond to modal ages at death. Data have been smoothed for illustrative purposes.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

Importantly, Figure 5.1 further motivates the development of a methodology that can leverage the features and the changes of age-at-death distributions with the purpose of modelling and forecasting human mortality. Clearly intelligible and demographically meaningful mortality developments readily emerge from this Figure. The left panel shows the increase of the modal

age at death for the cohorts born after the 1860. In the right panel, the same distributions of the left panel are plotted over a common y -axis: decreasing age-at-death variability is evident for more recent cohorts.

Age-at-death distributions thus provide informative insights into mortality patterns and development of the population, and they further allow one to study the shifting and compression dynamics of mortality changes (see, e.g., Fries, 1980; Kannisto, 2001; Bongaarts, 2005; Janssen and de Beer, 2019). We therefore propose an approach to model and forecast adult cohort mortality that is based on age-at-death distributions. A relational model inspired by the seminal contribution of Brass (1971) serves our purposes well. Specifically, the combination of a reference distribution and its transformation over different cohorts allows us to simultaneously: (i) fit the observed data, and (ii) obtain estimates for the unobserved data, thereby completing the mortality experience of partially observed cohorts. In the following Section, we introduce our proposed methodology.

5.3 Methods & Data

5.3.1 The C-STAD model

Suppose we have two adult age-at-death distributions defined on the age range $x \geq 40$. Specifically, let $f(x)$ be a “standard”, or reference, distribution and $g(x)$ an observed distribution. Let $t(x; \boldsymbol{\theta})$ be a transformation function of the age axis and a vector of parameters $\boldsymbol{\theta}$ such that:

$$g(x) = f[t(x; \boldsymbol{\theta})], \quad (5.1)$$

i.e. the distribution $g(x)$ is derived from a warping transformation of the age axis of $f(x)$. Specifically, the term “warping” originates in Functional Data Analysis (Ramsay and Silverman, 2005) and refers to the transformation of a time axis to achieve close alignment of functions.

We propose a parsimonious yet flexible transformation function $t(x; \boldsymbol{\theta})$ that rigorously captures adult mortality developments across cohorts. Let $\boldsymbol{\theta}' = [s, b_L, c_L, d_L, b_U]$ be a vector containing the model’s parameters, where $s = M^g - M^f$ is the difference between the modal ages at death of the unimodal distributions $g(x)$ and $f(x)$. The proposed *Cohort Segmented Transformation Age-at-death Distributions* (C-STAD) model can be written as:

$$t(x; \boldsymbol{\theta}) = \begin{cases} M^f + b_L \tilde{x} + c_L \tilde{x}^2 + d_L \tilde{x}^3 & \text{if } x \leq M^g \\ M^f + b_U \tilde{x} & \text{if } x > M^g \end{cases} \quad (5.2)$$

where $\tilde{x} = x - s - M^f$, and the subscripts L and U refer to the *Lower* and *Upper* parts of the segmented transformation (i.e. before and after M^g), respectively.

The warping function $t(x; \boldsymbol{\theta})$ takes the form of a segmented transformation model which breaks at the value of M^g . Below M^g , the transformation function is cubic, while it is linear above M^g . Although acting on $t(x; \boldsymbol{\theta})$, the model’s parameters are directly related to the summary measures of the age-at-death distributions. Specifically, while s captures the difference in modal ages between $g(x)$ and $f(x)$, b_L and b_U measure the change in variability before and after the modal ages of the two distributions. For the ages below M^g , c_L and d_L further measure differences

in terms of asymmetry and heaviness of the left tail between $g(x)$ and $f(x)$, respectively. In terms of mathematical moments, the parameters b_L and b_U can be related to the variance of the distribution before and after the mode, while c_L and d_L relate to the skewness and kurtosis of the distribution.

Figure 5.2 provides a graphical illustration of the C-STAD model. For ease of presentation, we start from the simpler case in which $b_L = b_U = 1$ and $c_L = d_L = 0$. Substitution of these parameters in Eq. (5.2) yields a unique transformation function $t(x; \theta) = x - s$, and a corresponding distribution $g_1(x) = f(x - s)$ via Eq. (5.1). In the left panel of Fig. 5.2, the standard distribution (grey line) is shifted to the right by an amount equal to s , and $g_1(x)$ (blue line) maintains the original shape of $f(x)$. The right panel shows the transformation function related to this plain shifting scenario. Note that a left-shift could be simply obtained with a negative value of s .

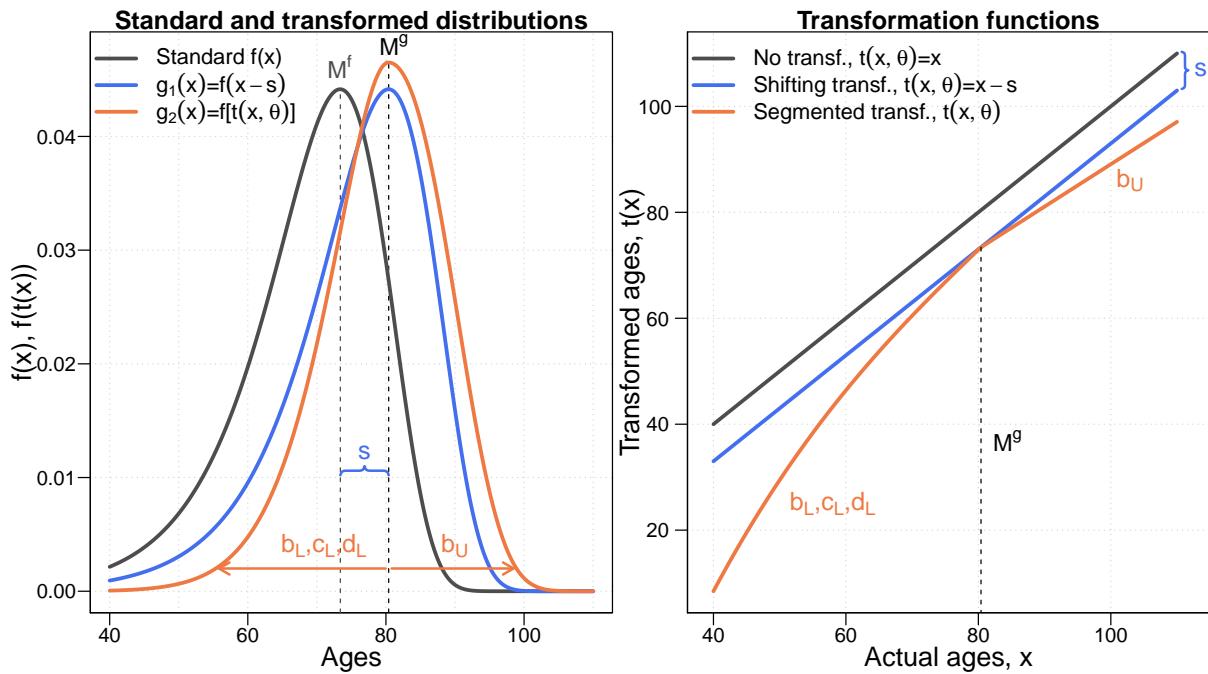


Figure 5.2: A schematic overview of the *Cohort Segmented Transformation Age-at-death Distributions* model.

Source: Authors' own elaborations.

Different parameters' values allow one to capture broader mortality developments than the shifting scenario described above. While b_L and b_U modify the variability of the distribution $g_2(x) = f[t(x; \theta)]$ (orange line, left panel) w.r.t. $f(x)$ before and after the modal age at death, c_L and d_L affect the asymmetry and heaviness of the left tail of $g_2(x)$ as compared to $f(x)$. In the example shown in Fig. 5.2, $b_L > 1$ reduces the variability of $g_2(x)$ before M^g w.r.t. $f(x)$, while $b_U < 1$ increases the variability of $g_2(x)$ after M^g w.r.t. $f(x)$. The effects of c_L and d_L are difficult to discern from the left panel. However, the right panel shows the warping transformation $t(x; \theta)$ applied to $f(x)$ to derive $g_2(x)$; the transformation (orange line) is composed of a cubic function (due to non-zero values of c_L and d_L) before the cut-off point M^g , and a linear function above M^g .

5.3.2 Data

For illustrative purposes we present outcomes from the proposed model on adult cohort mortality for females in two high-longevity countries, namely Sweden and Denmark. Long-term series of high quality data are available for both countries, even at very old ages ([Vaupel and Lundström, 1994](#); [Wilmoth and Lundström, 1996](#); [Andreev, 2002](#)), and the two countries display different mortality developments. We therefore test the goodness-of-fit and forecast accuracy of our model with respect to different mortality trajectories ([Christensen et al., 2010](#)). The data are derived from the [Human Mortality Database](#) (HMD, 2019), which provides free access to detailed, consistent and high quality historical mortality data for 41 different areas and countries ([Barbieri et al., 2015](#)).

Our interest in this article is restricted to the senescent component of mortality, hence we start our analyses from age 40. We therefore cover the age range that is of greater interest for pension and insurance funds. Specifically, we work with two $m \times n$ matrices $\mathbf{D} = (d_{x,c})$ and $\mathbf{E} = (e_{x,c})$, defined for ages $\mathbf{x}' = [40, \dots, 110+]$ and cohorts $\mathbf{c}' = [1835, \dots, 1970]$. Figure 5.3 offers a schematic overview of the data structure by means of two Lexis diagrams: the first (left panel) shows the conventional age-period structure, while the second (right panel) illustrates the age-cohort perspective that we adopt in this paper. On the one hand, we select 1835 as starting cohort of analysis for both populations because it is the first cohort with observed data at all ages in Denmark, and to compare the results for the two countries using the same fitting period. On the other hand, 1970 is the final cohort because it contains enough observed data points in both countries (seven in Sweden, six in Denmark) to accurately estimate the three parameters of the lower part of the transformation function in Eq. (5.2) (i.e. b_L , c_L and d_L , cf. Subsection 5.3.4).

Estimation and forecasting of the C-STAD parameters (Subsection 5.3.4) is performed on three different groups of cohorts:

$$\mathbf{c}' = [\underbrace{1835, \dots, \check{c}}_{\mathbf{c}_1}, \underbrace{\check{c} + 1, \dots, \tilde{c}}_{\mathbf{c}_2}, \underbrace{\tilde{c} + 1, \dots, 1970}_{\mathbf{c}_3}] . \quad (5.3)$$

Therefore the data are partitioned as follows:

$$\mathbf{D} = [\mathbf{D}_1 : \mathbf{D}_2 : \mathbf{D}_3] \quad \mathbf{E} = [\mathbf{E}_1 : \mathbf{E}_2 : \mathbf{E}_3] . \quad (5.4)$$

The first group, denoted by \mathbf{c}_1 , contains the fully observed cohorts $1835, \dots, \check{c}$, where \check{c} corresponds to the last cohort for which all data have been observed. As such, \mathbf{D}_1 and \mathbf{E}_1 have been observed at all ages x for all cohorts in \mathbf{c}_1 . The second group, denoted by \mathbf{c}_2 , is composed by cohorts $\check{c} + 1, \dots, \tilde{c}$, where \tilde{c} corresponds to the last cohort for which two age-groups above the adult modal age at death have been observed. In other words, \mathbf{D}_2 and \mathbf{E}_2 are incomplete, i.e. data are not available for higher ages and more recent cohorts. However this group of cohorts is selected such that associated $d_{x,c}$ and $e_{x,c}$ have been observed for at least two data points above the modal age $x = M$ for all cohorts in \mathbf{c}_2 . The choice of having two age groups above M is imposed by the estimation of the parameter above the mode (cf. Subsection 5.3.4). Finally, the third group \mathbf{c}_3 is composed of the remaining cohorts $\tilde{c} + 1, \dots, 1970$, in which data are only

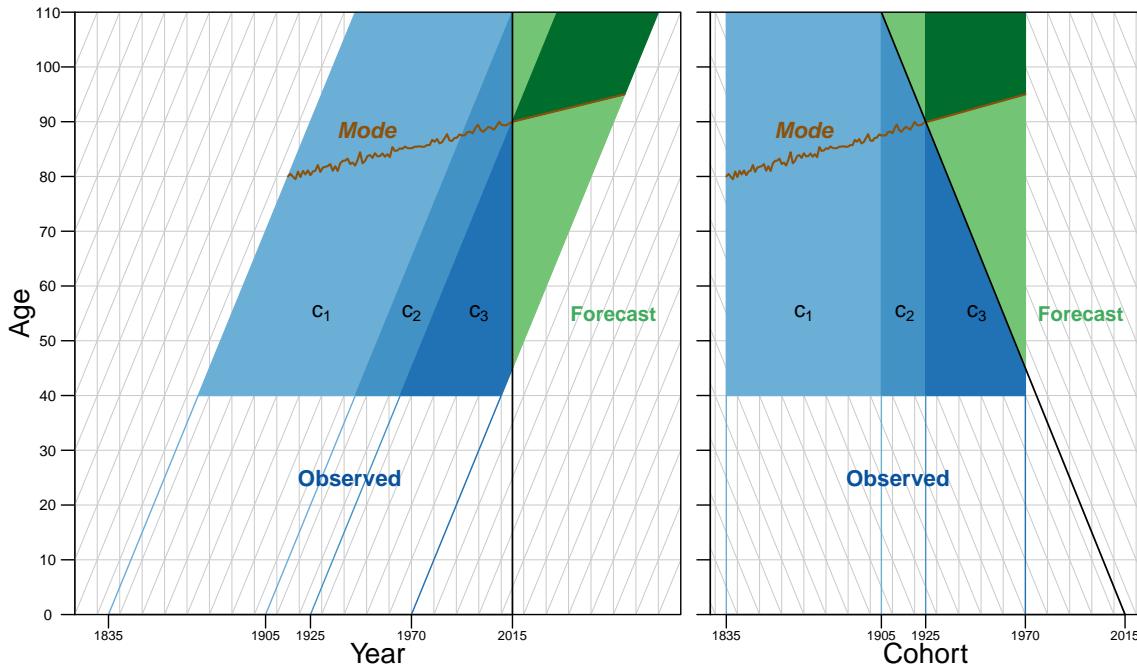


Figure 5.3: Conventional age-period (left panel) and age-cohort (right panel) Lexis diagrams illustrating the data structure and the division of cohorts into three groups. Here we assume that: (i) 2015 is the most recent year of data collection, (ii) $\check{c} = 1905$, and (iii) $\check{c} = 1925$. The three groups are then $c'_1 = [1835, \dots, 1905]$, $c'_2 = [1906, \dots, 1925]$ and $c'_3 = [1926, \dots, 1970]$. The two colours in the forecast years correspond to different parameters' derivations (cf. Subsection 5.3.4): estimation with incomplete data (light green) and forecasting (dark green).

Source: Authors' own elaborations.

partially available and modal age at death is not observed. An illustration of the divisions of cohorts into the three groups is provided in Figure 5.3. Figure 5.4 shows an example of the observed and missing data for three age-at-death distributions belonging to the different groups of cohorts.

5.3.3 The standard distribution

The first step in the estimation of the C-STAD model is the derivation of the standard distribution $f(x)$. The C-STAD can be interpreted as a relational model (Brass, 1971), hence it is desirable to include the representative features of the observed data for all cohorts in the computation of $f(x)$. Meanwhile, we also wish to remove the small random fluctuations that characterize the mortality pattern of age-at-death distributions. To achieve both goals at the same time, we derive the age-at-death distribution for each cohort 1835–1970 by a two-dimensional (2D) P -splines smoothing approach to cohort mortality (Eilers and Marx, 1996; Currie et al., 2004). Specifically, we assume that observed death counts $d_{x,c}$ at given age x and cohort c are realizations of the random variable $D_{x,c}$ which follows a Poisson distribution (Brillinger, 1986):

$$D_{x,c} \sim \mathcal{P}(e_{x,c} \mu_{x,c}), \quad (5.5)$$

where exposures-to-risk $e_{x,c}$ are given and $\mu_{x,c}$ denote the hazard or force of mortality (such as in, for example, Brouhns et al., 2002). We smooth observed death counts using a tensor product of B -

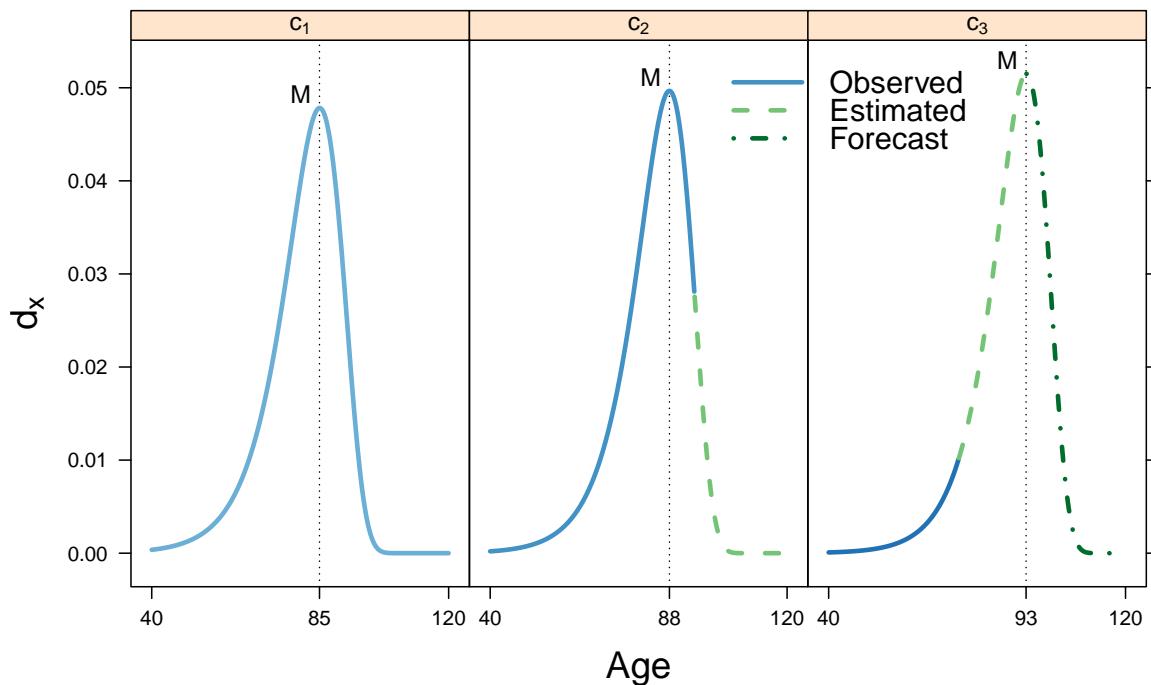


Figure 5.4: Example of observed, estimated and forecast data for three age-at-death distributions belonging to the groups of cohorts c_1 , c_2 and c_3 .

Source: Authors' own elaborations.

spline bases over ages and cohorts, and exposures as an offset. To account for incomplete cohorts and still preserve the rectangular structure in the data, we include regression weights $\mathbf{W} = (w_{x,c})$ whose elements are equal to one if the corresponding death counts $d_{x,c}$ and exposures $e_{x,c}$ have been observed, and zero otherwise. Smoothing parameters over ages and cohorts are chosen by Bayesian Information Criterion minimization. The R package `MortalitySmooth` ([Camarda, 2012](#)) provides a direct implementation of this procedure.

The estimated smooth mortality surface allows us to derive smooth (partial) distributions. To compute the standard, we first employ a landmark registration procedure, a technique often used in Functional Data Analysis with the aim of aligning important features, or landmarks, of the curves analysed ([Ramsay and Silverman, 2005](#)). For age-at-death distributions, the modal age at death is an obvious landmark. As such, we align the distributions so that their modal ages are equal to the mode of the first observed distribution (the 1835 cohort), maintaining their variability unchanged. In practice, the alignment is achieved by a plain shifting transformation of the observed distributions, which preserves all their features except the modal age-at-death.

Having aligned the observed distributions to a common modal age, we derive the standard $f(x)$ as the mean of the aligned distributions. Figure 5.9 in Appendix 5.6.2 illustrates the alignment procedure and the derivation of the standard distribution for the Swedish female population analysed in our article. This landmark registration procedure enhances the representativeness of $f(x)$ while improving the goodness-of fit of the model (for additional details, see [Basellini and Camarda, 2019b](#)). Importantly, it should be noted that for cohorts in c_2 and c_3 , we only use the part of the aligned distribution corresponding to the observed data (i.e. where regression weights are not zero).

Finally, we express the standard $f(x)$ as a linear combination of equally spaced B -spline bases $\mathbf{B}(x)$ over ages x and coefficients $\boldsymbol{\beta}_f$ specific to the standard:

$$f(x) = \exp [\mathbf{B}(x) \boldsymbol{\beta}_f]. \quad (5.6)$$

In this last step, we chose a generous number of B -splines with a small penalty term. This whole procedure allows us to preserve all important features in the standard distribution (embodied in $\boldsymbol{\beta}_f$) after having removed unnecessary random fluctuations from the original data. In addition, the smoothing approach further allows us to evaluate mortality at any finer scale of the age axis, practically at a continuous level.

5.3.4 Estimation and forecast of the C-STAD parameters

Given the estimated standard distribution, we can derive the C-STAD parameters $\boldsymbol{\theta}' = [s, b_L, c_L, d_L, b_U]$ for each cohort in 1835–1970. As anticipated in Subsection 5.3.2, we use three different approaches to estimate $\boldsymbol{\theta}$, depending on the data available for each cohort; we thus divide cohorts into three groups, as shown in Figure 5.3.

For the first group of fully observed cohorts c_1 , we start by estimating the parameters vector \boldsymbol{s} . To properly capture cohort-specific mortality fluctuations, we employ a one-dimensional P -spline approach, i.e. we smooth mortality for each cohort independently, numerically compute the corresponding density and extract the modal ages at death for each cohort (for a similar approach in a period perspective, see [Ouellette and Bourbeau, 2011](#)). From the modal ages we estimate the parameter $\hat{\boldsymbol{s}} = (\hat{s}_c = M_c - M_f)$ over cohorts in c_1 , where M_f denotes the mode of the standard distribution, which by construction corresponds to the modal age at death for the cohort born in 1835.

