University of the Aegean

Doctoral Thesis

A new approach to dynamic stochastic mortality modelling for managing mortality-longevity risk

Author

Aliki E. Sagianou

Supervisor

Associate Prof. Peter Hatzopoulos

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

at the

Department of Statistics and Actuarial-Financial Mathematics

University of the Aegean

Samos, November 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μια νέα προσέγγιση στη δυναμική στοχαστική μοντελοποίηση θνησιμότητας για τη διαχείριση του κινδύνου θνησιμότητας-μακροζωίας

Συγγραφέας
Αλίκη Ε. Σαγιάνου

Επιβλέπων
Αναπλ. Καθ. Πέτρος Χατζόπουλος

ΔΙΑΤΡΙΒΗ
για την απόκτηση Διδακτορικού Διπλώματος
στο

Τμήμα Στατιστικής και Αναλογιστικών-Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου

Σάμος, Νοέμβριος 2022

Υπεύθυνη Δήλωση

Εγώ η Αλίκη Ε. Σαγιάνου, δηλώνω ότι είμαι η αποκλειστική συγγραφέας της υποβληθείσας Διδακτορικής Διατριβής με τίτλο «Μια νέα προσέγγιση στη δυναμική στοχαστική μοντελοποίηση θνησιμότητας για τη διαχείριση του κινδύνου θνησιμότητας-μακροζωίας». Η συγκεκριμένη Διδακτορική Διατριβή είναι πρωτότυπη και εκπονήθηκε αποκλειστικά για την απόκτηση του Διδακτορικού Διπλώματος του Τμήματος Στατιστικής και Αναλογιστικών-Χρηματοοικονομικών Μαθηματικών. Κάθε βοήθεια, την οποία είχα για την προετοιμασία της, αναγνωρίζεται πλήρως και αναφέρεται επακριβώς στην εργασία.

Επίσης, επακριβώς αναφέρω στην εργασία τις πηγές, τις οποίες χρησιμοποίησα, και μνημονεύω επώνυμα τα δεδομένα ή τις ιδέες που αποτελούν προϊόν πνευματικής ιδιοκτησίας άλλων, ακόμη κι εάν η συμπερίληψη τους στην παρούσα εργασία υπήρξε έμμεση ή παραφρασμένη. Γενικότερα, βεβαιώνω ότι κατά την εκπόνηση της Διδακτορικής Διατριβής έχω τηρήσει απαρέγκλιτα όσα ο νόμος ορίζει περί διανοητικής ιδιοκτησίας και έχω συμμορφωθεί πλήρως με τα προβλεπόμενα στο νόμο περί προστασίας προσωπικών δεδομένων και τις αρχές της Ακαδημαϊκής Δεοντολογίας.

Υπογραφή:

Ημερομηνία: Νοέμβριος 28, 2022

Declaration of Authorship

I, Aliki E. Sagianou, declare that this thesis entitled, "A new approach to dynamic stochastic mortality modelling for managing mortality-longevity risk" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: November 28, 2022

Advising Committee of this Doctoral Thesis:

Professor Alexandros Karagrigoriou, Advisor
Department of Statistics and Actuarial-Financial Mathematics
University of the Aegean, Greece

Associate Professor Efstathios Chadjiconstantinidis, Advisor
Department of Statistics and Insurance Science
University of Piraeus, Greece

Associate Professor Peter Hatzopoulos, Supervisor
Department of Statistics and Actuarial-Financial Mathematics
University of the Aegean, Greece

University of the Aegean, Greece
2022

Approved by the Examining Committee:

Alexandros Karagrigoriou
Professor, University of the Aegean, Greece

Anastasia Kostaki
Professor, Athens University of Economics and Business, Greece

---

Efstathios Chadjiconstantinidis
Associate Professor, University of Piraeus, Greece

Peter Hatzopoulos
Associate Professor, University of the Aegean, Greece

Stelios Xanthopoulos
Associate Professor, University of the Aegean, Greece

Stelios Zimeras
Associate Professor, University of the Aegean, Greece

Georgios Pitselis
Assistant Professor, University of Piraeus, Greece

University of the Aegean, Greece
2022

Aliki E. Sagianou

Department of Statistics and Actuarial-Financial Mathematics
University of the Aegean

This thesis contains original work published or submitted for publication to international journals. The following papers are included:

1. Peter Hatzopoulos and Aliki Sagianou (2020) Introducing and Evaluating a New Multiple-Component Stochastic Mortality Model, North American Actuarial Journal, 24:3, 393-445, DOI: 10.1080/10920277.2019.1658606

2. Aliki Sagianou and Peter Hatzopoulos (2022) Extensions on the Hatzopoulos–Sagianou Multiple-Components Stochastic Mortality Model, Risks, 10(7):131, DOI:10.3390/risks10070131

3. Peter Hatzopoulos and Aliki Sagianou. HSTool: A Matlab toolbox for the Hatzopoulos–Sagianou Multiple-Component Stochastic Mortality Model. [**Working paper - To be submitted**]

This thesis has been funded by the "YPATIA" Scholarship Program for Doctoral Studies, of the University of the Aegean.

Abstract

Department of Statistics and Actuarial-Financial Mathematics
University of the Aegean

Doctor of Philosophy
by Aliki E. Sagianou

The continuous increase in life expectancy observed in recent years, especially in developed countries, as well as the actuarial risks raised due to the ageing population have triggered scientists and actuaries to develop appropriate models in order to model and project mortality rates. A great enabler to achieve this is the stochastic mortality models which serve as a steppingstone for the development of plans for managing the mortality risk for the sake of safeguarding the solvency of insurance organisations.

In this context, this doctoral thesis aims to further explore the domain of mortality modelling and identify the methodologies empowering the existing models, highlight their advantages and disadvantages and use them to draw a research line for further contributing to the domain.

Based on this analysis, this thesis introduces and evaluates a new multiple-component stochastic mortality model, namely the Hatzopoulos-Sagianou (HS) model, which goes beyond the prominent solutions, aiming to address the identified limitations. HS offers a dynamic model structure and is based on a parameter estimation methodology, which aims to reveal significant and distinct age clusters by identifying the optimal number of incorporated period and cohort effects. The latter maximise the captured variance of the mortality data, enable the attribution of an identified mortality trend to a unique age cluster and contribute to the interpretability of the modelling results.

Having HS as the basis, this thesis further extends the model through the use of various link functions, and diverse distributions in the model's estimation method and differentiates the HS approach by modelling the number of deaths using the Binomial model. In addition, the HS is reformed by introducing a new form of link functions with a particular focus on the use of heavy-tailed distributions leading to an enhanced estimation methodology, improving even more the model's *goodness-of-fit,* and producing more fine-grained age clusters.

The above-mentioned developments lead to a software implementation of the HS in the form of a Matlab toolbox called: *HS-Tool*. The latter provides the means for replicating the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projections. The *HS-Tool* introduces codebase-related improvements that contribute to the model's ability to provide a "plug 'n' play" user experience, while the integration of multi-criteria decision making methods in the model's workflow achieve increased tool automation and convergence to "optimal" values for critical model parameters.

The model's qualities are supported by experimental results showing that the HS model is able to achieve high scores over diverse qualitative and quantitative evaluation metrics, while the identified mortality rates come into agreement with well–established findings of the mortality literature.

Greek Abstract

(Εκτεταμένη Περίληψη)

Τμήμα Στατιστικής και Αναλογιστικών-Χρηματοοικονομικών Μαθηματικών
Σχολή Θετικών Επιστημών
Πανεπιστήμιο Αιγαίου

Διδακτορική διατριβή
της Αλίκης Ε. Σαγιάνου

Η συνεχής αύξηση του προσδόκιμου ζωής που παρατηρείται τα τελευταία χρόνια, ιδιαίτερα στις ανεπτυγμένες χώρες, καθώς και οι αναλογιστικοί κίνδυνοι που εγείρονται λόγω της γήρανσης του πληθυσμού έχουν ωθήσει την επιστημονική κοινότητα και τους αναλογιστές στην ανάπτυξη κατάλληλων μοντέλων προκειμένου να αναλύσουν και να προβλέψουν τους ρυθμούς θνησιμότητας των πληθυσμών. Ένα σημαντικό εργαλείο για να επιτευχθεί αυτό είναι τα στοχαστικά μοντέλα θνησιμότητας που αξιοποιούνται στην ανάπτυξη ολοκληρωμένων σχεδίων για τη διαχείριση του κινδύνου θνησιμότητας με απώτερο στόχο τη διασφάλιση της φερεγγυότητας των ασφαλιστικών οργανισμών.