Having derived an estimate of the shifting parameter $\hat{\boldsymbol{s}}$, we can estimate the remaining parameters $\boldsymbol{\alpha}' = [b_L, c_L, d_L, b_U]$. We take advantage of the Poisson assumption in Eq. (5.5) and maximise the following log-likelihood function:

$$\ln \mathcal{L}_{\boldsymbol{\alpha}} (\boldsymbol{\alpha} | d_{x,c}, e_{x,c}, w_{x,c}, \hat{s}_c, \boldsymbol{\beta}_f) \propto \sum_x w_{x,c} [d_{x,c} \ln (\mu_{x,c}^{\text{C-STAD}}) - e_{x,c} \mu_{x,c}^{\text{C-STAD}}] \quad (5.7)$$

for each cohort in $c'_1 = [1835, \dots, \check{c}]$, where regression weights $w_{x,c}$ are zero in the case of unobserved data at the highest ages, and $\mu_{x,c}^{\text{C-STAD}}$ denotes the estimated hazard of the C-STAD model. In words, the optimization procedure looks for a combination of parameters $\hat{\boldsymbol{\alpha}}$ that produces, for each cohort, an age-at-death distribution whose corresponding hazard maximises the log-likelihood in Eq. (5.7). The associated age-at-death distribution $\hat{g}_c(x)$ can be written as follows:

$$\hat{g}_c(x) = \exp [B(x_t) \boldsymbol{\beta}_f] \quad \text{where } x_t = t(x; \hat{s}_c, \hat{\boldsymbol{\alpha}}). \quad (5.8)$$

The hazard $\hat{\mu}_c(x)$ corresponding to $\hat{g}_c(x)$ is computed using standard life-table and survival analysis formulas:

$$\hat{\mu}_c(x) = \frac{\hat{g}_c(x)}{\hat{\ell}_c(x)} = \frac{\hat{g}_c(x)}{\int_x^\infty \hat{g}_c(t) dt}, \quad (5.9)$$

where $\hat{\ell}_c(x)$ denotes the life-table probability of surviving to age x (i.e. the survival function, Preston et al., 2001; Klein and Moeschberger, 2003), which can be computed numerically from the smooth distribution $\hat{g}_c(x)$, evaluated at an extremely fine level.

For the second group of partially observed cohorts c_2 , we start again from the shifting parameter s . We use the same estimation approach used in c_1 : data are available until the ages above the mode, therefore the smoothing approach produces an estimate of M_c and \hat{s} over cohorts in c_2 . With respect to the remaining parameters, we also follow the same approach: we maximize Eq. (5.7) for each cohort in $c'_2 = [\tilde{c} + 1, \dots, \tilde{c}]$, the only difference being that zero regression weights correspond to the missing data above the mode of the partially observed cohorts. It should be noted here that the unobserved data only influence the estimation of b_U , as complete data are observed below the mode for all cohorts in this group.

For the third group of partially observed cohorts c_3 , we employ a mixture of forecasting and estimation to determine the C-STAD parameters. The lack of data above the modal age at death makes it impossible to estimate the parameter s and compute the log-likelihood in Eq. (5.7). Hence, we start from the time-series of the estimated parameters \hat{s} and \hat{b}_U over cohorts in c_1 and c_2 to compute their forecasts for cohorts c_3 .

From a theoretical perspective, these two parameters are related by the fact that only mortality changes occurring above the mode can modify its value (cf. Appendix B in Canudas-Romo, 2010). Correlation analyses for the two countries under study confirm the strong relation between the two series (Pearson correlation of 0.96 and 0.90 for the time-series in first differences for Sweden and Denmark, respectively). As such, we specify a vector autoregressive (VAR) model of order one with a constant for the two (differenced) parameters, and we forecast their values for all cohorts c_3 . The R package `vars` allows us to perform model selection and estimation (Pfaff, 2008a,b).

Then, we take the forecast values of \hat{s} and \hat{b}_U as given, and we estimate the remaining parameters $\check{\alpha}' = [b_L, c_L, d_L]$ by maximizing the log-likelihood:

$$\ln \mathcal{L}_{\check{\alpha}} (\check{\alpha} | d_{x,c}, e_{x,c}, w_{x,c}, \hat{s}_c, \hat{b}_{U_c}, \beta_f) \propto \sum_x w_{x,c} \left[d_{x,c} \ln \left(\mu_{x,c}^{\text{C-STAD}} \right) - e_{x,c} \mu_{x,c}^{\text{C-STAD}} \right] \quad (5.10)$$

for each cohort in $c'_3 = [\tilde{c} + 1, \dots, 1970]$. In contrast to the estimation procedure in c_2 , here the unobserved data influence the estimation of the parameters associated to the ages below the modal age at death.

The estimate $\hat{\theta}$ for each cohort in 1835–1970 allows us to derive a complete set of age-specific mortality measures, i.e. we can complete the mortality experience for the partially observed cohorts of our analysis. In order to derive the C-STAD confidence intervals (CI)¹, we employ a bootstrapping procedure (Efron and Tibshirani, 1994). As suggested by Keilman and Pham (2006), we consider the uncertainty related to: (i) the estimated parameters, and (ii) the forecast values of s and b_U . The first source of uncertainty is accounted for by generating bootstrap death counts from the C-STAD deviance residuals (as in, for example, Koissi et al., 2006; Renshaw and Haberman, 2008; Ouellette et al., 2012). Appendix 5.6.1 provides more details on the computation of deviance residuals and bootstrap death counts. The second source of uncertainty

¹to avoid confusion, we use the general term CI for all cohorts analysed, even when intervals are constructed from the mixture of forecast and estimated parameters (i.e. cohorts c_3).

is considered by simulating future values of the VAR model. We employ 40 different matrices of bootstrap death counts, and for each of these, we refit the C-STAD model and simulate 40 future values of \mathbf{s} and \mathbf{b}_U . From the 1600 resulting simulations, we take the lowest and highest deciles to construct 80% pointwise confidence intervals.

Finally, routines developed to fit and forecast the C-STAD model were implemented in R (R Development Core Team, 2019) and are publicly available, and all the results presented in the following Section are fully reproducible at <https://github.com/ubasellini/C-STAD> [this GitHub repository will be made public upon eventual acceptance of the manuscript].

5.4 Results

5.4.1 Out-of-sample validation of the C-STAD model

Before estimating the proposed C-STAD model to complete partially observed cohorts, we first assess the accuracy of the C-STAD model by performing six predictive out-of-sample validation exercises on Swedish and Danish adult females. We compare the forecast life expectancy at age 40 (e_{40}) and the Gini coefficient at age 40 (G_{40}) with the observed out-of-sample values. Both measures of longevity (the former) and lifespan inequality (the latter) are useful to evaluate the accuracy of mortality forecasts (Bohk-Ewald et al., 2017).

Formally, the Gini coefficient at age 40 is defined as:

$$G_{40} = 1 - \frac{1}{e_{40} (\ell_{40})^2} \int_{40}^{\omega} [\ell(x)]^2 dx, \quad (5.11)$$

where ℓ_{40} is the life-table radix, which we set equal to one without loss of generality, and ω is the highest age attained in the population (Hanada, 1983; Shkolnikov et al., 2003).

Originally proposed to measure income or wealth inequality (Gini, 1912, 1914), the Gini coefficient is today one of the most common statistical indices employed for measuring concentration in the distribution of a positive random variable. In recent years, the coefficient has been used to measure lifespan inequality within and between populations (see, e.g., Shkolnikov et al., 2003; Smits and Monden, 2009; van Raalte and Caswell, 2013; Gigliarano et al., 2017) and to evaluate mortality forecasts (Diaz et al., 2018; Basellini and Camarda, 2019b). The coefficient takes values between 0 and 1, which correspond to the limit cases of perfect equality and perfect inequality, respectively. For an age-at-death distribution, Gini is equal to zero if all individuals die at the same age, and equal to 1 if all people die at age 0 and one individual dies at a positive age (Shkolnikov et al., 2003). In this article, we compute G_{40} in Eq. (5.11) with the approximation formula proposed by Shkolnikov et al. (2003), and we multiply G_{40} by 100 in order to have a comparable magnitude with e_{40} .

Specifically, we pretend that the last year of collected data is $2015 - h$, where $h = 10, 15, 20, 25, 30$ and 35 years. We then fit the C-STAD model to the fully observed cohorts $\mathbf{c}'_1 = [1835, \dots, 1905 - h]$, and we forecast mortality h years ahead. If the modal age is always observed for the cohorts to be completed, the exercise is restricted to the group \mathbf{c}_2 . Otherwise, the group \mathbf{c}_3 is considered too.

An explicative example of the validation procedure is useful to clarify the out-of-sample exercises. Let us consider $h = 10$: then, the last year of fully observed data is 2005. We fit the C-STAD to the fully observed cohorts $\mathbf{c}'_1 = [1835, \dots, 1895]$, and we forecast 10-year ahead. By doing so, we complete the mortality experience of the partially observed cohorts $1896, \dots, 1905$, and for each of these, we compute and compare the estimated e_{40} and G_{40} with their observed values.

It is worth mentioning at this point that, for the lower values of h , forecasting is achieved simply by fitting the C-STAD on the partially observed cohorts \mathbf{c}_2 . In the explicative example above, where the last data available occurred in 2005, the cohort 1896, for instance, has been observed at all ages except 110. We thus take advantage of the nature of cohort data and consider all possible observations to complete the mortality experience of this partial cohort. Conversely, for higher values of h , forecasting is achieved by considering also the cohorts \mathbf{c}_3 , which require the combination of forecasting and estimation of the C-STAD parameters.

In addition to the C-STAD, we perform the same out-of-sample exercises with the 2D P -splines approach of [Currie et al. \(2004\)](#). This is the only model that, to our knowledge, has been employed to forecast cohort mortality from a cohort perspective ([CMI Committee, 2007](#)) and can be implemented in the R software after data manipulation (in the `MortalitySmooth` package, [Camarda, 2012](#)).

Furthermore, we compare our results with the standard procedure of: (i) forecasting mortality in an age-period fashion, and (ii) extracting cohort patterns from the diagonals of the age-period mortality surface. For these comparisons, we employ the benchmark model of [Lee and Carter \(LC, 1992\)](#). For consistency, in each exercise we select the shortest fitting period that includes all the cohorts needed to assess the LC forecasts. Fitting periods for the six exercises are 1936–2005, 1931–2000, 1926–1995, 1921–1990, 1916–1985 and 1911–1980: the starting year is computed by adding 40 (the starting age of the analysis) to the first forecast cohort, while the last year is given by $2015 - h$. It should be noted that the fitting age-period surface for these exercises was derived from the original age-cohort data, to exclude potential bias related to differences in the computation of period vs. cohort mortality rates (see [Wilmoth et al., 2019](#), pp. 29–33).

Table 5.1 presents the results of our analysis. The first and second columns contain the cohorts used for fitting and forecasting the C-STAD and 2D P -splines models, respectively. The third column contains the forecast horizon of the out-of-sample exercise, while the fourth column indicates the measure analysed (e_{40} and g_{40}). Results are shown in the last six columns. We assess the accuracy of the point forecasts by computing the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{h} \sum_{c=1}^h (\hat{y}_c - y_c)^2},$$

where h is the forecasting horizon, and \hat{y}_c and y_c are the forecast and observed out-of-sample values of either e_{40} or G_{40} .

The table shows that the C-STAD forecasts are accurate in completing the mortality experience of partially observed cohorts. The RMSE values of both e_{40} and g_{40} are low across the six exercises, and they do not increase significantly with the forecasting horizon. Additionally,

Fitting cohorts	Forecast cohorts	Horizon	Measure	Sweden			Denmark		
				C-STAD	2D P-spline	LC (period)	C-STAD	2D P-spline	LC (period)
1835–1895	1896–1905	10y	e_{40} G_{40}	0.08 0.09	0.08 0.10	0.22 0.32	0.08 0.08	0.08 0.08	0.41 0.38
1835–1890	1891–1905	15y	e_{40} G_{40}	0.07 0.08	0.09 0.10	0.26 0.40	0.07 0.07	0.08 0.12	0.41 0.44
1835–1885	1886–1905	20y	e_{40} G_{40}	0.05 0.09	0.08 0.11	0.35 0.53	0.06 0.06	0.08 0.12	0.44 0.45
1835–1880	1881–1905	25y	e_{40} G_{40}	0.04 0.10	0.08 0.11	0.41 0.57	0.03 0.11	0.08 0.14	0.47 0.47
1835–1875	1876–1905	30y	e_{40} G_{40}	0.06 0.10	0.09 0.14	0.48 0.63	0.03 0.11	0.11 0.19	0.52 0.52
1835–1870	1871–1905	35y	e_{40} G_{40}	0.14 0.04	0.08 0.13	0.56 0.59	0.05 0.06	0.14 0.22	0.55 0.49

Table 5.1: Root mean square error (RMSE) of the C-STAD, 2D P -spline and LC (period) forecasts of e_{40} and G_{40} for adult females in Sweden and Denmark in six out-of-sample validation exercises: forecast horizon of 10, 15, 20, 25, 30 and 35 years. Lower values of the RMSE (in bold, assessed using all available decimals) correspond to greater forecast accuracy.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

C-STAD forecasts are more accurate than those of the 2D P -spline model, and both age-cohort models significantly outperform the standard LC age-period based approach. Employing different fitting periods in the age-period exercises, such as starting the analysis from the same year (for example, 1911) in all cases, would result in even greater RMSE for the LC forecasts for the shorter horizon exercises. Very similar results are obtained by employing different prediction accuracy measures, such as the MAPE and MAE (see Appendix 5.6.2).

5.4.2 Mortality developments for Swedish and Danish females, cohorts 1835–1970

In this Subsection we show the results of employing the C-STAD model to estimate and forecast adult female cohort mortality in Sweden and Denmark for the cohorts 1835–1970. The estimated and forecast parameters are shown in Appendix 5.6.2. Figure 5.5 shows the observed and fitted remaining life expectancies at age 40 (e_{40}) and Gini coefficient at age 40 (g_{40}) in the two population analysed for the fully observed cohorts c_1 (1835– \check{c} , where \check{c} is 1906 for Sweden and 1905 for Denmark). The two graphs provide evidence on the goodness-of-fit of the C-STAD model, whose estimates are very close to the observed values for both measures in the two populations. Inspection of the deviance residuals (shown in Appendix 5.6.2) provides additional evidence for the adequacy of the C-STAD model.

Figure 5.6 shows the observed (cohorts c_1) and completed (c_2 and c_3) e_{40} and g_{40} computed with the C-STAD (with 80% pointwise confidence intervals) and 2D P -spline model for the two population analysed. Despite sharing similar country trends in the fully observed cohorts c_1 , it is interesting to observe the different mortality developments in the partially observed cohorts c_2 : while Swedish adult females show continuous improvements in longevity and lifespan equality, Danish ones display a stagnation of e_{40} and an increase in lifespan inequality. The trends of the mortality measures for the partially observed cohorts are similar across the models, with the exception of Danish e_{40} : the increase of the 2D P -spline forecast e_{40} is much faster than

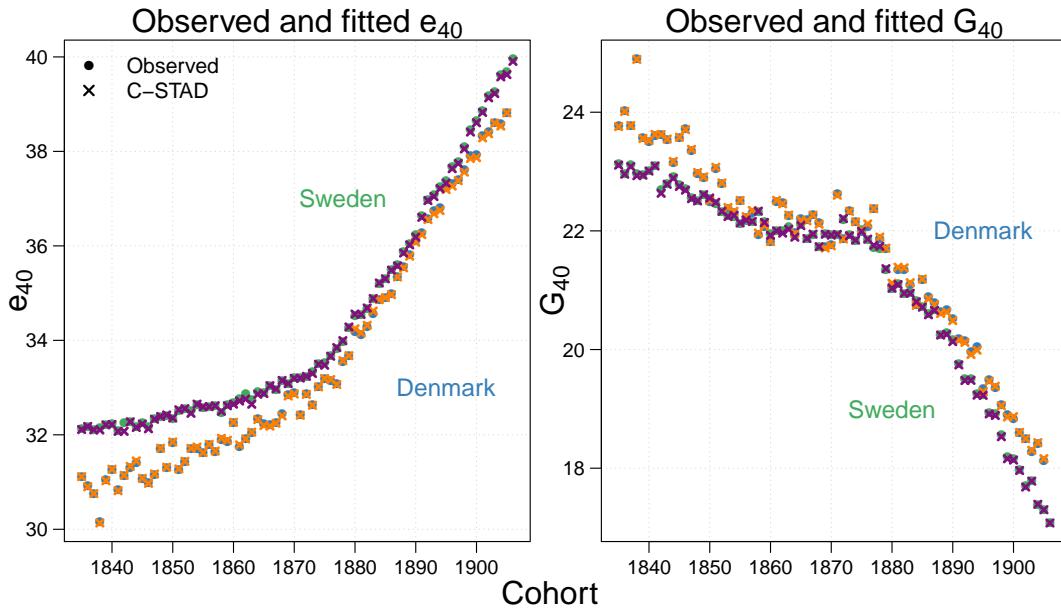


Figure 5.5: Observed and C-STAD estimated remaining life expectancies at age 40 (e_{40} , left panel) and Gini coefficient at age 40 (g_{40} , right panel) for adult females in Sweden and Denmark for the fully observed cohorts 1835– \tilde{c} (where \tilde{c} is 1906 for Sweden and 1905 for Denmark).

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

for the C-STAD model, resulting in a crossover between the two populations. Moreover, it is interesting to observe that the C-STAD confidence intervals are rather narrow for both countries in c_2 (as the great majority of data is observed for these cohorts), while they increase in the cohorts c_3 proportional to the amount of missing data. Note that \tilde{c} is 1925 and 1927 for Sweden and Denmark, respectively.

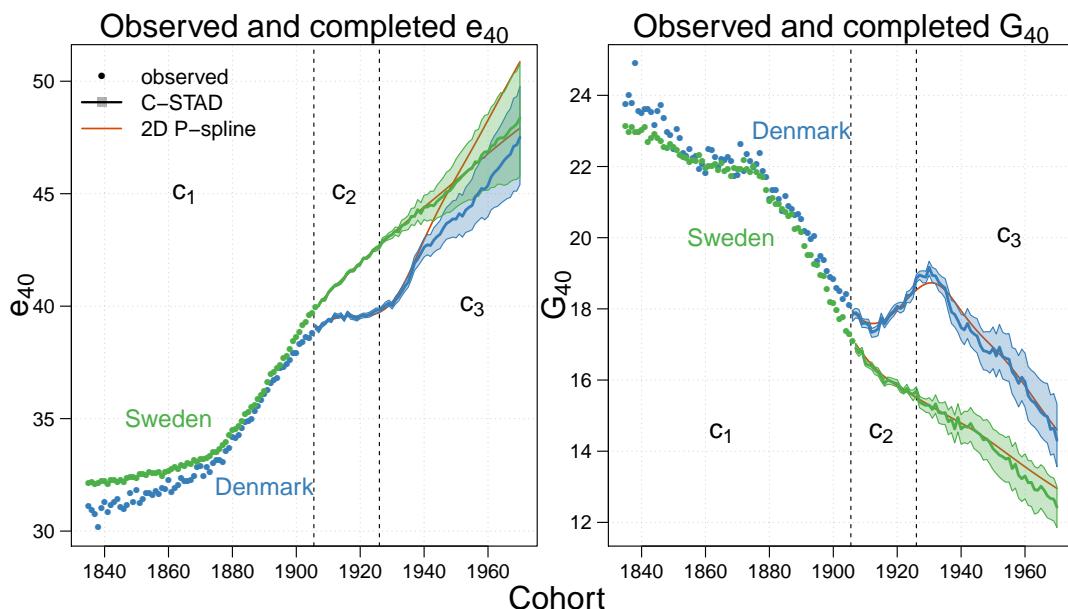


Figure 5.6: Observed (cohorts c_1) and completed (c_2 and c_3) remaining life expectancies at age 40 (e_{40} , left panel) and Gini coefficient at age 40 (g_{40} , right panel) for the C-STAD (with 80% confidence intervals) and 2D P-spline models for adult females in Sweden and Denmark for the cohorts 1835–1970.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

The age-specific mortality rates analysis shown in Figure 5.7 offers additional insights into cohort mortality developments for the two populations. In the top panels, observed, fitted and forecast mortality rates over cohorts are shown for some selected ages. In addition to the goodness-of-fit of the C-STAD model, the graphs highlight diverse age-specific developments in the two countries: for example, mortality at ages 40 and 60 for Danish cohorts born at the beginning of the twentieth century did not improve, resulting in the atypical trends of the summary measures shown in Figure 5.6 (stagnation of e_{40} and increase of g_{40}). In the bottom panels, mortality rates over all ages are shown for some selected cohorts. This second perspective shows how the shape of the mortality curve, appropriately captured by the C-STAD model, changed over time: for example, mortality at young adult ages was still relatively high in both countries for the 1835 cohort, with the curve being rather flat in the age range 40–50. The subsequent mortality decline at all ages, mainly attributable to improvements in sanitary environment, public hygiene and nutrition (McKeown, 1976), clearly emerges from Figure 5.7. An additional interesting observation is that the confidence intervals of the C-STAD widen as expected: for example, variability increases with age for the completed cohorts, as fewer age-specific data have been observed at higher ages. Finally, Figure 5.8 shows the observed and C-STAD age-at-death distributions for the three cohorts analysed in the previous panels.

5.5 Discussion

Mortality forecasting has drawn considerable interest in recent decades among academics and financial sector practitioners due to the increasing challenges posed by population ageing. Advances in the field have been made almost exclusively on period mortality, as the most recent and innovative techniques are based on modelling and forecasting different functions of period life tables (see, for example, Lee and Carter, 1992; Cairns et al., 2006; Raftery et al., 2013). When considered, cohort effects in mortality modelling and forecasting are typically analysed within an age-period perspective (Renshaw and Haberman, 2006; Cairns et al., 2009; Plat, 2009; Dokumentov et al., 2018).

In this article, we take an alternative perspective and introduce a new methodology to model and forecast mortality from cohort data. An important advantage of cohort forecasts is that they allow one to complete the mortality experience of non-extinct cohorts, thus enabling the derivation of their mortality developments. Our approach focuses on cohort age-at-death distributions: specifically, we propose a warping transformation of the age-axis of a standard distribution to describe and forecast adult mortality developments across cohorts. Since we focus on the cohort perspective, we denote our methodology *Cohort Segmented Transformation Age-at-death Distributions* (C-STAD) model. Warping transformations and skewing procedures have already been fruitfully employed to model distributional changes (see, e.g., Fernández and Steel, 1998; Camarda et al., 2008).

Our methodology is inspired by the Segmented Transformation Age-at-death Distributions (STAD) model recently proposed by Basellini and Camarda (2019b) to forecast adult age-at-death distributions. In addition to shifting the focus from period to cohort mortality, our methodology extends the STAD to a cubic transformation before the modal age at death. The additional parameters c_L and d_L are necessary to adequately describe cohort mortality devel-

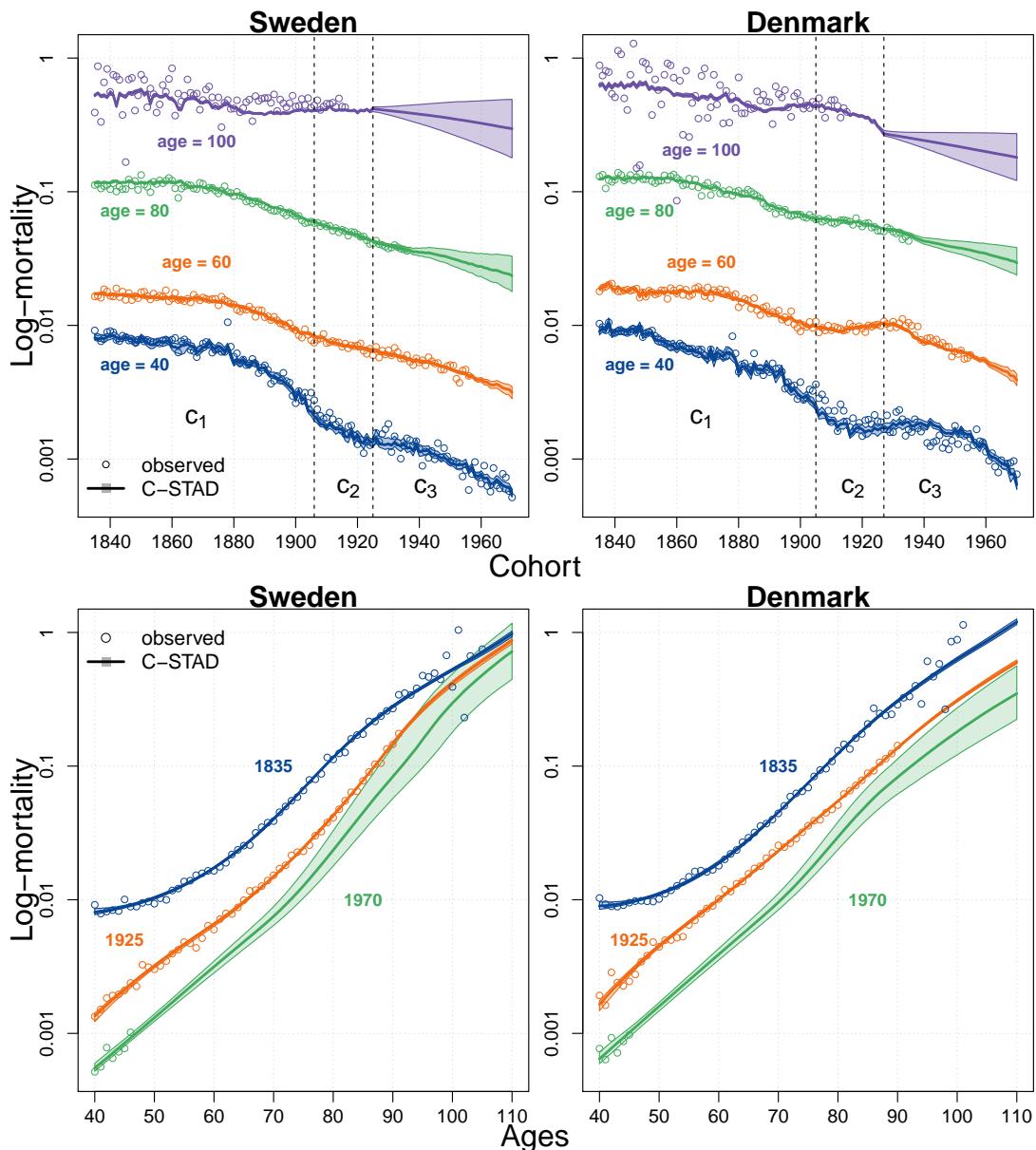


Figure 5.7: Observed, fitted and forecast age-specific mortality rates for selected ages (top panels) and for selected cohorts (bottom panels) with 80% confidence intervals for females in Sweden and Denmark aged 40-110+ for the cohorts 1835-1970.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

opments at young adult ages. A possible explanation for this is related to the significant improvements in mortality, especially at younger adult ages, across the cohorts that we analyse (cf. Fig. 5.7). Non-linear transformation functions above the mode were tested, but they did not provide a better fit compared to a linear transformation function.

Only a handful of models have been proposed to directly forecast cohort mortality so far ([Chiou and Müller, 2009](#); [Zanotto and Mazzucco, 2017](#); [Rizzi et al., 2019](#)). One of the main reasons for the limited efforts in this direction is the heavy data demands that such models require. However, this problem is reduced when only adult mortality is considered ([Booth, 2006](#)). As such, the issue does not affect us to a great extent, as our interest in this article is restricted to adult mortality only.

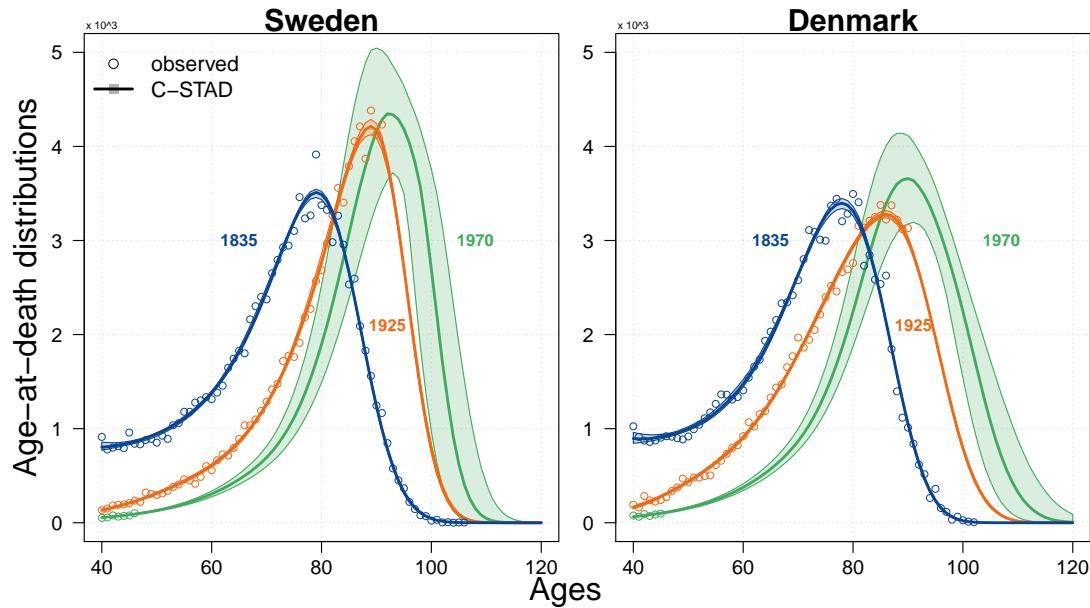


Figure 5.8: Observed, fitted and forecast age-at-death distributions for selected cohorts (bottom panels) with 80% confidence intervals for females in Sweden and Denmark aged 40–110+.

Source: Authors’ elaborations on data from the [Human Mortality Database \(2019\)](#).

We have shown the results of fitting and forecasting cohort mortality with the C-STAD model for Swedish and Danish adult females aged 40–110 for the cohorts 1835–1970. Our methodology is accurate from a point forecast perspective: for each population, we performed six out-of-sample validation exercises of different forecast horizons. The resulting point forecast errors are generally small, even for the longer forecast horizons. Additionally, the C-STAD forecasts are consistently more precise than those of the 2D *P*-spline model of [Currie et al. \(2004\)](#), which has already been used to directly forecast cohort mortality ([CMI Committee, 2007](#)).

Furthermore, both age-cohort approaches (the C-STAD and 2D *P*-spline) to forecast cohort mortality perform significantly better than the standard approach of extracting cohort patterns from the diagonals of a projected [Lee and Carter \(1992\)](#) age-period surface. More generally, age-cohort models are more efficient and parsimonious: looking at Figure 5.3, the objective of a cohort forecast is to project the green area, which is directly obtained by age-cohort approaches. Conversely, age-period models would also need to forecast all data below the green area as a preliminary step, and then extract the cohort mortality of interest.

Our results allow us to derive age-specific and summary measures of mortality, such as remaining life expectancy and the Gini coefficient at age 40 (e_{40} and g_{40}), for all cohorts of the population analysed. Although following similar trends to the 2D *P*-spline model, C-STAD estimates of e_{40} seem to be more coherent when considered together, lacking the rapid increase and cross-over of Danish forecasts displayed by the 2D model. With respect to Danish forecasts, it is interesting to observe a stagnation of e_{40} and an increase of g_{40} for the cohorts 1910–1930. Such results are consistent with other findings in the literature (see, e.g., Fig. 4 in [Jacobsen et al., 2002](#)), which have been attributed to the smoking behaviour of Danish women ([Jacobsen et al., 2006; Lindahl-Jacobsen et al., 2016a](#)).

Since [Thiele \(1871\)](#), actuaries and demographers have decomposed the age-pattern of mortality into three independent components that mainly operate at childhood, middle and old ages,

respectively. Let us denote the three components Childhood, Early-Adulthood and Senescence. Our proposed methodology is specifically designed to model and forecast the Senescent component of mortality. As such, applying the C-STAD from age 40 produces satisfactory results for the female populations analysed in this article because the Childhood and Early-Adulthood components are negligible with respect to the Senescent one in the age range that we study.

Although theoretically and practically possible, we do not recommend applying the model to a (much) wider age range. This would result in a reduction of goodness-of-fit and forecast accuracy, because the C-STAD cannot capture and disentangle the combination of different components at younger ages. A more suitable approach for modelling and forecasting the entire age range would be to decompose and model the mortality pattern by specialized versions of the C-STAD on the component-specific distributions. See [Basellini and Camarda \(2019a\)](#) for an example of this procedure from the age-period perspective.

To conclude, the C-STAD model offers great prospects for mortality forecasting from the cohort perspective. Public and private institutions would benefit from employing our model, as it provides a direct approach to complete the mortality experience of non-extinct cohorts. The R code provided with this article allows a fast and freely available opportunity for this purpose.

5.6 Appendix

5.6.1 Deviance residuals and bootstrap death counts

Model residuals are routinely analysed to explore the goodness-of-fit of a model as well as the adequacy of assumptions about error terms. Within a GLM setting (such as the Poisson considered here), deviance residuals are often used to measure discrepancy between fitted and actual data. For the Poisson distribution they are given by:

$$r_D = \text{sign}(d_{x,c} - \hat{d}_{x,c}) \sqrt{2} \left[d_{x,c} \ln \left(\frac{d_{x,c}}{\hat{d}_{x,c}} \right) - (d_{x,c} - \hat{d}_{x,c}) \right]^{1/2} \quad (5.12)$$

where $d_{x,c}$ and $\hat{d}_{x,c}$ denote the observed and fitted death counts at age x and for cohort c , respectively ([McCullagh and Nelder, 1989](#)).

Deviance residuals can be further employed to take into account the uncertainty related to the estimation of model parameters as suggested by [Koissi et al. \(2006\)](#). Specifically, bootstrap death counts can be computed by resampling deviance residuals with replacement and mapping them to corresponding death counts. We refer the interested reader to [Renshaw and Haberman \(2008\)](#) for details of the inverse formulas, which are based on the seminal work of [Efron and Tibshirani \(1994\)](#).

5.6.2 Section 5.3: additional results

In this appendix, we present additional results related to Section 5.3.

First, Figure 5.9 provides an illustration of the landmark registration procedure that we employ to compute the standard from the aligned distributions. The left panel shows the observed

smooth distributions derived from the 2D P -spline model; in the right panel, the same distributions have been aligned to a common modal age at death, corresponding to the mode of the first cohort (1835). The standard is computed as the mean of the aligned distributions.

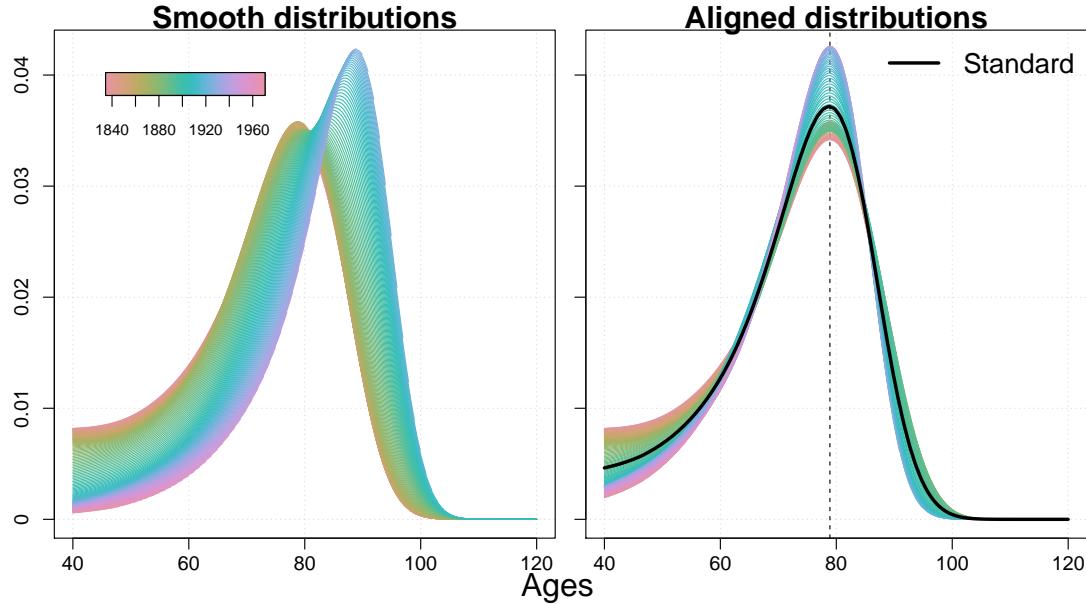


Figure 5.9: Observed smooth (left panel) and aligned distributions for Swedish females aged 40–110+ for cohorts 1835–1970. The standard distribution (black line) is computed as the mean of the aligned distributions.

Source: Authors’ elaborations on data from the [Human Mortality Database \(2019\)](#).

Next, we report the out-of-sample results of Subsection 5.3.1 derived from employing two different prediction accuracy measures. In addition to the root mean square error, we computed the mean absolute error (MAE) and the mean absolute percentage error (MAPE):

$$\text{MAE} = \frac{1}{h} \sum_{c=1}^h |\hat{y}_c - y_c| ,$$

$$\text{MAPE} = \frac{100}{h} \sum_{c=1}^h \left| \frac{\hat{y}_c - y_c}{y_c} \right| ,$$

where h is the forecasting horizon, and \hat{y}_c and y_c are the forecast and observed out-of-sample values of either e_{40} or g_{40} .

Tables 5.2 and 5.3 show the out-of-sample results obtained using the MAE and the MAPE, respectively. The results are very similar to those obtained with the RMSE shown in Table 5.1: the C-STAD forecasts are accurate in completing the mortality experience of partially observed cohorts, with forecasts errors generally low and smaller than the 2D P -spline model. Moreover, both age-cohort models significantly outperform the standard LC age-period based approach.

Finally, we provide additional results with regard to Subsection 5.3.2. Figure 5.10 shows the fitted and forecast C-STAD parameters with 80% confidence intervals for Swedish and Danish adult females for cohorts 1835–1970.

Fitting cohorts	Forecast cohorts	Horizon	Measure	Sweden			Denmark		
				C-STAD	2D P-spline	LC (period)	C-STAD	2D P-spline	LC (period)
1835–1895	1896–1905	10y	e_{40} G_{40}	0.08 0.07	0.06 0.08	0.20 0.28	0.07 0.06	0.07 0.07	0.37 0.34
1835–1890	1891–1905	15y	e_{40} G_{40}	0.07 0.07	0.07 0.08	0.22 0.35	0.06 0.06	0.07 0.09	0.35 0.38
1835–1885	1886–1905	20y	e_{40} G_{40}	0.05 0.06	0.07 0.09	0.29 0.45	0.06 0.05	0.07 0.09	0.38 0.39
1835–1880	1881–1905	25y	e_{40} G_{40}	0.03 0.07	0.06 0.08	0.37 0.52	0.03 0.08	0.07 0.10	0.42 0.39
1835–1875	1876–1905	30y	e_{40} G_{40}	0.04 0.08	0.07 0.12	0.41 0.55	0.02 0.07	0.08 0.13	0.46 0.44
1835–1870	1871–1905	35y	e_{40} G_{40}	0.09 0.03	0.07 0.10	0.48 0.52	0.04 0.05	0.10 0.17	0.49 0.43

Table 5.2: Mean absolute error (MAE) of the C-STAD, 2D P-spline and LC (period) forecasts of e_{40} and G_{40} for adult females in Sweden and Denmark in six out-of-sample validation exercises: forecast horizon of 10, 15, 20, 25, 30 and 35 years. Lower values of the MAE (in bold, assessed using all available decimals) correspond to greater forecast accuracy.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

Fitting cohorts	Forecast cohorts	Horizon	Measure	Sweden			Denmark		
				C-STAD	2D P-spline	LC (period)	C-STAD	2D P-spline	LC (period)
1835–1895	1896–1905	10y	e_{40} G_{40}	0.20% 0.42%	0.16% 0.45%	0.51% 1.57%	0.19% 0.35%	0.19% 0.36%	0.96% 1.83%
1835–1890	1891–1905	15y	e_{40} G_{40}	0.17% 0.36%	0.19% 0.45%	0.58% 1.93%	0.16% 0.30%	0.19% 0.47%	0.93% 2.04%
1835–1885	1886–1905	20y	e_{40} G_{40}	0.13% 0.34%	0.19% 0.46%	0.77% 2.40%	0.15% 0.27%	0.18% 0.44%	1.02% 2.02%
1835–1880	1881–1905	25y	e_{40} G_{40}	0.08% 0.40%	0.17% 0.44%	0.98% 2.71%	0.08% 0.42%	0.18% 0.51%	1.14% 2.05%
1835–1875	1876–1905	30y	e_{40} G_{40}	0.10% 0.42%	0.20% 0.59%	1.12% 2.84%	0.07% 0.39%	0.23% 0.63%	1.28% 2.21%
1835–1870	1871–1905	35y	e_{40} G_{40}	0.22% 0.14%	0.19% 0.51%	1.33% 2.62%	0.12% 0.23%	0.29% 0.81%	1.36% 2.09%

Table 5.3: Mean absolute percentage error (MAPE) of the C-STAD, 2D P-spline and LC (period) forecasts of e_{40} and G_{40} for adult females in Sweden and Denmark in six out-of-sample validation exercises: forecast horizon of 10, 15, 20, 25, 30 and 35 years. Lower values of the MAPE (in bold, assessed using all available decimals) correspond to greater forecast accuracy.

Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

We performed diagnostic checks on the fitted C-STAD model for the two populations analysed in this paper by using Eq. (5.12). Poisson deviance residuals for the two populations are shown in Figure 5.11. No clear patterns emerge from this graphical analysis, with the exception of the years corresponding to the Spanish' flu and World War II.