Στο πλαίσιο αυτό, η παρούσα διδακτορική διατριβή στοχεύει να διερευνήσει περαιτέρω τον τομέα της μοντελοποίησης θνησιμότητας και να μελετήσει τις μεθοδολογίες που χρησιμοποιούνται από τα υπάρχοντα μοντέλα, να καταγράψει τα πλεονεκτήματα και τα μειονεκτήματά τους και να τα αξιοποιήσει για να χαράξει μια ερευνητική γραμμή που θα συνεισφέρει στον τομέα.

Με βάση αυτή την ανάλυση, η παρούσα διατριβή εισάγει και αξιολογεί ένα νέο στοχαστικό μοντέλο θνησιμότητας πολλαπλών παραγόντων ηλικίας-περιόδου-γενεάς, με την ονομασία Χατζόπουλος-Σαγιάνου (HS), το οποίο ξεπερνά τις υπάρχουσες λύσεις και στοχεύει στην αντιμετώπιση των αδυναμιών που εντοπίστηκαν στην βιβλιογραφία. Το HS προσφέρει μια δυναμική δομή και βασίζεται σε μια μεθοδολογία εκτίμησης παραμέτρων που στοχεύει στην αποκάλυψη παραγόντων που αντιστοιχούν σε σημαντικούς (από άποψη πληροφορίας) ρυθμούς θνησιμότητας διακριτών ηλικιακών ομάδων, προσδιορίζοντας τον βέλτιστο αριθμό παραγόντων ηλικίας-περιόδου-γενεάς. Σκοπός της μεθόδου είναι να εξάγει τη μέγιστη δυνατή πληροφορία από τα δεδομένα θνησιμότητας, να αντιστοιχίσει μία μοναδική ηλικιακή ομάδα σε μία ξεκάθαρη τάση θνησιμότητας και να καταστήσει δυνατή την εύκολη ερμηνεία των αποτελεσμάτων της μοντελοποίησης.

Έχοντας ως βάση το HS, αυτή η διατριβή επεκτείνει περαιτέρω το μοντέλο με τη χρήση διάφορων συναρτήσεων σύνδεσης (link functions) και διαφορετικών κατανομών στη μέθοδο εκτίμησης του μοντέλου και διαφοροποιεί το HS μοντελοποιώντας τον αριθμό των θανάτων χρησιμοποιώντας το Διωνυμικό μοντέλο.

Επιπλέον, το HS αναδιαμορφώνεται με την εισαγωγή μιας νέας μορφής συναρτήσεων σύνδεσης με ιδιαίτερη έμφαση στη χρήση κατανομών βαριάς ουράς που, βελτιώνοντας ακόμη περισσότερο την απόδοση του μοντέλου. Τα παραπάνω οδηγούν στη δημιουργία μιας εργαλειοθήκης Matlab με το όνομα: HSTool. Το τελευταίο δίνει τη δυνατότητα της αναπαραγωγής όλων των απαραίτητων διαδικασιών ενός μοντέλου θνησιμότητας, δηλαδή την προσαρμογή του μοντέλου στα δεδομένα θνησιμότητας, την αξιολόγηση της απόδοσης του και την πρόβλεψη ρυθμών θνησιμότητας. Το HSTool εισάγει βελτιώσεις που συμβάλλουν στην ευχρηστία του μοντέλου, ενώ ενσωματώνει μεθόδους λήψης αποφάσεων πολλαπλών κριτηρίων που αυξάνουν το επίπεδο αυτοματισμού του εργαλείου και βοηθού στον προσδιορισμό των «βέλτιστων» τιμών κρίσιμων παραμέτρων.

Τα πλεονεκτήματα του HS υποστηρίζονται από πειραματικά αποτελέσματα που δείχνουν την αποτελεσματικότητα του βάσει ποσοτικών και ποιοτικών κριτηρίων αξιολόγησης, ενώ οι ρυθμοί θνησιμότητάς συμφωνούν με γνωστά ευρήματα της σχετικής βιβλιογραφίας.

Acknowledgements

I would like to thank my supervisor Peter Hatzopoulos for his guidance, the continuous support and his patience during this long journey.

I would also like to thank the other members of my advising committee Prof. Alexandros Karagrigoriou and Assoc. Prof. Efstathios Chadjiconstantinidis.

In addition, I would like to thank Prof. Anastasia Kostaki, Assoc. Prof. Stelios Xanthopoulos, Assoc. Prof. Stelios Zimeras, and Assist. Prof. Georgios Pitselis, members of the examining committee, for investing some of their time in reviewing this thesis and offering advice.

I am deeply grateful to my colleagues and friends, Christos Merkatas and Kostas Kaloudis, for sharing constructive ideas and moments of joy. I hope they will fulfil their goals and have a fortunate life.

I would also like to acknowledge my friends for their emotional support, particularly Alexandra P., Anna K., Annita S., Argiro G., Artemis, Eleni V., Giorgos L, Ilias P., Kostas T., Maria P. Michalis S., Niki K., Panagiotis K., Stathis K., and Villy K..

A special thanks goes to Dimitris Papamartzivanos for being always next to me during this long journey. His love and support gave me the strength to keep working hard for my goals. Without his valuable guidance, this work would not have been materialised. I would like also to thank Christos, Dina and Athanasia for their encouragement and support.

Nothing would have been possible without the emotional and moral support of my family. Words alone cannot express the thanks I owe to Vaggelis, Stella, Panos. Their belief in me, their encouragement and unconditional love gave me strength throughout my research.

Dedicated to my family...

Contents

List of Figures

List of Tables

Abbreviations

AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
APC	Age–Period–Cohort
BD–test	Bonferroni–Dunn test
BIC	Bayesian Information Criterion
CBD	Cairns–Blake–Dowd
CDF	Cumulative Distribution Function
cll	Complementary Log–Log
DLR	Dynamic Linear Regression
DM	Decision Matrix
E&W	England and Wales
ER	Explanation Ratio
F–test	Friedman's statistic test
GLM(s)	Generalised Linear Model(s)
HH	Hatzopoulos–Haberman
HS	Hatzopoulos–Sagianou
LC	Lee–Carter
lgt	logit
MAPE	Mean Absolute Percentage Error
MCDM	Multi-criteria decision making
MSPE	Mean Squared Percentage Error
npar	Number of parameters
PC	Principal Component
PCA	Principal Component Analysis
PE	Percentage Error

PL	Plat
prbt	Probit
prt	Generalised Pareto
RH	Renshaw–Haberman
RWD	Random Walk with Drift
SAW	Simple Additive Weighting
SCR	Solvency Capital Requirement
SPCA	Sparse Principal Component Analysis
SPCs	Sparse Principal Components
SVD	singular Value Decomposition
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
UV	Unexplained Variance
UVR	Unexplained Variance Ratio
WLC	Weighted Linear Combination
WSM	Weighted Sum Model

Chapter 1

Introduction

The continuous increase in life expectancy observed in recent years, especially in developed countries, as well as the actuarial risks raised due to the ageing population, have triggered scientists and actuaries to develop appropriate models in order to model and project mortality rates. Although the increase in life expectancy is a bright sign of social, medical and scientific progress, it poses challenges to governments, private pension plans and life insurers, as it significantly affects the cost of pensions. To address those challenges the insurance industry needs to adopt policies and tools that can counterbalance the negative effects of the so-called longevity risk. In this direction, major effort has been put by the insurance industry and academia to analyse, model and control the inherent market risks in order to provide the means to safeguard the solvency of insurance organisations. In this regard, the need for managing the longevity risk reinforces the investigation for novel techniques in order to understand and model the mortality dynamics.