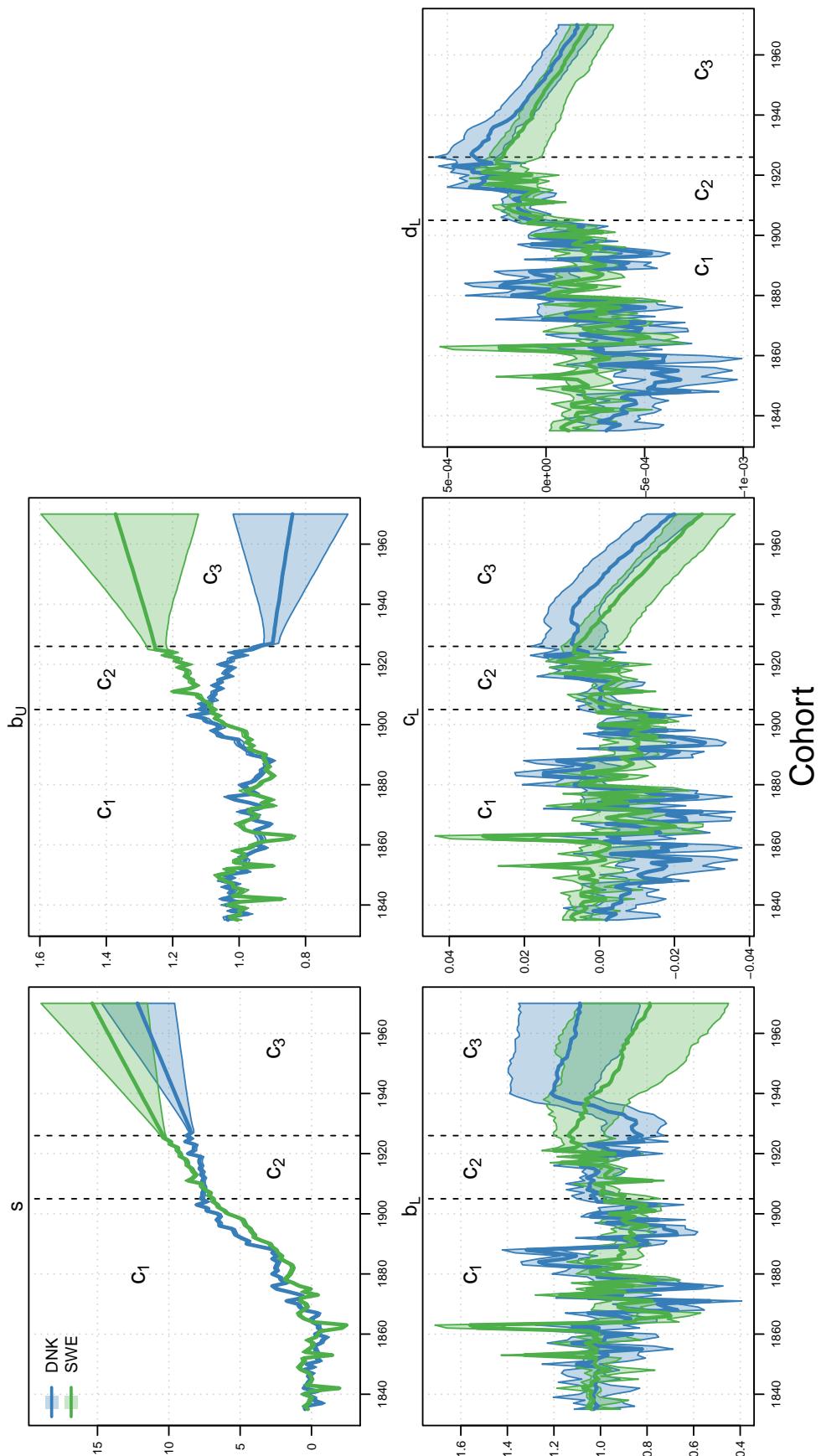


Figure 5.10: Estimated and forecast C-STAD parameters for adult females in Sweden and Denmark for the cohorts 1835–1970.
Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

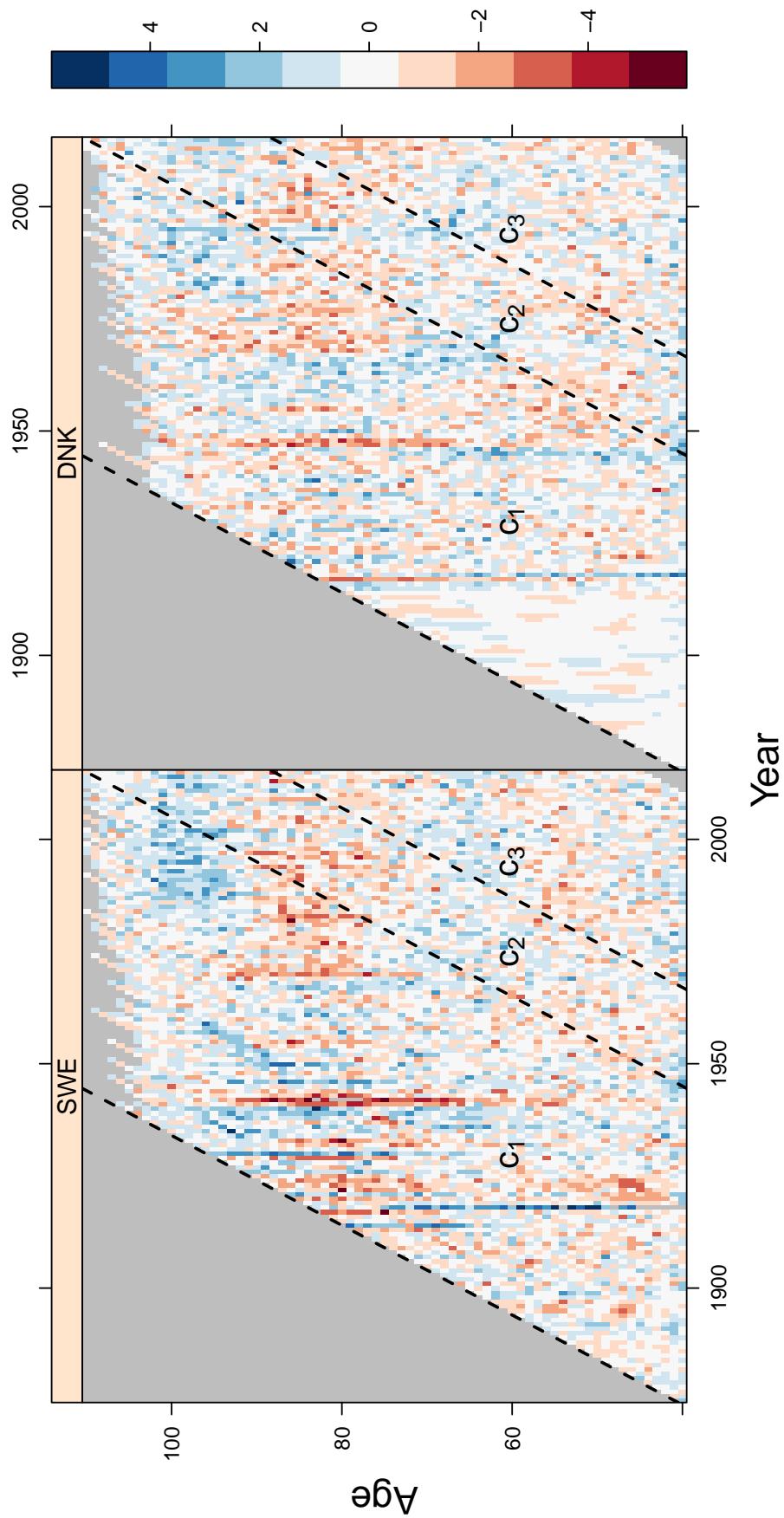


Figure 5.11: Poisson deviance residuals of the C-STAD model for adult females in Sweden and Denmark for the cohorts 1835–1970.
Source: Authors' elaborations on data from the [Human Mortality Database \(2019\)](#).

Chapter 6

Smoothing, decomposing and forecasting mortality rates

Carlo Giovanni Camarda

Ugo Filippo Basellini

Manuscript under review.

Smoothing, decomposing and forecasting mortality rates

Carlo Giovanni Camarda^{1,†} and Ugofilippo Basellini^{1,2}

¹*Institut national d'études démographiques (INED), Paris, France*

²*Interdisciplinary Centre on Population Dynamics (CPop) and Department of Public Health,
University of Southern Denmark, Odense, Denmark*

[†]*Corresponding author: carlo-giovanni.camarda@ined.fr*

Abstract

The Lee-Carter (LC) model represents a landmark paper in mortality forecasting. Whilst having been widely accepted and adopted, the model has some limitations that hinder its performance. Some variants of the model have been proposed to deal with these drawbacks individually, none coped with them all at the same time. In this paper, we propose a Three-Component smooth Lee-Carter (3C-sLC) model which overcomes all the issues simultaneously. It decomposes mortality development into childhood, early-adult and senescent mortality, which are described, individually, by a smooth variant of the LC model. Smoothness is enforced to avoid irregular patterns in projected life tables, and complexity in the forecasting methodology is unaltered with respect to the original LC model. Component-specific schedules are considered in projections, providing additional insights into mortality forecasts. We illustrate the proposed approach to mortality data for the USA and Switzerland. The 3C-sLC captures mortality developments better than a smooth improved version of the LC model, and it displays wider prediction intervals. The proposed approach provides demographers, actuaries, epidemiologists and social scientists in general with a unique and valuable tool to simultaneously smooth, decompose and forecast mortality.

Keywords: mortality decomposition · mortality forecasting · smoothness · rate of mortality improvement · Lee-Carter model

6.1 Background

An accurate knowledge of future mortality patterns and their uncertainty is crucial in almost all areas of a society, including the health-care system, health and life insurance, as well as pension schemes. Since [Bonynge \(1852\)](#), demographers, actuaries and statisticians have developed several methodologies and techniques to predict mortality. However, up until the 1980s, mortality forecasts were based on deterministic scenarios and, generally, some variants would be assumed at the life expectancy level as input for producing population projections. This approach was used almost without exception by international bodies, such as the United Nations and World Bank, and national statistical offices ([Booth, 2006](#)).

One of the main limitations of this scenario-based approach is that the probability of different scenarios cannot be quantified nor specified *a priori*, and the structure of mortality developments over age and time is only partially considered. For this reason, the last thirty years witnessed an increasing use of more sophisticated statistical methodologies for projections, which enable the construction of forecast probability distributions and probabilistic prediction intervals, mainly using time changes in age-specific death rates ([Booth and Tickle, 2008](#)).

The use of stochastic models has flourished in recent decades (for comprehensive reviews, see [Booth, 2006](#); [Booth and Tickle, 2008](#); [Stoeldraijer et al., 2013](#)) and in [1992](#), a breakthrough was achieved by [Lee and Carter](#), who proposed an elegant and powerful methodology to forecast mortality that later became widely accepted and employed. The Lee-Carter (LC) model is based on linear extrapolations of the logarithms of age-specific death rates, using principal component techniques. In particular, a matrix of logged age-specific mortality rates over time is summarized by two vectors of age-specific parameters α_x and β_x , which describe the general shape of mortality and the fixed rate of mortality improvement at age x , respectively, and by a time-varying index κ_t , which captures the general level of mortality. Mortality forecasts are then derived from the projection of κ_t using standard time series methods.

While the LC model works reasonably well, several limitations have been revealed in the succeeding literature. A central assumption of the LC model is the fixed rate of age-specific mortality improvement over time ([Lee and Miller, 2001](#)). Already recognized by [Alho \(1992\)](#), this assumption has been violated in several low-mortality countries in recent decades, as rates of mortality improvements have tended to decline over time at younger ages, and they have risen at older ages ([Kannisto et al., 1994](#); [Vaupel et al., 1998](#); [Wilmoth and Horiuchi, 1999](#)). As a result, projections that ignore this “rotation” of mortality declines will lead to errors, particularly in the projected age patterns of future death rates. To overcome this limitation, [Li et al. \(2013\)](#) proposed a rotation of the β_x schedule for long-term projections. This allows one to capture the observed trend of deceleration of infant and child mortality decline and the acceleration of old-age mortality reduction in low-mortality countries during the most recent decades. A similar approach has been used in other projection models ([Ševčíková et al., 2016](#)). Nevertheless, the approach proposed by [Li et al. \(2013\)](#) relies strongly on a subjective choice of the model parameters, and it does not account for the stochastic process behind mortality.

The original LC model is estimated by minimizing the residual sum of squares using singular value decomposition (SVD), which is as simple as inappropriate when dealing with mortality data ([Wilmoth, 1993](#); [de Jong and Tickle, 2006](#); [Koissi and Shapiro, 2006](#)). [Brouhns et al. \(2002\)](#)

embedded the LC model within a Poisson setting using Maximum Likelihood Estimation and a Newton-Raphson algorithm to estimate model parameters. Other papers used this Poisson assumption for estimating the LC model (Renshaw and Haberman, 2003b; Brouhns et al., 2005; Czado et al., 2005; Haberman and Renshaw, 2008; Li and Li, 2017). Despite the appropriate estimation procedure, the LC model leads to outcomes which become less smooth and more unrealistic when projected, and age-patterns tend to deviate from any given baseline described by α_x (Grosi and King, 2007, 2008). To prevent irregular projected life tables, smoothing techniques have been implemented within an LC framework. Whereas Renshaw and Haberman (2003c) suggested smoothing estimated series of β_x , Hyndman and Ullah (2007) employed a functional data approach and smoothed the mortality data as a preliminary step. Alternatively, a penalized likelihood approach has been proposed by Delwarde et al. (2007) and Currie (2013). Their approach allows them to frame the LC model within a Poisson setting and simultaneously obtain smooth β_x and, consequently, regular forecast age-patterns.

None of the variants and generalizations of the LC model attempted to account for the complex nature of the human mortality age-pattern. Specifically, we refer to the hypothesis that human mortality can be decomposed into three different groups that operate principally upon childhood, middle and old ages, respectively. This idea goes back at least one and a half centuries (Thiele, 1871) and it leads to a so-called “bath-tub shaped” mortality age-pattern. In most developed populations, during the first years of life, mortality decreases steeply from a relatively high value at age 0 (Levitis, 2011). A minimum is commonly reached at ages around 10-15 (Ebeling, 2018). Afterwards, especially for men, hazard rates show a hump at young-adult ages (Goldstein, 2011; Remund, 2015). Starting from about age 30, mortality rises exponentially and eventually it tends to level off at ages above 80 (Preston, 1976; Vaupel, 1997; Thatcher et al., 1998). The idea of mortality constructed from three competing components has been extensively used for modeling purposes (Siler, 1979, 1983; Heligman and Pollard, 1980; Kostaki, 1992; Rogers and Little, 1994; Dellaportas et al., 2001; de Beer and Janssen, 2016; Camarda et al., 2016; Mazzuco et al., 2018; Remund et al., 2018). However, very few attempts have been made to forecast the three components. Only Forfar and Smith (1987) employed the Heligman and Pollard model for forecasting, as the large number of parameters limits the model’s potential for projections (McNow and Rogers, 1992). In addition, the model’s parameters are typically highly correlated (Hartmann, 1987), and these high correlations further compromise their interpretability (Booth and Tickle, 2008).

Here, we propose a novel extension of the LC model that (i) is framed within a Poisson setting; (ii) enforces smoothness in the outcomes and (iii) addresses the drawback of the fixed rates of mortality improvement from a different perspective, i.e. by separately modeling childhood, early-adult and high-ages mortality. In particular, we simultaneously estimate a smooth variant of the LC model for all three components of human mortality. We therefore call our approach the “Three-Component smooth Lee-Carter” (3C-sLC) model. In other words, given a set of observed death counts and exposure-to-risk data, we derive the standard LC parameters α_x , β_x and κ_t for each component, with both α_x and β_x smoothly varying over x . Forecasting is achieved by extrapolating component-specific time-indexes via standard Box-Jenkins methods, maintaining the simplicity of the original LC model in projecting future trends. The main advantage of our

model is the greater flexibility of mortality patterns, which do not depend on a single fixed rate of age-specific mortality improvement, but rather on the combination of different component-specific schedules. The methodology that we propose also allows us to forecast overall mortality as a sum of future component-specific patterns. Consequently, future trends in childhood, early-adult and old-age mortality can be analyzed independently, adding great explanatory value to the proposed approach.

This paper is organized as follows. In Section 6.2, we start by introducing the data that we use in the paper, which are derived from the [Human Mortality Database \(2019\)](#). We present the original LC model and its smooth version in Section 6.2.1; then, Section 6.2.2 introduces the 3C-sLC model, as well as the procedures for estimating its parameters and obtain mortality forecasts. Throughout Section 6.2, we present outcomes of the smooth LC and 3C-sLC models on a specific population for illustrative purposes. In Section 6.3, we show and compare the results of fitting and forecasting mortality in four populations using the two models. Section 6.4 provides a discussion of our methodology and its related results, and Section 6.5 concludes the paper.

6.2 Data and methods

We assume that we have data on deaths and exposures to the risk of death, arranged in two matrices, $\mathbf{Y} = (y_{x,t})$ and $\mathbf{E} = (e_{x,t})$, each $m \times n$ where rows and columns are classified by single age at death, \mathbf{x} , $m \times 1$, and single year of death, \mathbf{t} , $n \times 1$, respectively. We aim to forecast mortality for future years \mathbf{t}_F , $n_F \times 1$. We assume that the number of deaths $y_{x,t}$ are realizations of the random variable $Y_{x,t}$, which follows a Poisson distribution:

$$Y_{x,t} \sim \mathcal{P}(e_{x,t} \mu_{x,t}), \quad (6.1)$$

i.e. the expected value is the product of exposures and the force of mortality ([Brillinger, 1986](#)). Description of the mortality development over age and time is commonly provided by portraying the so-called death rates: $m_{x,t} = y_{x,t}/e_{x,t}$.

In the following, we will present outcomes from the proposed model for four different populations. Specifically, we analyze both female and male populations of the USA and Switzerland (CHE). Data are taken from the [Human Mortality Database \(2019\)](#). We model mortality from ages 0 to 100 over the period 1960-2016, forecasting up to 2050.

In this section and only for illustrative purposes, methodology is presented on Swiss males data. Figure 6.1 shows mortality rates on a log scale over ages for all years as well as time trends for selected ages over time. The typical mortality age-pattern is portrayed in the left panel: a rapid declining mortality from birth, a roughly log-linear trajectory starting from ages 30-40, and an intermediate phase in early-adult ages often described as “accident hump”.

Time trends are also identifiable in the right panel of Figure 6.1. A general mortality improvement is clear with decreasing rates over time. However, differences in the speed of mortality reductions are evident for different ages, and some ages have also experienced stagnation and increasing mortality, e.g. age 30 from 1960 to early 1990s. Whereas looking at death rates helps to have a good image of mortality developments, it is essential to develop statistical models to

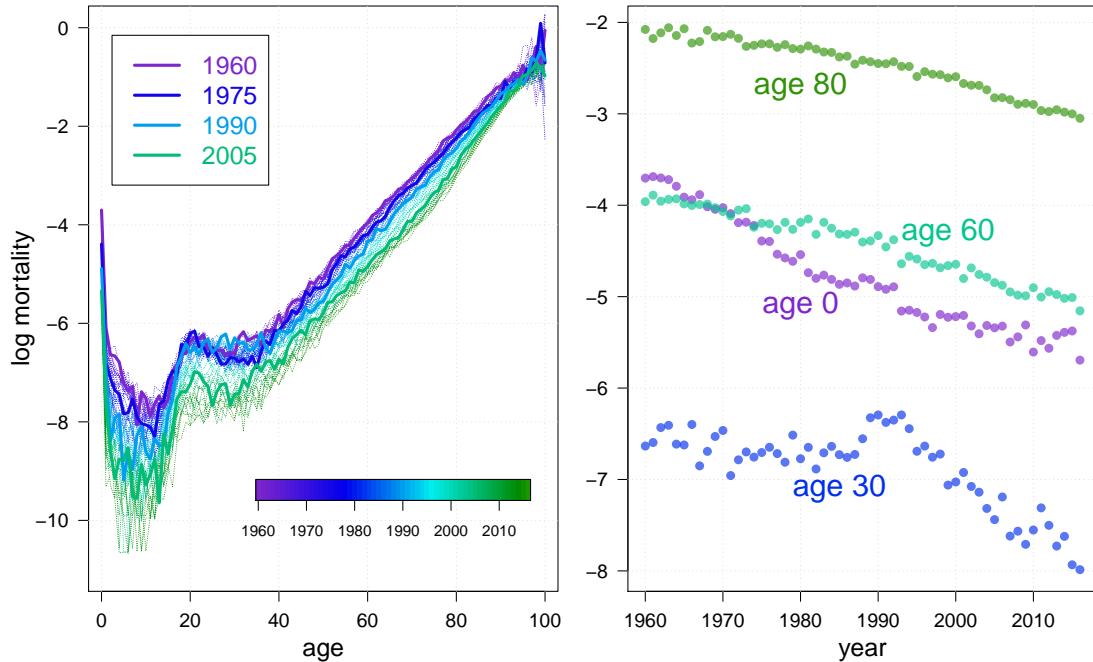


Figure 6.1: Actual death rates over age and time on a log scale for Swiss males. Ages 0-100, years 1960-2016. Left panel: rates over age for all years. Selected years are depicted by thicker lines. Right panel: selected ages over years.

obtain an accurate and parsimonious description as well as a reliable prediction of the observed trends. The following pages will be devoted to this purpose.

6.2.1 The Lee-Carter model

Lee and Carter (1992) pioneered an elegant and powerful methodology to model and forecast mortality based on a simple formula for the logged death rates:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \quad (6.2)$$

where α_x describes the average shape of the age profile, β_x the rate of mortality improvement at age x , and κ_t the general level of mortality at time t . The $\varepsilon_{x,t}$ are error terms that reflect residual age-specific historical influences not captured by the model. To ensure model identification, the parameters are commonly subjected to the following constraints:

$$\sum_x \beta_x = 1 \quad \text{and} \quad \sum_t \kappa_t = 0. \quad (6.3)$$

The original Lee-Carter (LC) model is estimated by minimizing the residual sum of squares:

$$\sum_{x,t} (\ln(m_{x,t}) - \alpha_x - \beta_x \kappa_t)^2. \quad (6.4)$$

In particular, the authors employ a singular value decomposition (SVD) method to find the least squares solution of (6.4); as such, the estimated $\hat{\alpha}_x$ is the average of the observed $\ln(m_{x,t})$, while $\hat{\beta}_x$ and $\hat{\kappa}_t$ are the first left- and right-singular vectors of the SVD of the matrix $\ln(m_{x,t}) - \hat{\alpha}_x$.

Furthermore, the parameter $\hat{\kappa}_t$ is adjusted in a second-step estimation so that the resulting fitted deaths match the total number of deaths observed in the data at each year t .

An advantage of the LC model is that its time-index κ_t can be easily forecast since it is able to capture the linear decline of the observed mortality development: see the grey dashed line in the right panel of Figure 6.3 (note that the figure also plots parameters of our proposed model, which should be ignored for now). In particular, forecasting in the LC model is performed by modelling and extrapolating fitted values of κ_t by an autoregressive integrated moving average (ARIMA) process. A random walk with drift (RWD) is used almost exclusively in applications with low-mortality populations and in the following we will follow this approach. Finally the extrapolated κ_t and its uncertainty are combined with the previous estimations to forecast the future death rates and associated prediction intervals.

With respect to the estimation procedure, the main drawback of the ordinary least-squares estimation via SVD is that the errors are assumed to be homoskedastic and normally distributed, which is a fairly unrealistic assumption for human mortality (Alho, 2000). Indeed, the logarithm of the observed force of mortality is much more variable at older than at younger ages because of the much smaller number of deaths (Brouhns et al., 2002).

To overcome this limitation, Brouhns et al. (2002) embed the LC model in a Poisson setting. The logarithm of the force of mortality in (6.1) can then be written as follows:

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \quad (6.5)$$

or in matrix notation:

$$\ln \mathbf{M} = \boldsymbol{\alpha} \mathbf{1}'_n + \boldsymbol{\beta} \boldsymbol{\kappa}' \quad (6.6)$$

where $\mathbf{M} = (\mu_{ij})$ is the matrix of the underlying force of mortality, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ denote the LC parameter-vectors and $\mathbf{1}_n$ is a $n \times 1$ matrix of 1s. Estimates of the parameters can be achieved by maximizing the associated log-likelihood function:

$$\ln \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa} | \mathbf{Y}, \mathbf{E}) \propto \sum_{x,t} [y_{x,t} \ln(\mu_{x,t}) - e_{x,t} \mu_{x,t}] . \quad (6.7)$$

The model parameters are subject to the same constraints as in (6.3) and the log-likelihood in (6.7) is maximized by a uni-dimensional iterative Newton-Raphson method. The advantage of this assumption is that the error terms in the model now follow a non-additive heteroscedastic error structure, which is more appropriate for modeling human mortality. Moreover maximum-likelihood estimation does not require any second-step adjustment of $\boldsymbol{\kappa}$, as the fitted number of deaths already matches the observed deaths at each year t .

Regardless of the estimation procedure, irregular age-patterns in future mortality have been observed when the LC model has been used, and increasing loss of smoothness over ages is a well known feature of this model (Girosi and King, 2007, 2008). This is mainly due to the $\boldsymbol{\beta}$, which often exhibits a fluctuating pattern that results in irregular projected life tables. Instead of smoothing the estimated $\boldsymbol{\beta}$ without modifying $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$, Delwarde et al. (2007) proposed a penalized likelihood approach to obtain smooth $\boldsymbol{\beta}$ within the estimation procedure. In addition, we propose smoothing the parameter-vector $\boldsymbol{\alpha}$ since, especially in small populations, the simple

average mortality age-pattern might present variable behaviour that can contribute to produce unsmooth future mortality patterns. In contrast, the time-index κ is free to vary over the years to better capture mortality fluctuations and provide an appropriate measurement of the uncertainty when forecasting is involved.

Smooth LC estimates could be achieved by adding a penalty term in the log-likelihood (6.7) which penalizes differences between adjacent α and β :

$$\ln \mathcal{L}^P(\alpha, \beta, \kappa | \mathbf{Y}, \mathbf{E}) = \ln \mathcal{L}(\alpha, \beta, \kappa | \mathbf{Y}, \mathbf{E}) - \frac{1}{2} \lambda_\alpha \alpha' \mathbf{D}' \mathbf{D} \alpha - \frac{1}{2} \lambda_\beta \beta' \mathbf{D}' \mathbf{D} \beta \quad (6.8)$$

where the smoothing parameters λ_α and λ_β control the amount of smoothness in the vector α and β and \mathbf{D} is the second order difference $m \times (m - 2)$ matrix:

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots \\ 0 & 1 & -2 & 1 & 0 & \dots \\ 0 & 0 & 1 & -2 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (6.9)$$

The order of difference in \mathbf{D} does not play a relevant role in this context. Unless specified otherwise, in the following we use second order differences. Moreover we can use the same \mathbf{D} matrix for both parameter-vectors since they both act on the age dimension.

In other words, the larger the smoothing parameters λ_α and λ_β , the smoother the estimates of α and β , respectively. Consequently fitted rates will show smoother patterns. Alternatively small λ s lead to estimates close to the original LC model. Here we optimized both smoothing parameters by minimizing the Bayesian Information Criterion (BIC, Schwarz, 1978).

Gray dashed lines in Figure 6.3 present estimated parameters from this smooth variant of the Lee-Carter model for Swiss males. The vector α shows the average mortality age-pattern, and the accident-hump is evident for early adult ages. This average trajectory is modulated over age and time by the parameters β and κ , respectively. A general linear time-trend is visible in the right panel and the average mortality improvement described by β_x presents large changes in younger ages and about age 70.

The time index is then extrapolated using a RWD and associated prediction intervals can be computed (grey areas in the right panel of Figure 6.3). From estimated and future values of κ_t as well as estimated α and β , the whole mortality development over both age and time is computed.

Figure 6.2 shows estimated and forecast death rates on a log scale for Swiss males. Whereas the smooth Lee-Carter is able to capture the general mortality improvement, future age-patterns still show fluctuating behaviour. Moreover, the fixed age-specific average rates of mortality improvement, described by the vector β , lead to unreasonable time-trends, i.e. all age groups show a constant decline in mortality that has not been observed in the past decades. These features are ultimately translated into future age-patterns that produce a poor description of past trends; see especially the fitted mortality for ages 0 and 30.

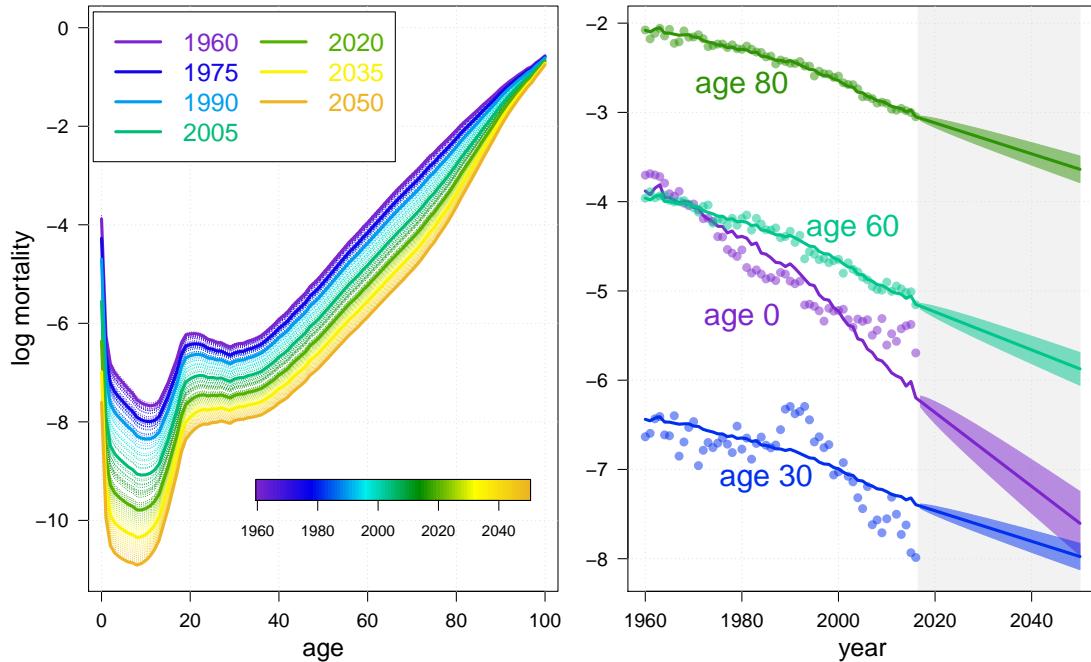


Figure 6.2: Actual, estimated and forecast death rates by a smooth variant of the Lee-Carter over age and time on a log scale for Swiss males. Ages 0-100, fitted years 1960-2016, forecast years 2017-2050. Left panel: rates over age for all years. Selected years are depicted by thicker lines. Right panel: selected ages over years.

6.2.2 The Three-Component smooth Lee-Carter model

Outcomes from the classic LC model have shown that the assumption of a single shape of mortality improvement over age described by β can often be wrong for mortality data. In other words, within a plain Lee-Carter framework, we assume that age-specific improvements are fixed throughout the observed period as well as in the forecast horizon. Available approaches attempted to modify the shape in β accounting for the projected life expectancy at birth and assuming relationships between estimated values at different ages (Li et al., 2013; Ševčíková et al., 2016). Here we opt for a different approach which does not require prior knowledge of future life expectancy and is not based on a subjective choice of the model parameters. Moreover, it allows us to simultaneously decompose and forecast the whole of childhood, early-adulthood and senescent mortality.

Since Thiele (1871), demographers have recognized that mortality trajectories over age can be conceived as the sum of three main components operating principally upon childhood, middle and old ages, respectively. We thus decide to incorporate this demographic knowledge within a Lee-Carter framework and decompose mortality into three components. Each component will be described by a distinct Lee-Carter model estimated in a smooth setting. Hence we call this a Three-Component smooth Lee-Carter (3C-sLC) model.

In formulas, let γ_C , γ_A and γ_S denote the vectors representing Childhood, early-Adulthood and Senescent mortality, respectively. The length of these vectors is equal to the product of the number of ages and years in the original data, $m n$. In a 3C-sLC model, each component is

defined as a Lee-Carter model:

$$\ln \gamma_k = \text{vec}(\alpha_k \mathbf{1}'_n + \beta_k \kappa'_k) \quad \text{for } k = C, A, S. \quad (6.10)$$

Here, the vec operator stacks the columns of a matrix in column order on top of each other. We note that with this definition the age suffix varies faster than the year suffix in all γ_k .

With our approach, three sets of α , β , and κ will be estimated. The advantages of this are twofold: on the one hand, we will have component-specific mortality improvement patterns which will mitigate the LC model drawback of a fix β and, on the other, we will be able to extract forecast mortality trajectories independently for childhood, middle-age and senescent components. These important enhancements come at a price: the 3C-sLC model is extremely flexible and highly non-linear. Consequently, prior knowledge of population-specific mortality development would be required to define the necessary amount of flexibility.

As in the classic LC model, future mortality can be obtained by extrapolating estimated κ_k by an ARIMA process and computing component-specific death rates by (6.10) using associated α_k and β_k . The sum of the three components provides the overall mortality. In the following, we opt to keep the original LC approach and apply a RWD to each time-index. Alternative procedures such as the selection of κ_k -specific ARIMA models would simply add an additional layer of complexity without improving the final outcomes, and lead to a disproportionate increase of prediction intervals.

Instead of decomposing the whole mortality data and successively estimating independent LC models, we opt for a simultaneous estimation of the three components. For the purpose of regression, we arrange the matrix of deaths by column order into a vector, that is, $\mathbf{y} = \text{vec}(\mathbf{Y})$. Likewise we arrange the matrix of exposures $\mathbf{e} = \text{vec}(\mathbf{E})$. We then express the expected values of the Poisson distribution in (6.1) as a sum of the three components:

$$\mathbf{y} \sim \mathcal{P}(\mathbf{e} * \boldsymbol{\mu} = \mathbf{C} \boldsymbol{\gamma}), \quad (6.11)$$

where $\boldsymbol{\gamma}' = (\gamma_C, \gamma_A, \gamma_S)'$ and $*$ denotes the element-wise product. The matrix \mathbf{C} additively combines the γ_k and also incorporates the exposures:

$$\mathbf{C} = \mathbf{1}_{1,3} \otimes \text{diag}(\mathbf{e}). \quad (6.12)$$

where $\mathbf{1}_{1,3}$ is a 1×3 matrix of ones, $\text{diag}(\mathbf{e})$ is the diagonal matrix of the exposure population and \otimes denotes the Kronecker product (Harville, 1997, p. 333).

Within this system, the force of mortality in each component is simultaneously multiplied by the associated exposures and summed up to obtain the final expected value. We thus hold the Poisson assumption, each component γ_k incorporates the associated force of mortality as described in (6.10) and the sum of the three components is done automatically by matrix multiplication.

Estimation by a Composite Link Model approach

The 3C-sLC model in (6.11) can be viewed as a Composite Link Model (CLM). Introduced by [Thompson and Baker \(1981\)](#) as an extension of the generalized linear model (GLM; [McCullagh and Nelder, 1989](#)), CLM is an elegant framework for modeling expected values which are described as a sum of components. A demographic application of this class of models is given in [Remund et al. \(2018\)](#). In their paper, they employed the Sum of Smooth Exponentials model (SSE, [Camarda et al., 2016](#)) for decomposing young adult excess mortality by cause-of-death. Specifically, the SSE model can be seen as a generalization of the 3C-sLC where decomposition is achieved in a completely non-parametric setting. However, none of these approaches deal with mortality forecasting which could be obtained by describing each component with an LC structure as in (6.10).

Unlike the SSE, each component of the 3C-sLC model is not linear with respect to all unknown parameters. Consequently, estimation cannot be carried out directly by a penalized Iterative re-Weighted Least Squares (IWLS) algorithm as proposed by [Eilers \(2007\)](#). Nevertheless, we linearize the system of equations with respect to each series of parameters and iteratively solve the associated IWLS. Similarly, [Currie \(2013\)](#) proposed estimating a classic LC model as a sequence of three constrained GLMs.

In general, we need to solve the following system of equations:

$$(\check{\mathbf{X}}' \tilde{\mathbf{W}} \check{\mathbf{X}} + \mathbf{P}) \boldsymbol{\theta} = \check{\mathbf{X}}' (\mathbf{y} - \boldsymbol{\mu}) + \check{\mathbf{X}}' \tilde{\mathbf{W}} \check{\mathbf{X}} \tilde{\boldsymbol{\theta}}, \quad (6.13)$$

where $\boldsymbol{\theta}$ denotes the combined vector with one of the triplets of the LC parameters: $\boldsymbol{\theta}' = (\boldsymbol{\alpha}_C, \boldsymbol{\alpha}_A, \boldsymbol{\alpha}_S)', \boldsymbol{\theta}' = (\boldsymbol{\beta}_C, \boldsymbol{\beta}_A, \boldsymbol{\beta}_S)' \text{ and } \boldsymbol{\theta}' = (\boldsymbol{\kappa}_C, \boldsymbol{\kappa}_A, \boldsymbol{\kappa}_S)'$. Moreover we have that:

$$\check{\mathbf{X}} = \mathbf{W}^{-1} \mathbf{C}_{\boldsymbol{\theta}} \boldsymbol{\Gamma} \mathbf{X}_{\boldsymbol{\theta}}; \quad \mathbf{W} = \text{diag}(e * \boldsymbol{\mu}); \quad \boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\gamma}). \quad (6.14)$$

A tilde, as in $\tilde{\boldsymbol{\theta}}$, indicates the current approximation to the solution. Note that only composite and design matrices ($\mathbf{C}_{\boldsymbol{\theta}}$ and $\mathbf{X}_{\boldsymbol{\theta}}$) change with respect to the series of parameters we currently estimate.

We start by fixing all triplets of parameters $\boldsymbol{\kappa}_k$ and $\boldsymbol{\beta}_k$ and estimate $\boldsymbol{\theta}' = (\boldsymbol{\alpha}_C, \boldsymbol{\alpha}_A, \boldsymbol{\alpha}_S)'$. Composite and design matrices for solving (6.13) are thus given by:

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\alpha}} &= [\mathbf{u}_C(\mathbf{1}_n \otimes \mathbf{I}_m) : \mathbf{u}_A(\mathbf{1}_n \otimes \mathbf{I}_m) : \mathbf{u}_S(\mathbf{1}_n \otimes \mathbf{I}_m)] \\ \mathbf{X}_{\boldsymbol{\alpha}} &= \mathbf{I}_{3m} \end{aligned} \quad (6.15)$$

where $\mathbf{u}_k = \text{diag}[e \text{vec}(\exp(\boldsymbol{\beta}_k \boldsymbol{\kappa}'_k))]$ for $k = C, A, S$ and \mathbf{I}_m is the identity matrix with m rows and columns.

Analogously, we then fix the other triplets of parameters and we solve (6.13) by changing the composite and design matrices:

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\beta}} &= \mathbf{C}_{\boldsymbol{\kappa}} = [\mathbf{u}_C : \mathbf{u}_A : \mathbf{u}_S] \\ \mathbf{X}_{\boldsymbol{\beta}} &= \text{diag}[\boldsymbol{\kappa}_C \otimes \mathbf{I}_m, \boldsymbol{\kappa}_A \otimes \mathbf{I}_m, \boldsymbol{\kappa}_S \otimes \mathbf{I}_m] \\ \mathbf{X}_{\boldsymbol{\kappa}} &= \text{diag}[\mathbf{I}_n \otimes \boldsymbol{\beta}_C, \mathbf{I}_n \otimes \boldsymbol{\beta}_A, \mathbf{I}_n \otimes \boldsymbol{\beta}_S] \end{aligned} \quad (6.16)$$

where $\mathbf{u}_k = \text{diag}[\mathbf{e} \text{vec}(\exp(\boldsymbol{\alpha}_k \mathbf{1}'_n))]$ for $k = C, A, S$.

Smoothness of the parameters is achieved via the penalty matrix \mathbf{P} in (6.13). As in the smooth LC variant in (6.8), smoothness is enforced only for triplets $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$. Time-indexes $\boldsymbol{\kappa}_k$ are free to change over time. In general the penalty term takes a block diagonal structure as follows:

$$\mathbf{P} = \text{diag}(\mathbf{P}_{\boldsymbol{\theta}_C}, \mathbf{P}_{\boldsymbol{\theta}_A}, \mathbf{P}_{\boldsymbol{\theta}_S}) \quad (6.17)$$

where $\mathbf{P}_{\boldsymbol{\theta}_k} = \lambda_{\boldsymbol{\theta}_k} \mathbf{D}' \mathbf{D}$. Difference matrices \mathbf{D} are constructed as in (6.9) and are used to measure roughness in the parameter-vectors. We choose second order differences, except for $\boldsymbol{\alpha}_A$, whose log-concave shape of the early-adult component suggests using $d = 3$.

As in the smooth LC variant, the smoothing parameter $\lambda_{\boldsymbol{\theta}_k}$ controls the roughness of the vector $\boldsymbol{\theta}_k$. Given the complexity and the flexibility of the 3C-sLC model, information criteria (e.g. BIC) do not lead to reasonable outcomes in terms of component-specific patterns. We thus opt for a choice of the smoothing parameters which relies on the prior information that users have about mortality development. However, we establish and suggest a rule of thumb which has worked for most of populations in the [Human Mortality Database \(2019\)](#). First, in order to reduce the computational burden, we assume that the smoothing parameter is the same within each triplet of parameters, i.e. $\lambda_{\boldsymbol{\alpha}_C} = \lambda_{\boldsymbol{\alpha}_A} = \lambda_{\boldsymbol{\alpha}_S}$ and $\lambda_{\boldsymbol{\beta}_C} = \lambda_{\boldsymbol{\beta}_A} = \lambda_{\boldsymbol{\beta}_S}$. However, we do not impose restrictions on $\lambda_{\boldsymbol{\theta}_k}$ across different triplets, i.e. in general $\lambda_{\boldsymbol{\alpha}_k} \neq \lambda_{\boldsymbol{\beta}_k}$. Then, we start from really large smoothing parameters (e.g. 10^8) and gradually reduce their values until the procedure provides reasonable outcomes without encountering convergence problems.

We believe that allowing users to tune the amount of smoothness, and consequently the effective dimension of the estimated 3C-sLC, is an asset of our approach, and not a drawback. On the one hand, users can adapt the model to diverse situations when in fact most of the alternative methods impose a number of parameters regardless the complexity of the data. On the other hand, more flexible models could be obtained by relaxing the previous procedure for selecting the smoothing parameters. Obviously, this additional flexibility would come at the price of a more complex selection of the smoothing parameters.

In demography, age 0 is commonly treated differently, e.g. in classic life-table construction ([Chiang, 1984](#)). This age constitutes a discontinuity in the mortality trajectory and we incorporate this feature, allowing discontinuity in the childhood component for the parameters operating over ages: $\boldsymbol{\alpha}_C$ and $\boldsymbol{\beta}_C$. This feature is obtained by not penalizing both $\alpha_{1,C}$ and $\beta_{1,C}$. Moreover, additional demographic knowledge is included in the model without enforcing a specific parametric structure of the parameter vectors. In particular, we impose the average shape of the age profile for early-adult and senescent components to be strictly log-concave and increasing with age, respectively. Such additional constraints can be implemented by a second penalty for the respective $\boldsymbol{\alpha}_A$ and $\boldsymbol{\alpha}_S$. Introduced by [Bollaerts et al. \(2006\)](#), these specialized penalties has proven useful in other demographic applications ([Camarda et al., 2016; Remund et al., 2018; Camarda, 2019](#)). Moreover incorporating additional knowledge about early-Adulthood and Senescent components is often necessary to ensure the identification of a complex model such as the 3C-sLC.

Finally the choice of the starting values is not as crucial as the non-linearity of the 3C-sLC model would suggest: the LC structure within each component ensures convergence of the proposed iterative penalized CLM algorithm from rather general starting parameters.

Inference and computation

Point estimates are only part of the results produced by a forecasting method. Uncertainty about the future needs to be measured properly. Conventionally, LC prediction intervals are obtained by applying uncertainty in future values of the extrapolated vector κ : upper and lower bounds for κ in t_F (see gray area in right panel of Figure 6.3) are combined with estimated α and β to compute prediction intervals for future death rates. Alternatively, instead of focusing on the variability in the time-varying parameter, [Brouhns et al. \(2005\)](#) and [Koissi et al. \(2006\)](#) proposed incorporating variability from all LC parameters using a bootstrap approach.

Here we combine both options to account for both sources of uncertainty, as previously suggested by [Keilman and Pham \(2006\)](#). Uncertainty in future mortality comes from variability in the estimated α_k , β_k and κ_k as well as from variability in the ARIMA model when extrapolation is performed on time-indexes κ_k .

In order to achieve this exhaustive measure of uncertainty, we first carry out a residual bootstrap procedure as in [Koissi et al. \(2006\)](#) for the LC framework, and [Ouellette et al. \(2012\)](#) and [Camarda \(2019\)](#) within a smoothing context. This allows us to simulate bootstrapped matrices of death counts $\mathbf{Y}_{(b)}$. We replicate this step 50 times and for each new $\mathbf{Y}_{(b)}$, together with the original matrix of exposure E , we estimate the 3C-sLC model. This procedure yields 50 new series of α_k , β_k and κ_k . Confidence intervals for the estimated parameters as well as for the fitted values during the past observed years can be obtained within this step.

We then apply a RWD model to each new vector κ_k . Instead of taking upper and lower bounds from the 50 estimated RWD models, we simulate 100 future κ_k based on the fitted drift parameter. By combining the 50 estimated time-indexes from the bootstrap procedure and the 100 extrapolated κ_k from the RWD simulation step, we obtain 5,000 new κ_k . From the distributions of future κ_k , we compute 5,000 matrices of future death rates for both overall and component-specific mortality. Empirical percentiles and confidence intervals can be extracted for death rates, as well as for any desirable summary measure of mortality, e.g., life expectancy at birth (cf. Section 6.3).

Finally, routines developed to model, decompose and forecast mortality data by a 3C-sLC model were implemented in R ([R Development Core Team, 2019](#)) and are publicly available¹. Using only basic R packages and for a single population, fitting a 3C-sLC model takes about 75 seconds, and it takes about 35 minutes to run the whole procedure to obtain confidence and prediction intervals for future mortality trends (portable personal computer, Intel i5-6300U processor, 2.4 GHz × 4 and 6 Gbytes random-access memory). With regard to computer data storage, a file of about 18 MB can store the R workspace with all outcomes for a single population.

¹ https://osf.io/rgf8y/?view_only=1608f0f8d1e9466cb33733bb03057b80. Note that the open science framework link is currently made anonymous for peer review.