Undoubtedly, a great tool to achieve this is the stochastic mortality models. Regardless of the peculiarities of each implementation, mortality models aim to analyse mortality by decomposing the mortality rates in the dimensions of age, period and cohort (or year of birth). The aforementioned dimensions enable the analysis of mortality rates of individuals as they age, aid to the understanding of medical and social progress, and act as enablers to shed light on the lifelong mortality effects. In this direction, mortality models can be used to support the analysis, modelling, and management of the inherent market risks for the sake of safeguarding the solvency of organisations.

Towards this direction, Solvency II was introduced, and came into effect on 1 January 2016, to tackle the lack of risk sensitivity of Solvency I which could lead to inaccurate assessment of risks and sub-optimal allocation of capital to deal with them (Rae et al. (2018)). That is, Solvency II instructs the so-called Solvency Capital Requirement (SCR) to be risk-based, and not just a fixed percentage of the mathematical reserve or the risk capital. However, in order to converge to a reliable estimation for the SCR, one needs to use as a ground truth a credible set of evidence.

The increasing use of complex quantitative models in applications throughout the financial world, has rendered the "Model Risk" a major concern. Such risk is generated by the potential inaccuracy and inappropriate use of models in business applications. The same applies to the insurance sector when it comes to the mortality models. Since the latter are a key enabler in the process of managing the longevity risk, it becomes clear that it is of paramount importance to ensure that the modelling process is performed with high fidelity, it produces accurate estimations, enables the results interpretation, and remains consistent with historical data.

Having identified several gaps in the mortality modelling literature, this doctoral thesis seeks new methods that can address the reported limitations and provide a new robust multiple-component mortality model with agile characteristics. Under this prism, the next section details the motivation and objectives of this thesis, while Section 1.2 highlights the contribution of this work.

1.1 Motivation and Objectives

Mortality modelling is a cardinal process in the context of insurance companies, as the modelling results and estimations are used as the basis for several other estimations, processes, and business operations such as pricing of life-related insurance products, pension products, social security and medical insurance management, longevity risk management and estimation of SCR.

Thus, since mortality models are a link of this operational chain, we need to guarantee their operational stability and accuracy when it comes to the identification of the mortality rates of studied populations. In fact, this is a requirement that falls under

the "Model Risk" notion. According to the Federal Reserve (2011) "*The use of models invariably presents model risk, which is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports. Model risk can lead to financial loss, poor business and strategic decision making, or damage to a bank's reputation*". Such a risk may occur as a result of potential fundamental errors in the model or by producing inaccurate outputs (Aggarwal et al. (2016)).

In this direction, it is of paramount importance to further investigate the mortality models of the literature in order to identify potential limitations that could lead to degraded performance, and can potentially trigger cascading effects to other depended business processes. To this end, a research question is posed on whether, and to which extend, the existing mortality models can accurately model the mortality rates considering the different mortality behaviour appearing among different age ranges on different populations. In fact, as it will be explained in detail in Chapter 2, the existing mortality models present a rather rigid structure and consider only 1 to 3 components (age-period and age-cohort components), and sometimes even with pre-specified age effect terms. Hence, the ability of such models to capture the necessary amount of information of the mortality data, and, in turn, to model the mortality rate of a population with high accuracy, warrants further investigation.

The limited number of components contributes to the parsimony of a model, to its simplicity, even its usability from the analyst's standpoint. In fact, those characteristics are claimed to be desirable for a mortality model according to Cairns et al. (2008), Plat (2009). Such an approach is good enough if one needs to perform a quick and superficial analysis. However, one could wonder whether there is a trade-off between a model's parsimony and its ability to accurately capture the dynamics of mortality in a given population. Considering that one dimension of model risk is a model's ability to perform accurately, it becomes clear that, in order to perform an in-depth analysis of mortality, we need to investigate the dynamic model structures that can adapt to the peculiarities of the mortality data, rather than expecting rigid models to perform adequately under different analysis testbeds.

In addition, the existing models come with limited "interpretability" when it comes to the extracted mortality trends. In other words, the identified components are not clear, and the extracted mortality trend is attributed to the whole age range or a mix of age

ranges of a population. This behaviour cannot ease the in-depth analysis of mortality on specific age clusters, as legacy models offer a generic representation of mortality dynamics without being evident which ages are reflected in an identified mortality trend. In this direction, there is a need to attribute a specific mortality trend to a unique age cluster in order to perform a fine-grained analysis and enable informed decision making for post-modelling operations.

Given the above-mentioned observations, the motivation of this PhD thesis is to investigate novel methodologies that tackle the limitations of existing models, can provide agile characteristics to mortality models and enable the in-depth analysis of mortality trends by increasing the interpretability of the generated results. In a nutshell, the objectives (and simultaneously the research pillars) of this PhD thesis are as follows:

Objective 1: We intend to review the mortality modelling-related literature in order to identify the methodologies empowering the existing mortality models and highlight their advantages and disadvantages. Through this analysis we intend to identify open research challenges than need to be addressed.

Objective 2: Propose a new mortality model which will go beyond the prominent solutions of the domain, aiming to address the identified limitations. The qualities of the model shall be advocated by comparative experimental results.

Objective 3: Investigate methods that can contribute to the mortality modelling domain and further improve models and their operational reliability in terms of model fitting and mortality forecasting.

Objective 4: Provide technical artefact to the community to support the insurance industry to its mission-critical operations towards achieving solvency and foster research in Academia.

As detailed in the next subsection, the novelties of this work mainly lie in the last three objectives, as the first one basically explores the related literature for identifying possible gaps, shortcomings, and research directions.

1.2 Contributions

As already pointed out, the main intent of this PhD work is to investigate new methodologies that will provide the missing flexibility to the mortality modelling in order to escape from the static modelling behaviour and to provide enhanced interpretability of the modelling results.

To this end, our work exploits the beneficial characteristics of Generalised Linear models (GLMs) and Sparse Principal Component Analysis (SPCA) to construct a dynamic multiple-component stochastic mortality model, namely the Hatzopoulos-Sagianou model (HS) and to address some of the limitations identified in the related literature. The modelling process is driven by the Unexplained Variance Ratio (UVR), a heuristic metric, which contributes to the identification of the most informative components in order to maximise the captured variance of the mortality data. In this way, the HS model gains a highly informative structure in an efficient way, and it is able to designate an identified mortality trend to a unique age cluster. The advantageous characteristics of the designed model are advocated by a thorough experimental testbed used to evaluate the efficiency of the proposed model in terms of fitting and forecasting performance over several datasets. The acquired results come into agreement with well-established findings in the mortality literature.

Initially, our work provides a comprehensive analysis of well-established mortality models, and offers an application that brings them into direct comparison with the HS model under diverse qualitative and quantitative evaluation metrics. In addition, we further extend the designed model by formulating it in terms of conditional probabilities of death, $q_{t,x}$, and by adopting various link functions. Given this, new HS extensions are derived using the off-the-shelf link functions, namely the complementary log–log, logit and probit, while we also reform the model by introducing a new form of link functions with a particular focus on the use of heavy-tailed distributions. The above-mentioned research actions, lead to the creation of the HSTool, a Matlab toolbox that offers to the community the implementation of the HS model. The tool extends further the operational capabilities of the HS model by enhancing the estimation process of HS with multi-criteria decision making methods in order to offer increased level of automation and increased usability.

More specifically, the contribution of this PhD thesis with respect to our publications in scientific journals is as follows:

- The introduction and evaluation of a new multiple component stochastic mortality model, namely the Hatzopoulos-Sagianou (HS) model.[1] The main pillars of this contribution are:
  - The design and development of a new stochastic mortality model, which utilises multiple age–period and age–cohort interaction terms based on a disruptive parameter estimation method which uses the Sparse Principal Component Analysis (SPCA) and Generalised Linear Models (GLMs).
  - Introduction of a novel methodology based on the Unexplained Variance Ratio (UVR) metric in order to pinpoint the most important age–period and age–cohort components and define the optimal sparsity factor of SPCA.
  - Extensive evaluation of the new model and comparative analysis of the obtained results against well-known existing mortality models.
  - Design of a backtesting testbed (out-of-sample comparison) in order to evaluate the ability of each model in terms of prediction.
- Extensions on the Hatzopoulos–Sagianou multiple-component stochastic mortality model based on the use of different link functions and probability distributions [2]. In this context, we:
  - We extend the stochastic mortality model HS formulated in terms of $q_{t,x}$, using Generalised Linear Models and by adopting various link functions. We illustrate through experimental results that the HS model remains robust and consistent under all modelling variations.
  - We introduce a new set of link functions, with particular focus on heavy-tailed distributions, and we evaluate their applicability in the context of mortality through the HS extensions. This approach leads to the definition of a new estimation methodology for the HS model.

[1] Peter Hatzopoulos and Aliki Sagianou (2020) Introducing and Evaluating a New Multiple-Component Stochastic Mortality Model, North American Actuarial Journal, 24:3, 393-445, DOI: 10.1080/10920277.2019.1658606

[2] Aliki Sagianou and Peter Hatzopoulos (2022) Extensions on the Hatzopoulos–Sagianou Multiple-Components Stochastic Mortality Model, Risks, 10(7):131, DOI:10.3390/risks10070131

  - We compare the efficiency of the new model extensions versus established mortality models in fitting and forecasting modes.
  - We highlight the lessons learned to inform the community on the adoption of the various link functions in a model's estimation methods having witnessed the beneficial impact of this approach to our model's efficacy.

- A software implementation and technical documentation of the HS model in the form of a Matlab toolbox under the name *HSTool* [3]. The contributions of this work are as follows:
  - The offering of a Matlab toolbox, namely HSTool, of the HS model presented in Hatzopoulos & Sagianou (2020) supported by the necessary material for the replication of mortality analysis testbeds and the use of model's commands.
  - Documentation of new codebase-related improvements that contribute to the model's stability and ability to provide a calibrated operation for a "plug 'n' play" user experience.
  - Introduction of multi-criteria decision making methods in the model's workflow to achieve increased tool automation and convergence to "optimal" values for critical model's parameters.
  - Documentation of the code commands for replicating the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projection.

[3] Peter Hatzopoulos and Aliki Sagianou. HSTool: A Matlab toolbox for the Hatzopoulos–Sagianou Multiple-Component Stochastic Mortality Model. [**Working paper - To be submitted**]

1.3 Thesis Structure

The next chapter presents the fundamental background of mortality modelling. More specifically, Chapter 2 elaborates on the data and notation used in mortality in order to introduce the reader to the domain specifics. In addition, it provides a thorough documentation of well-known mortality models, documents the theoretical background of commonly used quantitative and qualitative metrics, and concludes with a summary of the literature highlights.

Chapter 3 deals with the design, development and evaluation of the new multiple component stochastic mortality model, namely the Hatzopoulos-Sagianou model (HS model). More specifically, this chapter documents the model's disruptive estimation methodology that delivers a dynamically defined model structure, while an extensive comparative analysis against the existing models of the literature is performed.

Chapter 4 presents extensions of the HS model by deploying various link functions, using generalised linear models, and diverse distributions in the model's estimation method. More precisely, the HS approach is differentiated by modelling the number of deaths using the Binomial model, while new HS extensions are derived using off-the-shelf and newly introduced link functions.

Chapter 5 offers a handbook of the software implementation of the HS model in the form of a Matlab toolbox called: *HS-Tool*. The chapter offers a complete technical documentation and illustrates the capabilities of the toolbox through examples and applications to mortality data.

The last chapter provides a discussion over the results and the contributions of this PhD thesis. Additionally, it provides directions for future research.

Chapter 2

Background

This chapter provides an overview of the mortality modelling domain and offers the necessary preliminary information that will help the reader follow the methodologies and technical content presented in the next chapters of this thesis. Initially, the chapter elaborates on the data and notation commonly used in the mortality domain, provides an overview of the existing mortality models, and more details on the well-established models of the literature. This chapter also elaborates on the evaluation metrics used to ratify the *goodness-of-fit* performance and forecasting capabilities of the models. Those metrics will be used for benchmarking the existing models of the literature against the HS model and its extensions. Last but not least, Section 2.2 proceeds to the documentation of the specifics of the Generalised Linear Models, as those are a key ingredient of the HS parameter estimation methodology.

2.1 Data and Notation

The data used in the experimental testbeds of this thesis consist of the random variable of the number of deaths, $D_{t,x}$ where $d_{t,x}$ denotes the observed number of deaths, the central exposures to risk, $E^c_{t,x}$, and the corresponding initial exposures to risk, $E^0_{t,x}$, at age x last birthday during calendar year t. The data take the form of a rectangular structure (t, x) over a unit range of individual calendar years t $(t_1, \ldots, t_n)$ and individual ages, x, last birthday $(x_1, \ldots, x_a)$. That is, the probability that a person at age x and calendar year t will die within one year is defined as $q_{t,x} = d_{t,x}/E^0_{t,x}$. The crude central

mortality rate for any age x and calendar year t is defined as $m_{t,x} = d_{t,x}/E^c_{t,x}$. $E^c_{t,x}$ is often derived based on an approximation of the number of those aged x years old during their last birthday half way through year t, or based on an estimation of the average of said population at the beginning and end of that year. The $E^0_{t,x}$ is defined as the population size on January 1st, at age x in calendar year t. Initial exposures are then approximated by $E^0_{t,x} \approx E^c_{t,x} + \frac{1}{2}d_{t,x}$. We model the number of deaths as independent Poisson or Binomial realisations, i.e., $D_{t,x}$ follows the Poisson distribution with mean $E^c_{t,x} \times m_{t,x}$ (Brillinger (1986), Brouhns et al. (2002)), or the Binomial distribution with mean $E^0_{t,x} \times q_{t,x}$. When the context is clear, we may write $E_{t,x}$ to refer to $E^0_{t,x}$ or $E^c_{t,x}$.

2.2 Generalised Linear Model framework in mortality modelling

This section provides an overview of the Generalised Linear Model framework and how this is applied in the context of mortality modelling. In fact, GLM framework is a key enabler for the estimation methodology of the newly introduced mortality model presented in Chapter 3, as well as for the extensions presented in Chapter 4.

In a generalised linear model (see, e.g., McCullagh & Nelder (1989)), each outcome of a random variable $\mathbf{Y}$, whose components are independently distributed with mean $\boldsymbol{\mu}$, is considered to originate from certain distributions belonging to the exponential family. There is a wide range of probability distributions, such as the Binomial, Poisson, Gamma and Normal distributions. The explanatory variables, $\mathbf{X}$, define the mean, $\boldsymbol{\mu}$, of the distribution via:

$$\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

where $\mathbb{E}(\mathbf{Y})$ is the expected value of $\mathbf{Y}$, where $\mathbf{Y}$ is a vector of response variables. $\mathbf{X}\boldsymbol{\beta}$ is the linear predictor, a linear combination of parameters $\boldsymbol{\beta}$ whose values are usually unknown and have to be estimated from the data. $\mathbf{X}$ is a matrix of explanatory variables, and $\boldsymbol{\beta}$ is a vector of parameters. g is the link function.

The model's independent variables are incorporated through the linear predictor ($\boldsymbol{\eta}$). The link function relates it to the expected value of the data and it is expressed as a linear combination of the unknown parameters $\boldsymbol{\beta}$.

The linear predictor and the mean of the distribution function are related via the link function. A number of considerations need to be taken into account when choosing the most appropriate link function. The exponential of the response's density function determines a well-defined canonical link function. Although, for algorithmic purposes, in some cases, it makes sense to use a non-canonical link function.

Therefore, there are three components to any GLM (McCullagh & Nelder (1989)):

- The *random component*—refers to the probability distribution of the response variable ($\mathbf{Y}$). The components of $\mathbf{Y}$ have generated from an exponential family of probability distributions.
- The *systematic component*—specifies the explanatory variables ($\mathbf{X}$) in the model producing the so-called *linear predictor* $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$.
- The link function, g—specifies the connection between the random and systematic components. Specifically, it denotes how the expected value of the response relates to the linear predictor of explanatory variables, e.g., $\boldsymbol{\eta} = g(\boldsymbol{\mu})$ or $g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu}$.