6.3 Results

In this Section, we present the outcomes obtained from the proposed 3C-sLC model. First, we start by showing the estimated parameters and the age-specific results for the Swiss male population analysed in Section 6.2. Next, we present fitted and forecast summary measures of mortality for the four populations investigated in this article. Finally, we perform an exercise that demonstrates one demographic advantage of the 3C-sLC model: we estimate the potential gains derived from the hypothetical scenario of eliminating one or more mortality components from the age-pattern of mortality.

Figure 6.3 shows the estimated parameters of the 3C-sLC and the smooth variant of the Lee-Carter (LC) model for Swiss males. In the left panel, the average shape of mortality over age, captured by the parameter α in the LC model, is decomposed into component-specific average patterns. The three α_k display the expected shapes over age: decreasing, log-concave and increasing mortality with age for the Childhood, early-Adulthood and Senescent components, respectively. In the central and right panels, component-specific rates of mortality improvements β_k and time indexes κ_k are presented, together with estimates from the LC model. Here, it is interesting to observe the similarity between the LC estimates and the Senescent ones. Furthermore, the Childhood and Adulthood parameters capture additional dimensions of mortality developments that are overlooked and conflated in the LC framework.

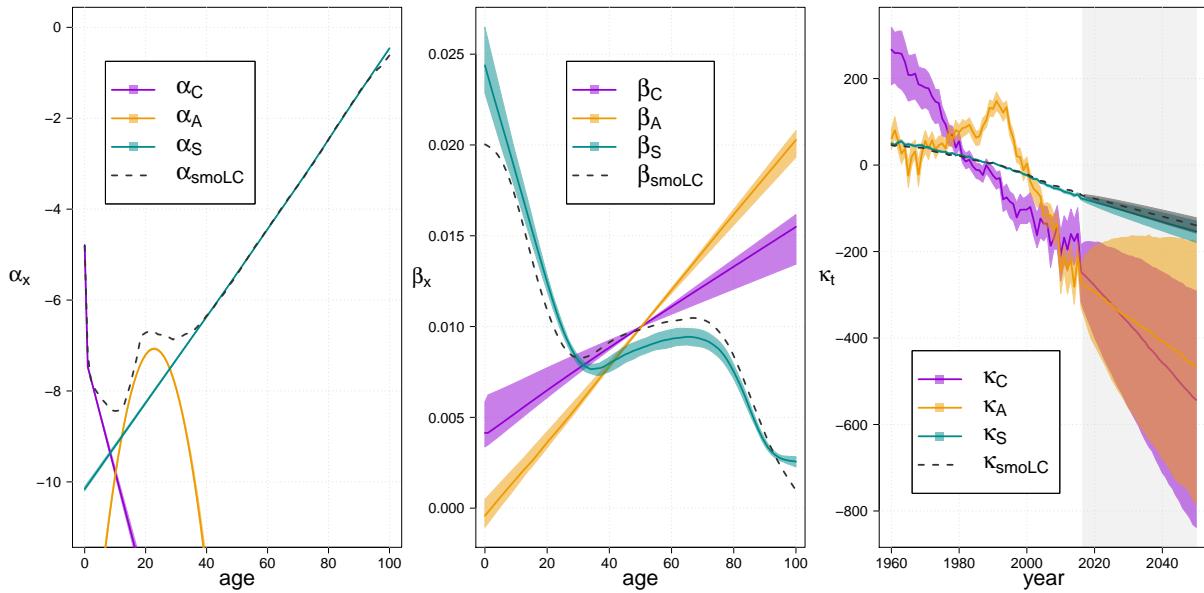


Figure 6.3: Estimated parameters from 3C-sLC model and a smooth variant of the Lee-Carter. Swiss males, ages 0-100, years 1960-2016, time index(es) κ forecast up to 2050 by RWD with 80% prediction intervals. Left panel: α , average shape of the age profile. Central panel: β , age-specific average rates of mortality decline. Right panel: κ , level of mortality at time t . For 3C-sLC parameters, 80% confidence intervals are also computed for the observed years.

Combining estimates and forecasts of the 3C-sLC parameters allows us to derive death rates for all ages and years of interest. Figure 6.4 shows estimated and forecast 3C-sLC death rates on a log scale for Swiss males. This figure can be directly compared with Figure 6.2, where the respective results are shown for the smooth LC model. Several observations can be drawn from comparing the two figures. In the left panel, fitted and forecast mortality age-profiles of

the 3C-sLC seem more plausible than LC ones, lacking the variable behavior and the constant age-specific rate of mortality decline of the smooth LC model. In the right panel, the superior fit of the 3C-sLC is evident, particularly at younger ages; this results from relaxing the assumption of a constant β in the LC framework. Age-specific mortality improvements are more flexible and produced by the combination of the component-specific parameters. The increased flexibility further translates into wider prediction intervals, which capture the uncertainty behind mortality developments at different ages.

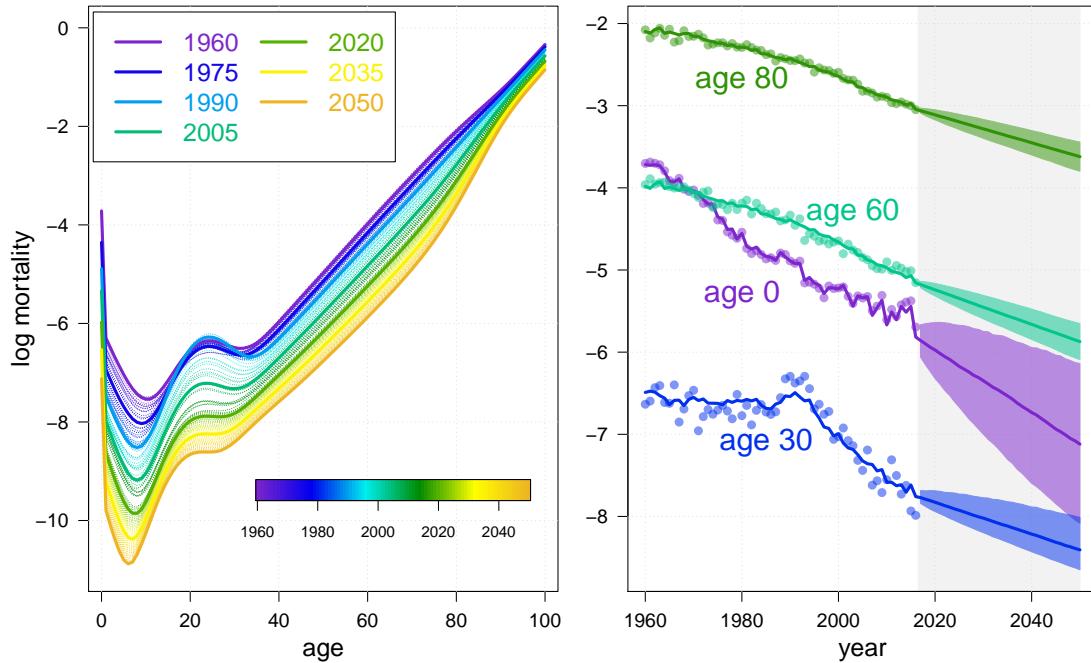


Figure 6.4: Actual, estimated and forecast death rates by 3C-sLC over age and time on a log scale for Swiss males. Ages 0-100, fitted years 1960-2016, forecast years 2017-2050. Left panel: rates over age for all years. Selected years are depicted by thicker lines. Right panel: selected ages over years.

A more rigorous assessment of the two models' goodness-of-fit is provided by comparing their deviance residuals, which are often employed as a diagnostic tool in Generalized Linear Models (GLM) settings (for an overview and the formulas for their computation, please refer to [McCullagh and Nelder, 1989](#)). Figure 6.5 shows the Poisson deviance residuals of the two models: the 3C-sLC displays smaller residuals than the smooth LC model, especially at childhood and early-adult ages. The red and blue clouds (corresponding to model misfit) at ages 20-40 of the smooth LC are considerably reduced in the 3C-sLC model.

The superior fit of the 3C-sLC is a direct result of the mortality decomposition, which allows one to better describe the age-pattern of mortality. Figure 6.6 displays the fitted and forecast Childhood, early-Adulthood and Senescent components on a log scale for Swiss males. The shapes observed in the α_k parameters are directly carried into the mortality components. Furthermore, the central panel clearly shows the increase in early-Adulthood mortality during the years 1980-1990s, related to the strong HIV epidemic that hit this population at that time ([Csete and Grob, 2012](#)).

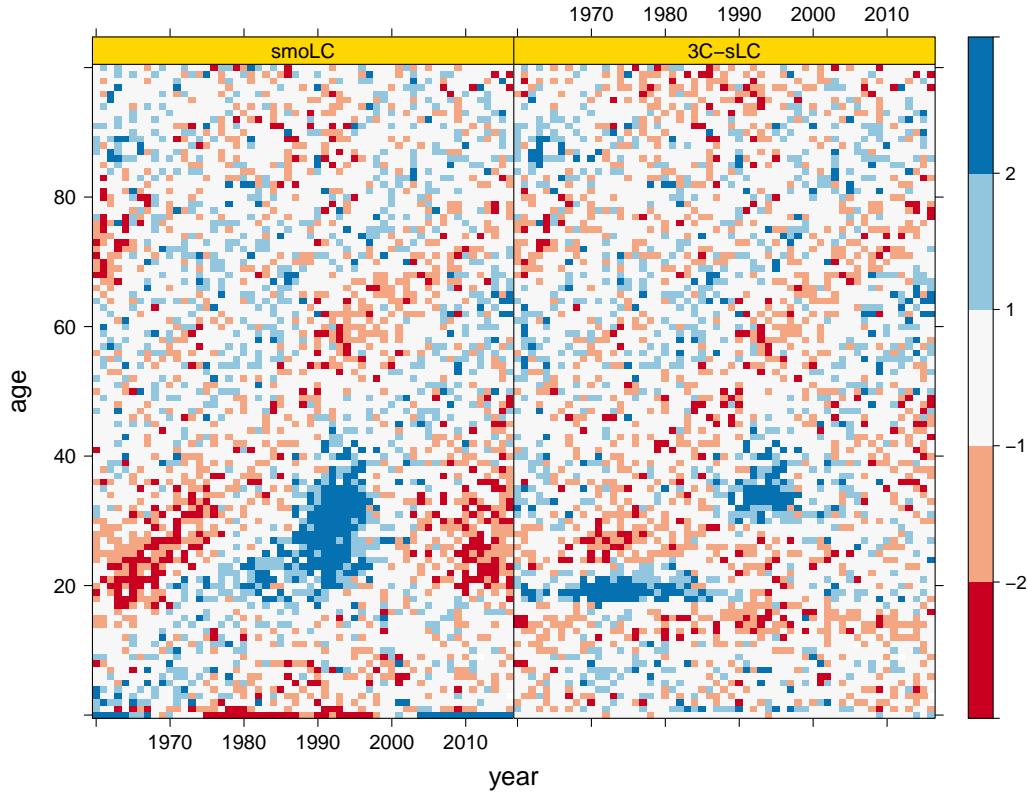


Figure 6.5: Poisson deviance residuals of the smooth LC and 3C-sLC models for Swiss males. Ages 0–100, fitted years 1960–2016.

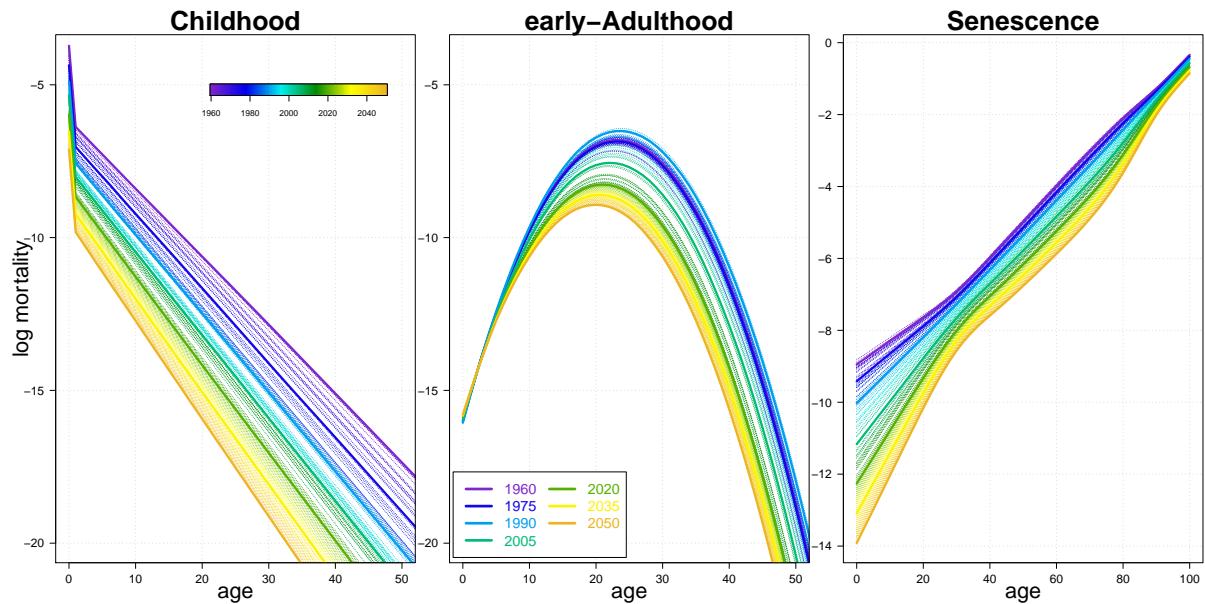


Figure 6.6: Estimated and forecast Childhood, Adulthood and Senescent components by 3C-sLC over age on a log scale for Swiss males. Ages 0–100, fitted years 1960–2016, forecast years 2017–2050. Axes of the left and central panels have been truncated for better displaying the mortality components.

Next, we show the results of applying the 3C-sLC and smooth LC models to the four populations introduced in Section 6.2, namely females and males in Switzerland (CHE) and the USA. We start by comparing the BIC values of the two models. The BIC is often used as a metric for

model selection as it conveniently captures the trade-off between goodness-of-fit and parsimony of a model. Table 6.1 reports the BIC values for the four populations and the two models, as well as the Deviance and effective dimension (which are used to compute the BIC).

	Deviance		ED		BIC	
	smoLC	3C-sLC	smoLC	3C-sLC	smoLC	3C-sLC
CHE males	9956	8116	160	316	11340	10851
CHE females	7631	7312	155	317	8975	10056
USA males	174910	161476	221	311	176822	164170
USA females	91825	93005	242	308	93916	95674

Table 6.1: Deviance, effective dimension (ED) and BIC values for the smooth LC and the 3C-sLC models for females and males in Switzerland (CHE) and the USA. Ages 0-100, fitted years 1960-2016. Lower BIC values (in bold) correspond to a better model in terms of trade-off between accuracy and parsimony.

The table shows that the 3C-sLC is chosen over the smooth LC model by the BIC for the two male populations. Furthermore, the 3C-sLC fits the observed data better than the smooth LC model for three populations (lower values of Deviance, except for USA females), and this superior fit is a consequence of the higher number of parameters (effective dimension) employed by the 3C-sLC model.

Furthermore, we present outcomes of the 3C-sLC and smooth LC in terms of two standard and complementary summary measures of mortality. We use life expectancy at birth e_0 as a traditional measure of longevity and population health ([Preston et al., 2001](#)), and the average number of life years lost at birth e_0^\dagger ([Vaupel and Canudas-Romo, 2003](#)) as a measure of lifespan disparity. In formulas:

$$e_0 = \frac{\sum_0^\omega \ell_x}{\ell_0} \quad \text{and} \quad e_0^\dagger = \frac{\sum_0^\omega e_x d_x}{\ell_0}, \quad (6.18)$$

where ℓ_x denotes the life-table probability of surviving from birth to age x , ℓ_0 is the life-table radix, and ω is the highest age attained in the life table. The vectors e_x and d_x denote the remaining life expectancy and the age-at-death distribution at age x , respectively. Whereas life expectancy at birth has a long tradition in measuring longevity, e_0^\dagger has been recently employed to assess the degree of lifespan inequality within and between populations (see, e.g., [Shkolnikov et al., 2011a](#); [Vaupel et al., 2011](#); [Aburto and van Raalte, 2018](#)), and it has also been used to evaluate mortality forecasts ([Bohk-Ewald et al., 2017](#); [Camarda, 2019](#)).

Figure 6.7 shows the actual e_0 and e_0^\dagger for the four populations, together with the estimated and forecast values with 80% prediction intervals for the smooth LC and 3C-sLC models. The recent flattening in e_0 as well as the increase in e_0^\dagger in the USA populations, mostly caused by the increase in drug overdoses and suicides ([Barbieri, 2018](#)), is evident from the figure. The wider prediction intervals of the 3C-sLC death rates (cf. Figure 6.4) translate into slightly larger intervals also for the summary mortality measures. Moreover, it is interesting to observe that forecast e_0 from both models are very similar, whereas forecast e_0^\dagger differ: specifically, the 3C-sLC predicts slower mortality compression (i.e. higher lifespan disparity) than the smooth LC. Finally, the figure shows that the models successfully capture the observed trends in e_0 , while they are less successful for e_0^\dagger : nevertheless, the 3C-sLC performs better than the smooth LC, particularly for the Swiss populations.

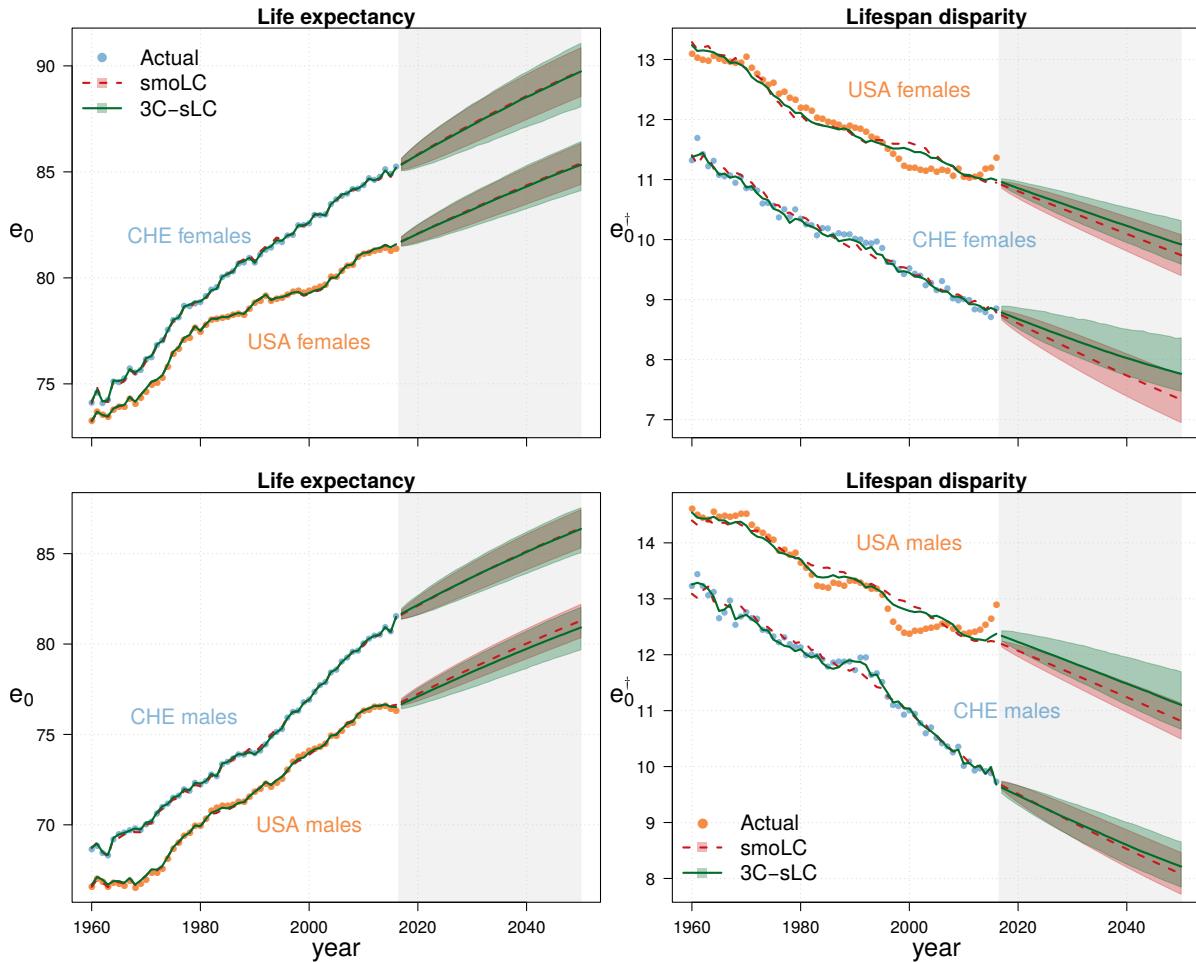


Figure 6.7: Actual, estimated and forecast life expectancy at birth (e_0 , left panels) and lifespan disparity (e_0^\dagger , right panels) by smooth Lee-Carter (red lines and shades) and 3C-sLC (dark green lines and shades) with 80% prediction intervals for females and males in Switzerland (CHE) and the USA. Ages 0-100, fitted years 1960-2016, forecast years 2017-2050.

In order to emphasize the uniqueness of the 3C-sLC model in terms of possible outcomes, we evaluate the potential gains resulting from the hypothetical scenarios of removing one or two mortality components from the age-pattern of mortality. Specifically, we estimate the potential gains in e_0 and reductions in e_0^\dagger derived from the elimination of: (i) the accident hump alone, and (ii) the accident hump together with the Childhood component. Within the 3C-sLC, these scenarios can be readily implemented by removing corresponding components (cf. Figure 6.6) from the overall mortality pattern. Figure 6.8 shows the estimated and forecast 3C-sLC summary measures of mortality for males (which have a more pronounced accident hump) in Switzerland and the USA, as well as the measures computed in the two hypothetical scenarios. Removing only the accident hump would result in an average increase over the fitted periods of 0.66 and 0.63 years in e_0 for males in Switzerland and the USA, respectively, and in an average reduction of 0.51 and 0.47 years in e_0^\dagger . If the Childhood component were eliminated too, gains in e_0 would rise to 1.60 and 1.78 years in the two populations, respectively, and reductions in e_0^\dagger would amount to 1.32 and 1.42 years.