Hence, as is the case for the generalised linear models, the same applies to the stochastic mortality models. The corresponding three components, plus a forth one on the parameter constraints, in a stochastic mortality model are (Villegas et al. (2018)):

1. The *random component*: the number of deaths $D_{t,x}$ follows the Poisson distribution with mean $E^c_{t,x} \times m_{t,x}$, or the Binomial distribution with mean $E^0_{t,x} \times q_{t,x}$, so that

$$D_{t,x} \sim Poisson(E^c_{t,x} \times m_{t,x})$$

or

$$D_{t,x} \sim Binomial(E^0_{t,x}, q_{t,x})$$

from which $\mathbb{E}(D_{t,x}) = E^c_{t,x} \times m_{t,x} \Rightarrow \mathbb{E}(D_{t,x}/E^c_{t,x}) = m_{t,x}$

or $\mathbb{E}(D_{t,x}) = E^0_{t,x} \times q_{t,x} \Rightarrow \mathbb{E}(D_{t,x}/E^0_{t,x}) = q_{t,x}$.

2. The *systematic component*: the effects of age x, calendar year t and cohort $c = t-x$ are captured through a predictor $\eta_{t,x}$ given by:

$$\eta_{t,x} = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{c(j)} \gamma_c^{(j)} \tag{2.1}$$

   $\eta_{t,x} = g(\mathbb{E}(D_{t,x}/E_{t,x}))$ is the link function used to transform the response variable (which is a measure of mortality rates) at age x and for year t, into a form suitable for modelling, and link it to the proposed predictor structure ((Villegas et al. 2018)). The term α_x represents the main age profile of mortality by age. β_x reflects the age effect for each period component, while β_x^c the age effect for each cohort component. The term κ_t refers to the period-related effects and corresponds to the mortality trend. The term γ_c reflects the cohort-related effects, where $c = t - x$. $p(\geq 1)$ and $q(\geq 0)$ are indexes referring to the period and cohort components included in the model structure.

   In fact, similar to Hunt & Blake (2021) and Villegas et al. (2018), who considered $q = 1$, formula 2.1 reflects the generic structure of an age–period–cohort stochastic mortality model, reflecting the methodology of generalised linear models (McCullagh & Nelder (1989)).

3. The *link function* g associating the random component and the systematic component, so that

$$g\left(\mathbb{E}\left(\frac{D_{t,x}}{E_{t,x}}\right)\right) = \eta_{t,x}$$

   for an appropriate exposure to risk, $E_{t,x}$.

   There are several link functions that can be used as suggested by Haberman & Renshaw (1996), Currie (2016) for the context of mortality models, and McCullagh & Nelder (1989) in the wider context of GLMs. As it will be explained in details in Section 4.2, the extensions of the HS model will be based on the $q_{t,x}$ formulation, using a wide set of canonical and non-canonical link functions.

4. The *set of parameter constraints*: most of the stochastic mortality models have identifiability problems in the parameter estimation. In an effort to avoid this issue, a set of parameter constraints is required to ensure unique parameter estimates. Notably, as will be documented in Chapter 3, the HS model does not need any

parameter constraints as the model does not face identifiability problems during parameter estimation and always provides a unique solution due to its estimation methodology (Hatzopoulos & Sagianou (2020)).

2.3 Stochastic Mortality models

In this section, we introduce the reader into the context of mortality models. Specifically, we outline some background information about the related models that take place in the comparative evaluations in Chapters 3 and 4 of this thesis.

A number of stochastic models have been developed to analyse the mortality improvements. Lee & Carter (1992) proposed a stochastic mortality model which is able to fit to the historical data using only one age-period component without any cohort effect. Even though, the Lee Carter (LC) model is considered a stepping stone in the mortality literature, as mentioned by many authors, LC model bears several weaknesses and limitations (Booth et al. (2006)). Numerous variations of the LC model have been introduced in an effort to deal with the reported limitations. Such a model variation is proposed by Renshaw & Haberman (2006) which, apart from time components, incorporates a cohort effect, resulting into an age–period–cohort version of the LC model. In this way, the Renshaw–Haberman (RH) model provides an improved fit to the historical data compared to the basic LC model. Similarly, Yue et al. (2008) proposed an age–shift model, that includes two second–order interaction terms (age–period and age–cohort). Yue et al. (2008) suggested that the reduction shift of ages for different time periods can be treated as a "cohort" effect, introducing an age–period–cohort model. Also, Hyndman & Shahid Ullah (2007) used several Principal Components (PCs) to capture the differential movements for specific–age mortality rates. Hyndman & Shahid Ullah (2007) smoothed the observed log–mortality rates, using constrained and weighted penalised regression splines, and decomposed the fitted curves using functional principal component analysis. Additionally, a simplified version of the RH model, is the Age-Period-Cohort (APC) model (Currie (2006)). In this model, the age, period and cohort effects influence mortality rates independently. APC model has a long–standing tradition in the fields of medicine and demography (Hobcraft et al. (1985), Clayton & Schifflers (1987)), but it had not been widely used in the actuarial literature until it was considered by Currie (2006). Other important stochastic mortality models, which are usually used in the

literature are the two–factor Cairns–Blake–Dowd (CBD) model introduced by Cairns et al. (2006*b*) and its extensions (Cairns et al. (2009)). These models are known in the literature for their ability to perform better for the higher ages. Moreover, Plat (2009) (PL) introduced a model which combines the advantageous features of the CBD and the LC models to produce a new one that operates under full age ranges and captures the cohort effect. Furthermore, Hatzopoulos & Haberman (2011) (HH) proposed a multiple component model as an extended version of the Hatzopoulos & Haberman (2009) by introducing an alternative estimation method for modelling the (age–period–cohort) mortality effects via Sparse Principal Component Analysis (SPCA).

In the following paragraphs we focus on the LC (Lee & Carter (1992)), RH (Renshaw & Haberman (2006)), APC (Currie (2006)) and PL (Plat (2009)) and CBD (Cairns et al. (2006*b*)) –and its extensions (Cairns et al. (2009))– mortality models, as those are the most well-known models and are part of the most known mortality modelling software packages, like Lifemetrics toolkit (Coughlan et al. (2007)) and StMoMo (Villegas et al. (2018)).

2.3.1 Lee - Carter (LC) model

In a seminal work, Lee & Carter (1992) proposed the following model for central mortality rates:

$$\log(\tilde{m}_{t,x}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} \tag{2.2}$$

i.e., $p = 1$ and $q = 0$, according to the generic formula (2.1).

However, there is an identifiability problem in parameter estimation and in order to ensure this property, the following set of parameter constraints are defined:

$$\sum_x \beta_x^{(1)} = 1 \qquad \text{and} \qquad \sum_t \kappa_t^{(1)} = 0$$

2.3.2 Renshaw - Haberman (RH) model

Renshaw & Haberman (2006) generalised the LC model and proposed a model which incorporates a cohort effect and it has the following form:

$$\log(\tilde{m}_{t,x}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{c\,(1)}\gamma_c^{(1)} \tag{2.3}$$

i.e., $p = q = 1$, according to the generic formula (2.1).

This model has similar identifiability problems as the LC model, and in order to ensure this property the following constraints are defined:

$$\sum_x \beta_x^{(1)} = 1, \qquad \sum_t \kappa_t^{(1)} = 0, \qquad \sum_x \beta_x^{c\,(1)} = 1, \qquad \sum_c \gamma_c^{(1)} = 0$$

2.3.3 Age - Period - Cohort (APC) model

Currie (2006) proposed a special case of the RH model and defined a simpler model called Age-Period-Cohort (APC) model with the following form:

$$\log(\tilde{m}_{t,x}) = \alpha_x + n^{-1}\kappa_t^{(1)} + n^{-1}\gamma_c^{(1)} \tag{2.4}$$

i.e., $p = q = 1$ and $\beta_x^{(1)} = \beta_x^{c\,(1)} = n^{-1}$, according to the generic formula (2.1).

In addition, the following constrains are imposed to ensure the identifiability of the model:

$$\sum_t \kappa_t^{(1)} = 0 \qquad \text{and} \qquad \sum_c \gamma_c^{(1)} = 0$$

One more constraint is needed in order to keep the overall shape of the fitted α_x's as close as possible to the historical average log death rate at age x, using a tilting parameter, δ. The tilting parameter is defined via an iterative scheme, as proposed by Cairns et al. (2009) and according to Lifemetrics implementation. This implies that $\delta = -\frac{\sum_x (x-\bar{x})(\alpha_x - \bar{\alpha}_x)}{\sum_x (x-\bar{x})^2}$ where, $\bar{\alpha}_x = n^{-1}\sum_t \log(\tilde{m}_{t,x})$.

2.3.4 Plat (PL) model

Plat (2009) proposed a multiple component model which includes three age-period effects and a cohort effect with pre-specified age effect terms. The PL model has the following form:

$$\log(\tilde{m}_{t,x}) = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_c^{(1)} \tag{2.5}$$

i.e., $p = 3, q = 1, \beta_x^{(1)} = 1, \beta_x^{(2)} = (\bar{x} - x), \beta_x^{(3)} = (\bar{x} - x)^+$ and $\beta_x^{c(1)} = 1$, according to the generic formula (2.1).

This model incorporates multiple age–period components. $\kappa_t^{(1)}$ represents changes in the level of mortality for all ages, while $\kappa_t^{(2)}$ allows changes in mortality to vary among ages as the improvement rates can differ for different age classes. $\kappa_t^{(3)}$ captures the dynamics of mortality rates at lower ages, which can be different at some times (i.e., the developments like AIDS, drugs and alcohol abuse, and violence).

Similarly to the above-mentioned models, in order to ensure the identifiability of the model, the following set of parameter constrains are imposed:

$$\sum_t \kappa_t^{(3)} = 0, \qquad \sum_c \gamma_c^{(1)} = 0 \qquad \text{and} \qquad \sum_c c\gamma_c^{(1)} = 0$$

Plat (2009) mentions that the above identifiability constraints are set using the same approach as in Cairns et al. (2009), model M6. Both the second and the third constraints are same as those in Cairns et al. (2009), model M6, whereas Plat (2009) considers one additional constraint, which is used in order to normalise the estimates for $\kappa_t^{(3)}$.

2.3.5 CBD models

Cairns et al. (2006*b*) proposed a model which incorporates only two age-period components with pre-specified age effect terms, while there are no static age function and cohort terms. The CBD model has the following form:

$$\text{logit}(\tilde{q}_{t,x}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} \tag{2.6}$$

i.e., $p = 2, q = 0, \beta_x^{(1)} = 1$ and $\beta_x^{(2)} = (x - \bar{x})$, according to the generic formula (2.1) and $\bar{x}$ is the mean age in the sample range.

There is no need for parameter constraints for the CBD model as this structure has no identifiability issues.

Cairns et al. (2009) extended the original CBD model, which was presented previously, and introduced several variations by adding a quadratic age effect and a cohort term. The Quadratic CBD model and its simpler predictor structures have the following form:

$$\text{logit}(\tilde{q}_{t,x}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \hat{\sigma}_x^2)\kappa_t^{(3)} + \gamma_c^{(1)} \tag{2.7}$$

i.e., $p = 3, q = 1, \beta_x^{(1)} = 1, \beta_x^{(2)} = (x - \bar{x}), \beta_x^{(3)} = ((x - \bar{x})^2 - \hat{\sigma}_x^2)$ and $\beta_x^{c\,(1)} = 1$, according to the generic formula (2.1). $\bar{x}$ is the mean age in the sample range and the constant $\hat{\sigma}_x^2$ is the mean of $(x - \bar{x})^2$.

$$\text{logit}(\tilde{q}_{t,x}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_c^{(1)} \tag{2.8}$$

i.e., $p = 2, q = 1, \beta_x^{(1)} = 1$, $\beta_x^{(2)} = (x - \bar{x})$ and $\beta_x^{c\,(1)} = 1$, according to the generic formula (2.1) and $\bar{x}$ is the mean age in the sample range.

$$\text{logit}(\tilde{q}_{t,x}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x_d - x)\gamma_c^{(1)} \tag{2.9}$$

i.e., $p = 2, q = 1, \beta_x^{(1)} = 1$ and $\beta_x^{(2)} = (x - \bar{x})$ and $\beta_x^{c\,(1)} = (x_d - x)$, according to the generic formula (2.1). $\bar{x}$ is the mean age in the sample range and x_d is a constant parameter.

Similarly to the above-mentioned models, in order to ensure the identifiability of the models, the following set of parameter constrains are imposed:

$$\sum_c \gamma_c^{(1)} = 0, \qquad \sum_c c\gamma_c^{(1)} = 0 \qquad \text{and} \qquad \sum_c c^2\gamma_c^{(1)} = 0$$

2.3.6 Model Fitting

In the mortality literature, several software implementations and packages have been developed and made available to the community for making the use of the models a

trivial process for mortality analysts and actuaries. Regardless of the peculiarities of each implementation, the aim is to provide the means to fit the mortality models and perform the parameter estimation.

As reported in Villegas et al. (2018), the **demography** package Hyndman et al. (2015) capitalises on the Lee & Carter (1992) model and its variants presented in Lee & Miller (2001), Booth et al. (2002) and Hyndman & Shahid Ullah (2007). The **ilc** package Butt et al. (2014) offers the Renshaw and Haberman model and the Lee–Carter model under a Poisson regression framework. The **LifeMetrics** R functions (Coughlan et al. (2007)) consider the Cairns–Blake–Dowd (CBD) models and their extensions introduced in Cairns et al. (2009), the Lee–Carter model (using Poisson maximum likelihood), the age–period–cohort model (Osmond (1985), Currie (2006)) and the Renshaw and Haberman model. **StMoMo** presented in Villegas et al. (2018) and has become a well-established mortality modelling R package which provides the tools for fitting stochastic mortality models, assessing their goodness-of-fit and performing mortality projections. In fact, **StMoMo** encompasses the vast majority of stochastic mortality projection models proposed to date.

Overall, each implementation or model may utilise a different parameters estimation method. For instance, Lee & Carter (1992) estimated model parameters using singular value decomposition (SVD) in the context of least squares fitting method, while Renshaw & Haberman (2003) minimised the deviance of their predictor structure. Maximisation of the log-likelihood has been used for the age, period and cohort parameters estimation in Brouhns et al. (2002), Renshaw & Haberman (2006) and Cairns et al. (2009)). Brouhns et al. (2002) described a fitting methodology for the Lee-Carter model based on a Poisson model. Since then, this method has been adopted for parameter estimation also for Renshaw & Haberman (2003), Renshaw & Haberman (2006) and Cairns et al. (2009).

In the context of this PhD thesis, the parameter estimation of the models (LC, RH, APC and PL), that take part in the comparative analysis of Chapters 3 and 4, is achieved by using the R-code of the software package **LifeMetrics** (Coughlan et al. (2007)). Note that, the PL model is not part of the **LifeMetrics** by-default. One can use the archive provided by the Pensions Institute[1], in order to calibrate the stochastic mortality model

[1] `https://www.pensions-institute.org/papers.html` (fitModelsPlat.rtf)

introduced in Plat (2009). This code complies with the specifications of the existing code of the **LifeMetrics** tool. In this tool, the parameters estimation is performed by maximising log–likelihood via the Newton-Raphson iterative process. The estimation process of a parameter utilises the following updating scheme (Goodman (1979), Brouhns et al. (2002)):

$$\hat{\theta}^{(n+1)} = \hat{\theta}^{(n)} - \frac{\partial \ell^{(n)}/\partial \theta}{\partial^2 \ell^{(n)}/\partial \theta^2}$$

starting with an initial value for each parameter. Under the assumption of a Poisson distribution, the log-likelihood is

$$\mathcal{L}(d_{t,x}, \hat{d}_{t,x}) = \sum_t \sum_x \left\{ d_{t,x} \log(\hat{d}_{t,x}) - \hat{d}_{t,x} - \log(d_{t,x}!) \right\} \tag{2.10}$$

and under the assumption of a Binomial distribution, is

$$\mathcal{L}(d_{t,x}, \hat{d}_{t,x}) = \sum_t \sum_x \left\{ d_{t,x} \log\left(\frac{\hat{d}_{t,x}}{E^0_{t,x}}\right) + (E^0_{t,x} - d_{t,x}) \log\left(\frac{E^0_{t,x} - \hat{d}_{t,x}}{E^0_{t,x}}\right) + \log\binom{E^0_{t,x}}{d_{t,x}} \right\} \tag{2.11}$$

In both cases, the expected number of deaths for each model is

$$\hat{d}_{t,x} = E_{t,x} g^{-1}\left(\alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{c(j)} \gamma_c^{(j)}\right) \tag{2.12}$$

with g^{-1} denoting the inverse of the link function g, for an appropriate exposure to risk, $E_{t,x}$.