Moreover, the 3C-sLC allows us to decompose summary measures in future years. For instance, we forecast USA males overall life expectancy in 2050 to reach 80.92 years, with a 80%

prediction interval of [79.68, 82.03]. The mean value of e_0 would increase to 81.64 if we removed only the early-Adulthood component, and to 81.89 if Childhood was excluded too. Smaller reductions are forecast for Swiss males: total life expectancy in 2050 is projected to 86.37 years [85.05, 87.54], which would increase to 86.54 years if the accident hump was removed, and to 86.65 if both components were theoretically eradicated.

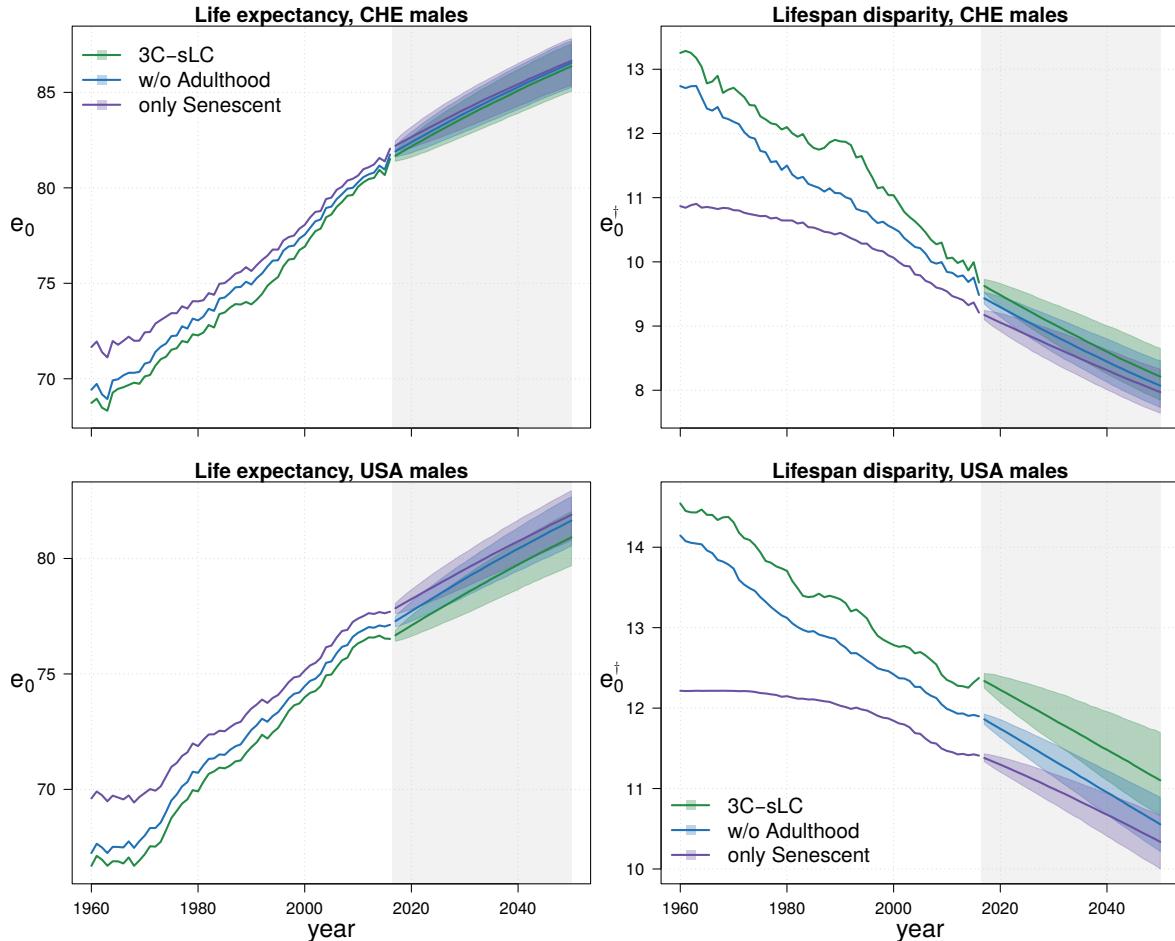


Figure 6.8: Estimated and forecast life expectancy at birth (e_0 , left panels) and lifespan disparity (e_0^\dagger , right panels) with 80% prediction intervals by: (i) 3C-sLC (green lines and shades), (ii) 3C-sLC removing the Early-Adulthood component (blue lines and shades), and (iii) 3C-sLC removing the Childhood and Early-Adulthood components (purple lines and shades). Males in Switzerland (CHE) and the USA, ages 0-100, fitted years 1960-2016, forecast years 2017-2050.

6.4 Discussion

Models to predict the future course of mortality of rather large populations have flourished during the most recent decades, stimulated by the influential contribution of [Lee and Carter \(1992\)](#). Although the number of available approaches to forecast mortality keeps increasing by the day, the Lee-Carter (LC) model arguably remains the most widely known and adopted methodology for this purpose. At the root of this success lie two important factors, namely power and simplicity, which have made LC the benchmark approach in mortality forecasting.

Despite its almost universal recognition and adoption, the original LC model suffers from some shortcomings (cf. Section 6.1). Several extensions of the model have been proposed to

mitigate its principal limitations (e.g., [Lee and Miller, 2001](#); [Booth et al., 2002](#); [Brouhns et al., 2002](#); [Renshaw and Haberman, 2006](#); [Li et al., 2013](#)). However, no attempts have been made to overcome all the recognized issues simultaneously.

In this article, we have introduced a novel generalization of the LC model that tackles all of its main limitations at once. Starting from the hypothesis that human mortality can be conceived as the sum of three independent components ([Thiele, 1871](#)), we propose a methodology in which mortality is simultaneously smoothed, decomposed and forecast within an LC framework. As such, we call our approach “Three-Component smooth Lee-Carter” (3C-sLC) model.

The 3C-sLC is embedded in a Poisson setting, which provides a reasonable assumption for the underlying mortality process. Smoothness in the outcomes is enforced within the estimation of the 3C-sLC parameters by employing a penalized likelihood approach. Overall mortality developments are described by the combination of three mortality components: Childhood, early-Adulthood and Senescent mortality. Each component is modelled and forecast with a smooth version of the LC model. As such, mortality developments are more flexible than in the original LC model, because they do not depend on the rather unreasonable assumption of a fixed rate of age-specific mortality improvement.

We have shown the results of fitting and forecasting mortality with the 3C-sLC model in four populations, namely females and males in Switzerland and the USA. We have compared the 3C-sLC outcomes with those of the improved smooth version of the LC model ([Delwarde et al., 2007](#)). We compared the 3C-sLC with this LC variant because both approaches are estimated in a Poisson setting with a penalized likelihood for enforcing smooth outcomes. Moreover, it should be noted that this comparison could have been made only at the overall mortality level: all LC variants as well as other existing forecasting models do not allow one to simultaneously forecast and decompose mortality into meaningful components. For the same reasons, additional comparisons and out-of-sample exercises would have assessed performances of the 3C-sLC only for one of its features, underrating its great explanatory value.

At the overall mortality level, the 3C-sLC death rates fit observed mortality developments better than the smooth LC model; moreover, 3C-sLC forecast rates are characterized by wider prediction intervals, which result from the enhanced flexibility of the model. This is a potential advantage of the 3C-sLC model, as LC prediction intervals have been criticized for being too narrow ([Alho, 1992](#)). Although the 3C-sLC generally fits better the observed mortality data, the smooth LC model was preferred by the BIC criterion for the female populations. A potential explanation for this is that the accident hump is not a relevant feature of the mortality pattern for females analysed in this article. We plan to investigate modelling and forecasting female mortality using only two components (i.e. Childhood and Senescence) in future work.

For both models, we have further computed and compared two complementary summary measures of mortality: life expectancy at birth (e_0) and lifespan disparity at birth (e_0^\dagger). The two measures provide information on the longevity and lifespan inequality of the population. Despite being characterized by rather similar fitted and forecast values of e_0 , the difference between the two models clearly emerges from e_0^\dagger . The 3C-sLC better captures observed trends in lifespan inequality, and it forecasts slower mortality compression than the smooth LC model. This is another advantage of the 3C-sLC, since LC forecasts have been shown to produce a (possibly)

too fast mortality compression (see, e.g., [Bardoutsos et al., 2018](#); [Basellini and Camarda, 2019b](#)). In terms of prediction intervals, the widening observed in 3C-sLC rates only partially carries forward to forecast e_0 and e_0^\dagger ; the reason is that larger 3C-sLC intervals occur at younger ages especially, whose influence on summary mortality measures is rather limited.

Comparing the two models' estimated parameters (cf. Figure 6.3) provides two interesting points for discussion. First, the 3C-sLC estimates of the Senescent component closely mirror those of the smooth LC model (with the exception of the $\hat{\alpha}$ values at young and early-adult ages). This hints at the fact that the LC model is driven by the Senescent component of mortality, which carries most weight in terms of death counts. Within the standard (or improved) LC framework, mortality developments of Childhood and early-Adulthood components are thus conflated with Senescent ones, resulting in a loss of fitting accuracy. The second point worthy of discussion concerns the variability of the 3C-sLC model. Figure 6.3 shows that confidence intervals for the α_k estimates are very narrow. Hence, uncertainty in the 3C-sLC model depends on the variability of the β_k and κ_k estimates.

The presence of three independent components could prompt a reader to establish an association between the 3C-sLC and LC variants with additional $\beta_x^{(i)} \kappa_t^{(i)}$ interaction terms ([Renshaw and Haberman, 2003a](#); [Hyndman and Ullah, 2007](#)). However, the two approaches present substantial differences in both methodological and demographic perspectives. Whereas the 3C-sLC attempts to decompose the mortality pattern into additive independent components, the LC model (with more interaction terms) tries to approximate it with additional terms that have decreasing explicative power. In statistical terms, the 3C-sLC models the expected values of the Poisson process, while the LC models the logarithm of the expected values (for a more detailed discussion in a general modelling framework, see [Camarda et al., 2016](#), p. 279).

The decomposition of mortality into its three independent components has received considerable interest during recent decades (cf. Section 6.1). However, only a few methodologies have been proposed to take into account the three components when projecting mortality. The well-established model of [Heligman and Pollard \(1980\)](#) has only been employed by [Forfar and Smith \(1987\)](#) in mortality projections due to the high number of parameters to forecast. Moreover, [Bardoutsos et al. \(2018\)](#) employed the ten-parameter CoDe model ([de Beer and Janssen, 2016](#)) to forecast the full age range of mortality. The authors find that accounting for the compression and delay (i.e. shift) of mortality improves point forecast accuracy with respect to the LC model when the modal age at death increases linearly. Finally, [Basellini and Camarda \(2019a\)](#) proposed the 3C-STAD model to analyze and forecast component-specific age-at-death distributions derived from a non-parametric decomposition of the mortality pattern. The authors show that 3C-STAD projections for high-longevity populations are more accurate in terms of point forecasts and prediction intervals than the LC model and its variants.

The mortality decomposition provides interesting demographic insights into fitted and forecast mortality schedules. The increased quantity of information on component-specific patterns allows one to obtain a more thorough overview on mortality developments. Furthermore, hypothetical exercises can readily be performed within this framework. For example, it is possible to estimate the potential gains in e_0 and reductions in e_0^\dagger resulting from the elimination of one or more mortality components. We have performed this exercise for the male populations analysed

in this article: for the observed period, we estimate an average 1.60 and 1.78 additional years of longevity in Switzerland and the USA, respectively, and an average 1.32 and 1.42 reduced years of lifespan disparity by theoretically eradicating the Childhood and early-Adulthood components. We present similar analysis for the forecast years too.

Two additional findings related to these exercises deserve mention. Firstly, Figure 6.8 shows that the Childhood and early-Adulthood components have a significant effect on e_0 and e_0^\dagger for both male populations during the fitted years only. In forecast years, the effect of the two components varies by population: in Switzerland, the two forecast scenarios converge to the baseline mortality pattern, while in the USA differences between the scenarios do not disappear. Forecast Childhood and early-Adulthood components thus seem to maintain a significant role only in the USA. Secondly, the bottom panels of Figure 6.8 highlight that lifespan disparity fluctuations are driven by the Childhood and early-Adulthood components: removing them from the age-pattern of mortality smooths the trends of e_0^\dagger in both populations.

6.5 Conclusions

The proposed 3C-sLC aims to provide demographers and actuaries with a new extension of the LC model that overcomes its main drawbacks while maintaining the factors that underpin its success. Conceptually, the 3C-sLC can be thought of as the combination of three LC models, each one targeted to model and forecast the Childhood, early-Adulthood and Senescent components of mortality. We thus keep the simplicity of the LC framework, namely the interpretability of the three set of parameters α_x , β_x and κ_t (one triplet for each mortality component). Furthermore, forecasting is performed as in the original LC model, without introducing additional layers of complexity. Although the formulas developed to fit the 3C-sLC are somewhat more elaborate than the original SVD decomposition suggested by [Lee and Carter \(1992\)](#), we provide the routines needed to fit and forecast mortality with our proposed model in the Supplementary Material to this article. Our hope is that the 3C-sLC could be employed by researchers and statistical agencies that currently use the LC model in an effort to obtain more desirable outcomes within the same LC framework.

Bibliography

- Aburto, J. M. and van Raalte, A. (2018). Lifespan Dispersion in Times of Life Expectancy Fluctuation: The Case of Central and Eastern Europe. *Demography*, 55(6):2071–2096.
- Aburto, J. M., Wensink, M., van Raalte, A., and Lindahl-Jacobsen, R. (2018). Potential gains in life expectancy by reducing inequality of lifespans in Denmark: an international comparison and cause-of-death analysis. *BMC Public Health*, 18(1):831.
- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.
- Alho, J. M. (1992). Modeling and Forecasting U.S. Mortality: Comment. *Journal of the American Statistical Association*, 87(419):673–674.
- Alho, J. M. (2000). “The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications”: Discussion. *North American Actuarial Journal*, 4(1):91–93.
- Alho, J. M. and Spencer, B. (2005). *Statistical Demography and Forecasting*. Springer New York.
- Anderson, J. L. and Dow, J. B. (1948). *Actuarial Statistics. Volume II: Construction of mortality & other tables*. Cambridge Univeristy Press.
- Andreev, K. F. (2002). *Evolution of the Danish population from 1835 to 2000*. Odense Monographs on Population Aging. Odense University Press, Odense - DK.
- Barbieri, M. (2018). The contribution of drug-related deaths to the US disadvantage in mortality. *International Journal of Epidemiology*, 48(3):945–953.
- Barbieri, M., Wilmoth, J. R., Shkolnikov, V. M., Glei, D., Jasillionis, D., Jdanov, D., Boe, C., Riffe, T., Grigoriev, P., and Winant, C. (2015). Data Resource Profile: The Human Mortality Database (HMD). *International Journal of Epidemiology*, 44(5):1549–1556.
- Bardoutsos, A., de Beer, J., and Janssen, F. (2018). Projecting delay and compression of mortality. *Genus*, 74. Article number: 17.
- Basellini, U. and Camarda, C. G. (2019a). A Three-component Approach to Model and Forecast Age-at-death Distributions. In Mazzuco, S. and Keilman, N., editors, *Developments in Demographic Forecasting*. Springer (forthcoming).
- Basellini, U. and Camarda, C. G. (2019b). Modelling and forecasting adult age-at-death distributions. *Population Studies*, 73(1):119–138.
- Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. *Biological Aspects of Demography*, 999:57–68.
- Bell, W. R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, 13(3):279–303.

- Bengtsson, T., Keilman, N., Alho, J. M., Christensen, K., Palmer, E., and Vaupel, J. W. (2019). Introduction. In *Old and New Perspectives on Mortality Forecasting*, pages 1–19. Springer Open.
- Bennett, S. (1983). Log-logistic regression models for survival data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 32(2):165–171.
- Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., and Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37:527–566.
- Bergeron-Boucher, M.-P., Canudas-Romo, V., Pascariu, M., and Lindahl-Jacobsen, R. (2018). Modeling and forecasting sex differences in mortality: a sex-ratio approach. *Genus*, 74. Article number: 20.
- Bergeron-Boucher, M.-P., Ebeling, M., and Canudas-Romo, V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research*, 33:391–424.
- Bergeron-Boucher, M.-P., Kjærgaard, S., Oeppen, J., and Vaupel, J. W. (2019). The impact of the choice of life table statistics when forecasting mortality. *Demographic Research*, 41:1235–1268.
- Blake, D., Boardman, T., and Cairns, A. (2014). Sharing longevity risk: Why governments should issue longevity bonds. *North American Actuarial Journal*, 18(1):258–277.
- Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan Disparity as an Additional Indicator for Evaluating Mortality Forecasts. *Demography*, 54(4):1559–1577.
- Bollaerts, K., Eilers, P. H. C., and van Mechelen, I. (2006). Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology*, 59:451–469.
- Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography*, 42(1):23–49.
- Bongaarts, J. and Feeney, G. (2002). How long do we live? *Population and Development Review*, 28(1):13–29.
- Bonynge, F. (1852). *The future wealth of America*. New York.
- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3):547–581.
- Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56(3):325–336.
- Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(1-2):3–43.
- Borgan, Ø. and Keilman, N. (2019). Do Japanese and Italian Women Live Longer than Women in Scandinavia? *European Journal of Population*, 35(1):87–99.

- Brass, W. (1971). On the Scale of Mortality. In Brass, W., editor, *Biological Aspect of Demography*, pages 69–110. Taylor & Francis, London.
- Brillinger, D. R. (1986). A biometrics invited paper with discussion: The natural variability of vital rates and associated statistics. *Biometrics*, 42(4):693–734.
- Brouhns, N., Denuit, M., and Van Keilegom, I. (2005). Bootstrapping the Poisson log- bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 3:212–224.
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.
- Cadena, M. and Denuit, M. (2016). Semi-parametric accelerated hazard relational models with applications to mortality projections. *Insurance: Mathematics and Economics*, 68:1–16.
- Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.
- Camarda, C. G. (2008). *Smoothing methods for the analysis of mortality development*. PhD thesis, Universidad Carlos III de Madrid, Department of Statistics.
- Camarda, C. G. (2012). MortalitySmooth: An R Package for Smoothing Poisson Counts with P-Splines. *Journal of Statistical Software*, 50:1–24. Available on www.jstatsoft.org/v50/i01.
- Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Demographic Research*, 41:1091–1130.
- Camarda, C. G., Eilers, P. H., and Gampe, J. (2008). A Warped Failure Time Model for Human Mortality. In *Proceedings of the 23rd International Workshop of Statistical Modelling*, pages 149–154.
- Camarda, C. G., Eilers, P. H., and Gampe, J. (2016). Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling*, 16(4):279–296.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19:1179–1204.
- Canudas-Romo, V. (2010). Three measures of longevity: Time trends and record values. *Demography*, 47(2):299–312.
- Canudas-Romo, V. and Schoen, R. (2005). Age-specific contributions to changes in the period and cohort life expectancy. *Demographic Research*, 13:63–82.

- Chatfield, C. (2000). *Time-series forecasting*. Chapman & Hall/CRC: London.
- Cheung, S. L. K. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of Japan. *Population Studies*, 61(1):85–97.
- Cheung, S. L. K., Robine, J.-M., Paccaud, F., and Marazzi, A. (2009). Dissecting the compression of mortality in Switzerland, 1876–2005. *Demographic Research*, 21:569–598.
- Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–258.
- Chiang, C. L. (1984). *The Life Table and its Applications*. Malabar, FL: Krieger.
- Chiou, J.-M. and Müller, H.-G. (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association*, 104(486):572–585.
- Christensen, K., Davidsen, M., Juel, K., Mortensen, L., Rau, R., and Vaupel, J. W. (2010). The Divergent Life-Expectancy Trends in Denmark and Sweden - and Some Potential Explanations. In Crimmins, E. M., Preston, S. H., and Cohen, B., editors, *International Differences in Mortality at Older Ages: Dimensions and Sources*. National Academies Press, Washington (DC).
- Clarkson, D. B., Fraley, C., Gu, C. C., and Ramsay, J. O. (2005). *S+functional data analysis: User's manual for windows*. Springer, New York.
- CMI Committee (1974). Considerations affecting the preparation of standard tables of mortality. *Journal of the Institute of Actuaries*, 101(2):135–216.
- CMI Committee (1976). The graduation of pensioners and of annuitants mortality experience 1967–70. Revised Working Paper 20 (November 2007), Institute of Actuaries and the Faculty of Actuaries.
- CMI Committee (2007). Stochastic projection methodologies: Further progress and *P*-Spline model features, example results and implications. Technical report, Institute of Actuaries and the Faculty of Actuaries.
- Coale, A. J., Demeny, P., and Vaughan, B. (1983). *Regional model life tables and stable populations*. New York, NY: Academic Press, second edition.
- Colchero, F., Rau, R., Jones, O. R., Barthold, J. A., Conde, D. A., Lenart, A., Nemeth, L., Scheuerlein, A., Schooley, J., Torres, C., Zarulli, V., Altmann, J., Brockman, D. K., Bronikowski, A. M., Fedigan, L. M., Pusey, A. E., Stoinski, T. S., Strier, K. B., Baudisch, A., Alberts, S. C., and Vaupel, J. W. (2016). The emergence of longevous populations. *Proceedings of the National Academy of Sciences*, 113(48):E7681–E7690.
- Cole, L. C. (1957). Sketches Of General and Comparative Demography. *Cold Spring Harbor Symposia on Quantitative Biology*, 22(0):1–15.

- Congdon, P. (1993). Statistical Graduation in Local Demographic Analysis and Projection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156(2):237–270.
- Csete, J. and Grob, P. J. (2012). Switzerland, HIV and the power of pragmatism: Lessons for drug policy development. *International Journal of Drug Policy*, 23(1):82–86.
- Currie, I. D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, 13(1):69–93.
- Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 4:356–383.
- Currie, I. D., Durban, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.
- Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics & Economics*, 36:260–284.
- Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, 27(1):65–81.
- de Beer, J. and Janssen, F. (2016). A new parametric model to assess delay and compression of mortality. *Population Health Metrics*, 14. Article number: 46.
- de Jong, P. and Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, 13:1–18.
- Dellaportas, P., Smith, A., and Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of Royal Statistical Society. Series A*, 164:275–291.
- Delwarde, A., Denuit, M., and Eilers, P. (2007). Smoothing the Lee–Carter and Poisson log–bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical Modelling*, 7(1):29–48.
- DeMoivre, A. (1725). Annuities on Lives: Or, the Valuation of Annuities Upon Any Number of Lives as also of Reversions. *William Person, London*.
- Diaz, G., Debón, A., and Giner-Bosch, V. (2018). Mortality forecasting in Colombia from abridged life tables by sex. *Genus*, 74. Article number: 15.
- Dokumentov, A., Hyndman, R. J., and Tickle, L. (2018). Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat*, 7(1):e199.
- Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2010a). Backtesting stochastic mortality models: an ex post evaluation of multiperiod-ahead density forecasts. *North American Actuarial Journal*, 14(3):281–298.
- Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. (2010b). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47(3):255–265.