2.4 Evaluation Metrics

One of the main objectives of this PhD thesis is to provide an extensive experimental testbed in order to evaluate the mortality models from multiple perspectives and reveal their advantages and disadvantages and their modelling capabilities. To do so, it is important to document the evaluation metrics that will be used throughout the experimental testbeds of this thesis. More specifically, we choose to conduct a thorough evaluation over several performance metrics, while we go for metrics that enable the qualitative and quantitative performance analysis of both the fitting and forecasting

processes of the models. Thus, in order to be aligned with the good practices of the domain, we make use of Information Criteria and Percentage Error Tests, as documented in Sections 2.4.1 and 2.4.2 respectively. It has to be noted, that in the context of Chapter 3, this thesis introduces a set of additional qualitative and quantitative metrics (e.g., the Unexplained Variance Ratio (UVR) in Section 3.3) which are tailored to the evaluation testbeds described in that chapter. That is, in this section we opt to document the performance metrics which are predominantly applied in the domain and are applied horizontally to all the testbeds of this thesis.

2.4.1 Information Criteria

For evaluating the *goodness-of-fit* of different models that have different number of parameters, it is common in the mortality literature to use information criteria driven by likelihood. Those criteria are known to penalise those models which include more parameters. In this direction, the Akaike Information Criterion (AIC) and its correction (AICc), which is more suitable for small samples and the Bayesian Information Criterion (BIC) are defined as follows:

$$\text{AIC} = -2l + 2\text{npar} \quad \text{and its correction} \quad \text{AICc} = \text{AIC} + \frac{2\text{npar}(\text{npar}+1)}{N - \text{npar} - 1}$$

and

$$\text{BIC} = -2l + \text{npar} \log N$$

where l is the log-likelihood, $N = n \times \text{a}$ is the number of observations and npar is the number of parameters being estimated. The model, for each dataset, having the lowest AIC, AICc and BIC values are preferable.

2.4.2 Percentage Error Tests

In order to quantify the fitting and forecasting quality of an estimator, the standard statistical approach is to use the Mean Squared Percentage Error (MSPE) test. MSPE measures the average of the squares of the errors, which is the difference between the estimator and what is estimated. MSPE is the second moment of the error, and thus incorporates both the variance of the estimator and its bias. MSPE can be written as the sum of the variance and the squared bias of the estimator. In the mortality

modelling context, a small MSPE value indicates low bias and variance of the estimated central death rates. In addition, another measure is the Mean Absolute Percentage Error (MAPE). MAPE is used to measure how close the estimations are to the actual values, in order to quantify the magnitude of the error. MAPE expresses the accuracy and when is multiplied by 100 represents a percentage error. These measures are defined for all ages from x_1 to x_a and for all years t_1 to t_n as:

$$\text{MSPE} = \frac{1}{(t_n – t_1 + 1) \cdot (x_a – x_1 + 1)} \sum_t \sum_x \left(\frac{\tilde{m}_{t,x} - m_{t,x}}{m_{t,x}} \right)^2$$

$$\text{MAPE} = \frac{1}{(t_n – t_1 + 1) \cdot (x_a – x_1 + 1)} \sum_t \sum_x \frac{|\tilde{m}_{t,x} - m_{t,x}|}{m_{t,x}}$$

where $\tilde{m}_{t,x}$ is either the in–sample or out–of–sample estimated central mortality rate and the $m_{t,x}$ is the observed crude central mortality rate.

2.5 Mortality model design criteria

The continuous developments in the stochastic mortality modelling, led the community to identify a set of criteria that the models need to meet in order to be considered reliable. The following checklist has been created by consolidating the various criteria which have been highlighted in Cairns et al. (2006*a*), Cairns et al. (2008), Cairns et al. (2009), Plat (2009), Cairns et al. (2011).

1. Mortality rates should be positive.
2. The model should be consistent with historical data.
3. Long-term dynamics under the model should be biologically reasonable.
4. Parameter estimates should be robust relative to the period of data and range of ages employed.
5. Model forecasts should be robust relative to the period of data and range of ages employed.
6. Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data.
7. The model should be straightforward to implement using analytical methods or fast numerical algorithms.

8. The model should be relatively parsimonious.
9. It should be possible to use the model to generate sample paths and calculate prediction intervals.
10. The structure of the model should make it possible to incorporate parameter uncertainty in simulations.
11. At least for some countries, the model should incorporate a stochastic cohort effect.
12. The model should have a non-trivial correlation structure.
13. The model should be applicable for a full age range.
14. The model should be transparent and not be treated as a "black box".
15. The model should provide a good fit to the historical data.
16. The model should rank well against other models by criteria such as the BIC.

As can be inferred, the above-mentioned points constitute a long list of requirements that research endeavours on the design of mortality models need to consider. As stated in the 2$^{\text{nd}}$ of objective of this doctoral thesis, our goal is to propose a new mortality model which will go beyond the prominent solutions aiming to address the identified limitations of existing models. That is, our research and development actions need to be aligned with the identified requirements and, if possible, to identify additional desired characteristics to extent this list.

Chapter 3

Introducing and Evaluating a New Multiple Component Stochastic Mortality Model

This chapter builds on top of the analysis conducted in Chapter 2, and introduces and evaluates a new multiple component stochastic mortality model, namely the Hatzopoulos-Sagianou model (HS). Our proposal is based on a parameter estimation methodology, which aims to reveal significant and distinct age clusters by identifying the optimal number of incorporated period and cohort effects. Our methodology adopts Sparse Principal Component Analysis and Generalised Linear Models (GLMs), which firstly introduced in Hatzopoulos & Haberman (2011), while it incorporates several novelties. Precisely, our approach is driven by the Unexplained Variance Ratio (UVR) metric to maximise the captured variance of the mortality data and to regulate the sparsity of the model with the aim of acquiring distinct and significant stochastic components. In this way, the introduced model gains a highly informative structure in an efficient way, while it is able to designate an identified mortality trend to a unique age cluster. In this chapter. we also provide an extensive experimental testbed to evaluate the efficiency of the proposed model in terms of fitting and forecasting performance over several datasets (Greece, England & Wales, France and Japan), while we compare our results to those of well-known mortality models (Lee–Carter, Renshaw–Haberman, Currie (APC), and Plat). The HS model is able to achieve high scores over diverse

qualitative and quantitative evaluation metrics and outperforms the rest of the models in the majority of the experiments. Our results advocate the beneficial characteristics of the proposed model and come into agreement with well–established findings of the mortality literature.

3.1 Introduction

As thoroughly analysed in Chapter 2, each of the mortality models bears beneficial and desirable characteristics, while, taking realistically, they are not free of weaknesses. Considering the extensive stochastic mortality literature, such a model should incorporate some fundamental and desirable characteristics. One of the most important characteristics of a model is its ability to fit accurately the most important features of the historical mortality data. Therefore, it is desirable for a model to have multiple factors in order to capture the improvement rates that differ at different age ranges. In this way, a model should reveal non-trivial correlation structures between changes in mortality rates among the different ages. A contemporary model should have the ability to incorporate cohort effects. According to the literature and based on several implementations on mortality data (Cairns et al. (2009), Plat (2009)), those models that do not incorporate cohort effects fail to provide a good fit to the historical data, especially for those countries where a cohort effect has been observed. In addition, a mortality model should be consistent among diverse datasets and the parameters' estimations should be robust across the range of the used data. However, the above mentioned desirable characteristics inevitably bring in a higher complexity to a model in terms of the number of the incorporated parameters and the implementation's stiffness. Thus, it is important to keep the balance between the effectiveness of a model and its overall complexity.