- Ebeling, M. (2018). How Has the Lower Boundary of Human Mortality Evolved, and Has It Already Stopped Decreasing? *Demography*, 55:1887–1903.
- Edwards, R. D. (2013). The cost of uncertain life span. *Journal of Population Economics*, 26(4):1485–1522.
- Edwards, R. D. and Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, 31(4):645–674.
- Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Eilers, P. H. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, 7(3):239–254.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with B -splines and Penalties (with discussion). *Statistical Science*, 11(2):89–102.
- Fernández, C. and Steel, M. F. J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441):359–371.
- Forfar, D. O. (2004). *Encyclopedia of actuarial science*, chapter Mortality laws, pages 1139–1145. John Wiley & Sons.
- Forfar, D. O., McCutcheon, J. J., and Wilkie, A. D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115(1):1–149.
- Forfar, D. O. and Smith, D. M. (1987). The changing shape of English life tables. *Transactions of the Faculty of Actuaries*, 40:98–134.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303(3):130–135.
- Gage, T. B. (1993). The Decline of Mortality in England and Wales 1861 to 1964: Decomposition by Cause of Death and Component of Mortality. *Population Studies*, 47(1):47–66.
- Gampe, J. (2010). Human mortality beyond age 110. In *Supercentenarians*, pages 219–230. Springer.
- Gigliarano, C., Basellini, U., and Bonetti, M. (2017). Longevity and concentration in survival times: the log-scale-location family of failure time models. *Lifetime Data Analysis*, 23(2):254–274.
- Gini, C. (1912). Variabilità e mutabilità. Contributi allo studio delle relazioni e delle distribuzioni statistiche. *Studi Economico-Giuridici della Università di Cagliari*.
- Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, 73:1203–1248.
- Girosi, F. and King, G. (2007). Understanding the Lee-Carter Mortality Forecasting Method. Technical report, RAND Corporation.

- Girosi, F. and King, G. (2008). *Demographic Forecasting*. Princeton University Press.
- Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151.
- Goldstein, J. R. (2011). A Secular Trend toward Earlier Male Sexual Maturity: Evidence from Shifting Ages of Male Young Adult Mortality. *PLoS ONE*, 6(8):e14826. doi:10.1371/journal.pone.0014826.
- Goldstein, J. R. and Wachter, K. W. (2006). Relationships between period and cohort life expectancy: Gaps and lags. *Population Studies*, 60(3):257–269.
- Gompertz, B. (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*, 115:513–583.
- Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. London. American edition edited by Walter F. Willcox, John Hopkins Press, Baltimore, 1939.
- Gupta, A. and Székely, G. (1994). On location and scale maximum likelihood estimators. *Proceedings of the American Mathematical Society*, 120(2):585–589.
- Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55.
- Haberman, S. and Renshaw, A. E. (2008). Mortality, longevity and experiments with the Lee–Carter model. *Lifetime Data Analysis*, 14:286–315.
- Halley, E. (1693). An estimate of the degrees of the mortality of mankind. *Philosophical Transactions*, 17:596–610.
- Hamilton, J. D. (1994). *Time Series Analysis*, volume 2. Princeton University Press.
- Hanada, K. (1983). A formula of Gini's concentration ratio and its application to life tables. *Journal of the Japan Statistical Society, Japanese Issue*, 13(2):95–98.
- Hartmann, M. (1987). Past and recent attempts to model mortality at all ages. *Journal of Official Statistics*, 3(1):19–36.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer.
- Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.
- Heuveline, P. (2002). An International Comparison of Adolescent and Young Adult Mortality. *The ANNALS of the American Academy of Political and Social Science*, 580(1):172–200.
- Horiuchi, S., Ouellette, N., Cheung, S. L. K., and Robine, J.-M. (2013). Modal age at death: lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, 11:37–69.

Human Mortality Database (2019). 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: 23 August 2019 (Chapter 1), 25 May 2018 (Chapter 2), 22 March 2018 (Chapter 3), 15 February 2019 (Chapter 4), 8 May 2019 (Chapter 5), July 2019 (Chapter 6).

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., and Yasmeen, F. (2018a). *forecast: Forecasting functions for time series and linear models*. R package version 8.4.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3):1–22.

Hyndman, R. J. and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24(3):323–342.

Hyndman, R. J., Booth, H., Tickle, L., and Maindonald, J. (2018b). *demography: Forecasting Mortality, Fertility, Migration and Population Data*. R package version 1.21.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1):261–283.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

Jacobsen, R., Keiding, N., and Lynge, E. (2002). Long term mortality trends behind low life expectancy of Danish women. *Journal of Epidemiology & Community Health*, 56(3):205–208.

Jacobsen, R., Keiding, N., and Lynge, E. (2006). Causes of death behind low life expectancy of Danish women. *Scandinavian Journal of Public Health*, 34(4):432–436.

Jacobsen, R., Osler, M., Lynge, E., and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different? *European Journal of Epidemiology*, 19(2):117–121.

Janssen, F. (2018). Advances in mortality forecasting: introduction. *Genus*, 74. Article number: 21.

Janssen, F. and de Beer, J. (2016). Projecting future mortality in the Netherlands taking into account mortality delay and smoking. In *Joint Eurostat/UNECE Work Session on Demographic Projections*.

Janssen, F. and de Beer, J. (2019). The timing of the transition from mortality compression to mortality delay in Europe, Japan and the United States. *Genus*, 75. Article number: 10.

Janssen, F., van Wissen, L. J. G., and Kunst, A. E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, 50(4):1341–1362.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions*, volume 2. Wiley series in probability and mathematical statistics: applied probability and statistics.

- Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, second edition.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research*, 3:doi: 10.4054/DemRes.2000.3.6.
- Kannisto, V. (2001). Mode and dispersion of the length of life. *Population: An English Selection*, 13:159–171.
- Kannisto, V., Lauritsen, J., Thatcher, A. R., and Vaupel, J. W. (1994). Reductions in mortality at advanced ages: several decades of evidence from 27 countries. *Population and Development Review*, 20(4):793–810.
- Keilman, N. (2008). European Demographic Forecasts Have Not Become More Accurate over the Past 25 Years. *Population and Development Review*, 34(1):137–153.
- Keilman, N. (2019). Mortality shifts and mortality compression in period and cohort life tables. *Demographic Research*, 41:1147–1196.
- Keilman, N. and Pham, D. (2006). Prediction intervals for Lee-Carter-based mortality forecasts. In *European Population Conference 2006, Liverpool, June 21–24, 2006*.
- Keyfitz, N. (1982). Choice of function for mortality analysis: Effective forecasting depends on a minimum parameter representation. *Theoretical Population Biology*, 21(3):329–352.
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, second edition.
- Koissi, M.-C. and Shapiro, A. F. (2006). Fuzzy formulation of the Lee–Carter model for mortality forecasting. *Insurance: Mathematics and Economics*, 39:287–309.
- Koissi, M.-C., Shapiro, A. F., and Högnäs, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1):1–20.
- Kontis, V., Bennett, J. E., Mathers, C. D., Li, G., Foreman, K., and Ezzati, M. (2017). Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *The Lancet*, 389(10076):1323–1335.
- Kostaki, A. (1992). A nine-parameter version of the Heligman-Pollard formula. *Mathematical Population Studies*, 3(4):277–288.
- Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*, volume 362. John Wiley & Sons.
- Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549.
- Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671.

- Lenart, A. and Missov, T. I. (2016). Goodness-of-fit tests for the Gompertz distribution. *Communications in Statistics-Theory and Methods*, 45(10):2920–2937.
- Levitis, D. A. (2011). Before senescence: the evolutionary demography of ontogenesis. *Proceedings of the Royal Society B: Biological Sciences*, 278(1707):801–809.
- Lexis, W. H. R. A. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. Strassburg: Trübner.
- Lexis, W. H. R. A. (1878). Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de Démographie Internationale*, 2(5):447–460.
- Li, H. and Li, J. S. (2017). Optimizing the Lee-Carter Approach in the Presence of Structural Changes in Time and Age Patterns of Mortality Improvements. *Demography*, 54(3):1073–1095.
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3):575–594.
- Li, N., Lee, R., and Gerland, P. (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051.
- Lindahl-Jacobsen, R., Oeppen, J., Rizzi, S., Möller, S., Zarulli, V., Christensen, K., and Vaupel, J. W. (2016a). Why did Danish women's life expectancy stagnate? The influence of interwar generations' smoking behaviour. *European Journal of Epidemiology*, 31(12):1207–1211.
- Lindahl-Jacobsen, R., Rau, R., Jeune, B., Canudas-Romo, V., Lenart, A., Christensen, K., and Vaupel, J. W. (2016b). Rise, stagnation, and rise of Danish women's life expectancy. *Proceedings of the National Academy of Sciences*, 113(15):4015–4020.
- Lotka, A. J. (1939). Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particulière à l'espèce humaine. *Actualités Scientifiques et Industrielles*, (780).
- Lynch, S. M. and Brown, J. S. (2001). Reconsidering mortality compression and deceleration: An alternative model of mortality rates. *Demography*, 38(1):79–95.
- Maddison, A. (2006). The Contours of World Development. In *The World Economy. Volume 1: A Millennial Perspective and Volume 2: Historical Statistics*. OECD Publishing.
- Makeham, W. M. (1860). 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.
- Makridakis, S., Hyndman, R. J., and Petropoulos, F. (2019). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, In press.
- Manton, K. G., Singer, B. H., and Suzman, R. M. (1993). The scientific and policy needs for improved health forecasting models for elderly populations. In *Forecasting the Health of Elderly Populations*. New York: Springer-Verlag.
- Mazzuco, S., Scarpa, B., and Zanotto, L. (2018). A mortality model based on a mixture distribution function. *Population Studies*, 72(2):191–200.

- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- McKeown, T. (1976). *The modern rise of population*. London: Edward Arnold.
- McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8:413–432.
- Meeker, W. Q. and Escobar, L. A. (2014). *Statistical methods for reliability data*. John Wiley & Sons.
- Meslé, F. and Vallin, J. (2011). Historical trends in mortality. In Rogers, G. R. and Crimmins, M. E., editors, *International Handbook of Adult Mortality*, pages 9–47. Springer.
- Michaelson, A. and Mulholland, J. (2014). Strategy for increasing the global capacity for longevity risk transfer: developing transactions that attract capital markets investors. *Journal of Alternative Investments*, 17(1):18–27.
- Missov, T. I. and Lenart, A. (2013). Gompertz–Makeham life expectancies: Expressions and applications. *Theoretical Population Biology*, 90:29–35.
- Missov, T. I., Lenart, A., Nemeth, L., Canudas-Romo, V., and Vaupel, J. W. (2015). The Gompertz force of mortality in terms of the modal age at death. *Demographic Research*, 32:1031–1048.
- Missov, T. I. and Vaupel, J. W. (2015). Mortality implications of mortality plateaus. *SIAM Review*, 57(1):61–70.
- Mukhopadhyay, N. (2000). *Probability and statistical inference*. CRC Press.
- Mullen, K., Ardia, D., Gil, D., Windover, D., and Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26.
- Myers, G. C. and Manton, K. G. (1984). Compression of Mortality: Myth or Reality? *The Gerontologist*, 24(4):346–353.
- Nepomuceno, M. R. and Canudas-Romo, V. (2019). A Cohort Survival Comparison of Central and Eastern European and High-Longevity Countries. *Population (English edition)*, 74(3):281–300.
- Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. In *Compositional Data Analysis Conference*.
- Oeppen, J. and Camarda, C. G. (2013). Coherent forecasting of multiple-decrement life tables: compositional models for French Cause of Death data, 1925–2008. In *Joint Eurostat/UNECE Work Session on Demographic Projections*.
- Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570):1029–1031.
- Ouellette, N. and Bourbeau, R. (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research*, 25:595–628.

- Ouellette, N., Bourbeau, R., and Camarda, C. G. (2012). Regional disparities in Canadian adult and old-age mortality: A comparative study based on smoothed mortality ratio surfaces and age at death distributions. *Canadian Studies in Population*, 39(3-4):79–106.
- Pascariu, M. D. (2018). *MortalityLaws: Parametric Mortality Models, Life Tables and HMD*. R package version 1.7.0.
- Pascariu, M. D., Canudas-Romo, V., and Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics*, 78:339–350.
- Pascariu, M. D., Lenart, A., and Canudas-Romo, V. (2019). The maximum entropy mortality model: forecasting mortality using statistical moments. *Scandinavian Actuarial Journal*, pages 1–25.
- Pawlowsky-Glahn, V. and Buccianti, A., editors (2011). *Compositional Data Analysis: Theory and Applications*. Wiley.
- Pearson, K. (1897). *The Chances of Death, and Other Studies in Evolution*, volume 1. E. Arnold.
- Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63(1):12–57.
- Pfaff, B. (2008a). *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.
- Pfaff, B. (2008b). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27(4):1–32.
- Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.
- Pletcher, S. D. (1999). Model fitting and hypothesis testing for age-specific mortality data. *Journal of Evolutionary Biology*, 12(3):430–439.
- Pollard, J. H. (1987). Projection of age-specific mortality rates. *Population Bulletin of the United Nations*, 21-22:55–69.
- Preston, S. H. (1976). *Mortality Patterns in National Populations. With special reference to recorded causes of death*. Academic Press.
- Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.
- R Development Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raftery, A. E., Alkema, L., and Gerland, P. (2014). Bayesian Population Projections for the United Nations. *Statistical Science*, 29(1):58–68.
- Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801.

- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, 2nd edition.
- Remund, A. (2015). *Jeunesses vulnérables? Mesures, composantes et causes de la surmortalité des jeunes adultes*. PhD thesis, University of Geneva.
- Remund, A., Camarda, C. G., and Riffe, T. (2018). A Cause-of-Death Decomposition of Young Adult Excess Mortality. *Demography*, 55(3):957–978.
- Renshaw, A. E. and Haberman, S. (2003a). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.
- Renshaw, A. E. and Haberman, S. (2003b). Lee–Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *Applied Statistics*, 52:119–137.
- Renshaw, A. E. and Haberman, S. (2003c). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, 32(3):379–401.
- Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.
- Renshaw, A. E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance: Mathematics and Economics*, 42(2):797–816.
- Riley, J. C. (2001). *Rising life expectancy: a global history*. Cambridge University Press.
- Rizzi, S., Kjærgaard, S., Bergeron-Boucher, M.-P., Lindahl-Jacobsen, R., and Vaupel, J. (2019). Forecasting mortality of not extinct cohorts. In *21st Nordic Demographic Symposium, Reykjavík (Iceland)*.
- Rogers, A. and Little, J. S. (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedule. *Mathematical Population Studies*, 4(3):175–195.
- Rothenberg, R., Lentzner, H. R., and Parker, R. A. (1991). Population aging patterns: the expansion of mortality. *Journal of Gerontology*, 46(2):S66–S70.
- Sasson, I. (2016). Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010. *Demography*, 53(2):269–293.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- Seligman, B., Greenberg, G., and Tuljapurkar, S. (2016). Convergence in male and female life expectancy: Direction, age pattern, and causes. *Demographic Research*, 34:1063–1074.
- Shang, H. L., Booth, H., and Hyndman, R. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, 25:173–214.

- Shkolnikov, V. M., Andreev, E. E., and Begun, A. Z. (2003). Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. *Demographic Research*, 8:305–358.
- Shkolnikov, V. M., Andreev, E. M., Zhang, Z., Oeppen, J., and Vaupel, J. W. (2011a). Losses of expected lifetime in the united states and other developed countries: methods and empirical analyses. *Demography*, 48(1):211–239.
- Shkolnikov, V. M., Jdanov, D. A., Andreev, E. M., and Vaupel, J. W. (2011b). Steep Increase in Best-Practice Cohort Life Expectancy. *Population and Development Review*, 37(3):419–434.
- Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60(4):750–757.
- Siler, W. (1983). Parameters of Mortality in Human Populations with Widely Varying Life Spans. *Statistics in Medicine*, 2:373–380.
- Smith, D. and Keyfitz, N. (1977). *Mathematical Demography. Selected Papers*. Berlin: Springer-Verlag.
- Smits, J. and Monden, C. (2009). Length of life inequality around the globe. *Social Science & Medicine*, 68(6):1114–1123.
- Stoeldraijer, L., van Duin, C., van Wissen, L., and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research*, 29:323–354.
- Sundhedsministeriet (1994). Levetiden i Danmark [Life Expectancy in Denmark] (Sundhedsministeriet, Copenhagen). Danish.
- Tabeau, E. (2001). A review of demographic forecasting models for mortality. In *Forecasting Mortality in Developed Countries*, pages 1–32. Springer.
- Terzani, T. (1995). *Un indovino mi disse*. RL libri.
- Thatcher, A. R., Cheung, S. L. K., Horiuchi, S., and Robine, J.-M. (2010). The compression of deaths above the mode. *Demographic Research*, 22:505–538.
- Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). *The Force of Mortality at Ages 80 to 120*. Monographs on Population Aging, 5. Odense University Press, Odense.
- Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the Danish Life Insurance Company of 1871. *Journal of the Institute of Actuaries and Assurance Magazine*, 16(5):313–329.
- Thompson, R. and Baker, R. (1981). Composite Link Functions in Generalized Linear Models. *Applied Statistics*, 30(2):125–131.
- Torri, T. and Vaupel, J. W. (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting*, 28(2):519–531.

- United Nations (2015). *World Population Ageing 2015 (ST/ESA/SER.A/390)*. Department of Economic and Social Affairs, Population Division, New York, NY.
- United Nations (2019). *World Population Prospects 2019: Highlights. ST/ESA/SER.A/423*. Department of Economic and Social Affairs, Population Division, New York, NY.
- Vallin, J. and Meslé, F. (2004). Convergences and divergences in mortality: A new approach of health transition. *Demographic Research*, 2:11–44.
- Vallin, J. and Meslé, F. (2016). Highest Life Expectancies: Which Leader After Japan? In *Annual Meeting of the Population Association of America, Washington D.C.*
- van Raalte, A. A. and Caswell, H. (2013). Perturbation analysis of indices of lifespan variability. *Demography*, 50(5):1615–1640.
- van Raalte, A. A., Nepomuceno, M. R., Camarda, C. G., and Myrskylä, M. (2018a). How is lifespan variation developing across cohorts? In *Tel Aviv Workshop on Inequality and Uncertainty in Length of Life*.
- van Raalte, A. A., Sasson, I., and Martikainen, P. (2018b). The case for monitoring life-span inequality. *Science*, 362(6418):1002–1004.
- Vaupel, J. and Lundström, H. (1994). Longer life expectancy? Evidence from Sweden of reductions in mortality rates at advanced ages. In *Studies in the Economics of Aging*, pages 79–102. University of Chicago Press.
- Vaupel, J. W. (1997). Trajectories of Mortality at Advanced Ages. In Wachter, K. W. and Finch, C. E., editors, *Between Zeus and the Salmon: The biodemography of longevity*, pages 17–37. National Academy Press, Washington, DC.
- Vaupel, J. W. and Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography*, 40(2):201–216.
- Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khzaaeli, A. A., Liedo, P., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365):855–860.
- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.
- Vaupel, J. W., Zhang, Z., and van Raalte, A. A. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ open*, 1(1):e000128.
- Villegas, A. M., Millossovich, P., and Kaishev, V. K. (2017). *StMoMo: An R Package for Stochastic Mortality Modelling*. R package version 0.4.0.
- Ševčíková, H., Li, N., Kantorova, V., Gerland, P., and Raftery, A. E. (2016). Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections. In Schoen, R., editor, *The Springer series on demographic methods and population analysis: Vol. 39. Dynamic demographic analysis*, pages 285–310. Springer.

- Wei, L.-J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11(14-15):1871–1879.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 103:293–297.
- Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, pages 253–270.
- Whelpton, P. K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association*, 31(195):457–473.
- Willekens, F. (2001). Gompertz in Context: The Gompertz and Related Distributions. In Tabeau, E., van den Berg Jeths, A., and Heathcote, C., editors, *Forecasting Mortality in Developed Countries*, volume 9 of *European Studies of Population*, pages 105–126. Springer Netherlands.
- Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., C., B., Bubenheim, M., Philipov, D., Shkolnikov, V., and Vachon, P. (2007). Methods Protocol for the Human Mortality Database. Last Revised: May 31, 2007 (Version 5).
- Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., Riffe, T., Boe, C., Bubenheim, M., Philipov, D., Shkolnikov, V., Vachon, P., Winat, C., and Barbieri, M. (2019). Methods Protocol for the Human Mortality Database. Last Revised: October 5, 2019 (Version 6).
- Wilmoth, J. R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report, Technical report, Department of Demography, University of California, Berkeley.
- Wilmoth, J. R. (1997). In Search of Limits. In Wachter, K. W. and Finch, C. E., editors, *Between Zeus and the Salmon: The biodemography of longevity*, pages 38–64. National Academy Press, Washington, DC.
- Wilmoth, J. R., Deegan, L. J., Lundström, H., and Horiuchi, S. (2000). Increase of maximum life-span in Sweden, 1861-1999. *Science*, 289(5488):2366–2368.
- Wilmoth, J. R. and Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography*, 36(4):475–495.
- Wilmoth, J. R. and Lundström, H. (1996). Extreme longevity in five countries. *European Journal of Population/Revue Européenne de Démographie*, 12(1):63–93.
- Wilson, C. (2001). On the Scale of Global Demographic Convergence 1950–2000. *Population and Development Review*, 27(1):155–171.
- Zanotto, L. (2017). *A mixture model to distinguish mortality components*. PhD thesis, Università degli Studi di Padova - Dipartimento di Scienze Statistiche.
- Zanotto, L. and Mazzuco, S. (2017). Reconstruction of cohort data via EM algorithm. In *XXVIII International Population Conference*.