To this end, in this chapter, a new multiple component stochastic mortality model is proposed. Our approach aims to combine the aforementioned advantageous characteristics to provide a model which is able, not only to accurately fit to historical data and provide a solid base for forecasting mortality rates, but also to capture the age-shifts of mortality improvements under a robust and parsimonious manner. To do so, our model is based on a new parameters estimation method, which adopts fundamental elements of the Hatzopoulos & Haberman (2011) (HH) model. We evaluate and we prove the efficiency of our model through an extensive experimental testbed. More specifically, this

chapter provides a comprehensive and comparative analysis of our model with respect to the Lee-Carter (LC), Renshaw-Haberman (RH), Age-Period-Cohort (APC), Plat (PL) models, utilising mortality data from various countries. Note that, the selected models are widely used in the literature (Coughlan et al. (2007), Cairns et al. (2009), Dowd et al. (2010), Mitchell et al. (2013)), and they are selected among others for their ability to operate in the whole age range. Our analysis is performed using data from Greece, England & Wales, France and Japan, for males, as those provided by Human-Mortality-Database and Eurostat. These datasets form a comprehensive testbed to fully explore the capacity of each mortality model, as they consist of diverse time periods and data quality.

In short, the contributions of the work presented in this chapter are as follows:

- We propose a new stochastic mortality model, which utilises multiple age–period and age–cohort interaction terms, and we introduce a parameter estimation method. This method adopts Sparse Principal Component Analysis (SPCA) and Generalised Linear Models (GLMs), which firstly introduced in Hatzopoulos & Haberman (2011), while it incorporates several novelties.
- We introduce a novel methodology based on Unexplained Variance Ratio (UVR) metric in order to pinpoint the most important age–period and age–cohort components incorporated in the proposed model. The optimal model structure is acquired by defining the optimal scalar s which regulates the sparsity in SPCA.
- We extensively evaluate our model over several metrics and diverse datasets, while we compare its efficiency against well-known existing mortality models.
- In the comparative evaluation, we use a backtesting framework (out-of-sample comparison) in order to evaluate the ability of each model in terms of prediction. We forecast the period and the cohort indexes using three different forecasting models, namely the Random Walk with Drift (RWD) model, optimum Arima models based on Bayesian Information Criterion (BIC) test and a specific class of Dynamic Linear Regression (DLR) models.

The rest of this chapter is organised as follows: Given the preliminaries presented in Chapter 2, Section 3.2 presents the proposed mortality model including its structure and the utilised estimation method. Section 3.3 outlines the evaluation methodology we followed, and the comparative results are analysed in Section 3.4. Note that, due to the

extensive analysis, the majority of the described results are given in the Appendices of this thesis (Appendices A, B and C). Section 3.5 provides an overall discussion on the findings and concludes this chapter.

3.2 The proposed Stochastic Mortality Model

In this section we present the proposed stochastic mortality model. In contrast to the models presented in Section 2.3, our model utilises a different semi–parametric estimation method. In the context of our model, we propose an estimation method following the one presented in Hatzopoulos & Haberman (2011), utilising Generalised Linear Models (GLMs) and Sparse Principal Component Analysis (SPCA). Even if we adopt these fundamental elements, the proposed estimation method bares major advantages which are discussed in Section 3.4.7. In addition, we present the Unexplained Variance Ratio (UVR) method, which is used to determine the sparsity factor (s–value) required in the SPCA method, as well as to define the most important age–period and age–cohort components which are incorporated in the proposed model.

3.2.1 Estimation Methodology

3.2.1.1 Age–Period effect estimation methodology

As described in the Section 2.1 of Chapter 2, our data consist of the number of deaths ($D_{t,x}$), with their corresponding central exposure to the risk ($E^c_{t,x}$), for each calendar year $t = t_1, \ldots, t_n$ and for each age $x = x_1, \ldots, x_{\text{a}}$ last birthday. The number of deaths, $D_{t,x}$, follows a Poisson distribution with mean $E^c_{t,x} \cdot m_{t,x}$ (Brillinger (1986), Booth et al. (2006)). The form of our data is shown in Figure 3.1 (parallelogram BCDE). Intuitively, each vertical line represents the data of each calendar year t, which are used to estimate the age–period effects. To do so, a GLM approach is applied, with response variables $D_{t,x}$, for each calendar year t independently, while the logarithmic exposure variables (i.e., $\log(E^c_{t,x})$) are handled as an offset by using the natural log link function:

$$\log(\hat{m}_t) = b_t \cdot L^T \tag{3.1}$$

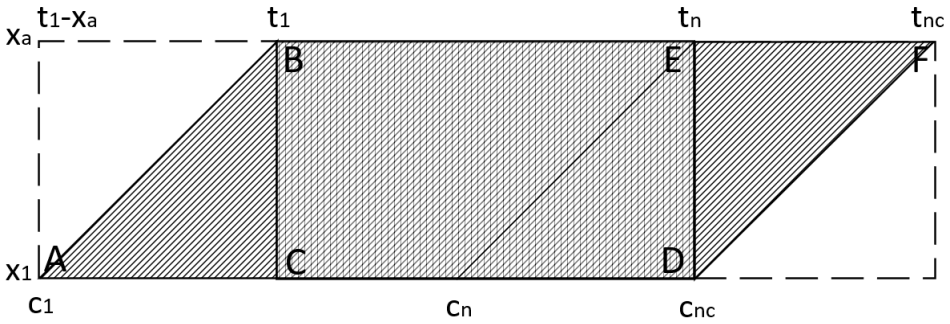

FIGURE 3.1: Data form.

where, $L = \{l_{x,k-1}\}$, $k = 1, \ldots, k_1$, is an (a $\times$ k_1)-dimensional matrix of orthonormal polynomials with degree of $k_1 - 1$ for age x, and b_t denotes the vector of the GLM–estimated parameters for calendar year t. Therefore, for all calendar years $t = t_1, \ldots, t_n$, a random (asymptotically normal) matrix $B = \{b_{t,k-1}\}$, $k = 1, \ldots, k_1$, of order $(n \times k_1)$, containing the GLM–estimated parameters, is produced using the GLM approach.

The overall optimum degree $k_1 - 1$ of the orthonormal polynomials is determined by the following tests: X^2 test, Sign test, Run test, Ljung–Box Q test, Likelihood ratio test and Bayesian information criterion. The overall optimum degree $k_1 - 1$ (considering all the calendar years), is chosen to be that value which minimises the total BIC value for all calendar years. An additional verification for defining the optimum degree is employed by detecting which k_1 value gives the maximum number of statistically acceptable p–values of the aforementioned tests (for calendar years). If these two methods (BIC, p–values) give different optimum k_1 values then we choose the larger one, in case we need to capture specific intrinsic mortality characteristics more accurately (for example the "accident hump" effect), otherwise we choose the smaller optimum k_1, if we need a more parsimonious model or the mortality data quality is pure.

For reducing further the dimensionality of the problem by keeping only those factors which explain most of the information in the data, Sparse Principal Component Analysis (SPCA) is applied to the associated covariance matrix of the GLM–estimated parameters B. According to Luss & d'Aspremont (2006)[1], given the covariance matrix $A = \text{Cov}(B)$, the (dual) problem of finding a sparse factor which explains the maximum amount of variance in the data can be written as follows:

$$\min \lambda^{\max}(A + U), \quad \text{s.t. } |U_{ij}| \leq s \tag{3.2}$$

[1]the associated Matlab package can be found in https://www.di.ens.fr/~aspremon/DSPCA.html

where s is a scalar which defines the sparsity. In order to define the appropriate "optimum" s–value, we propose the Unexplained Variance Ratio (UVR) approach which is discussed in Section 3.2.1.3.

The covariance matrix U in equation (3.2) is the solution to the problem. By applying eigenvalue decomposition on U, the associated matrix of eigenvectors P, of order $(k_1 \times k_1)$, is acquired. Thus, the log–graduated central mortality rates, for each calendar year t, can be decomposed as:

$$\begin{aligned} \log(\hat{m}_t) = b_t \cdot L^T = \left(\bar{b} + b_t^r\right) \cdot L^T = \bar{b} \cdot L^T + b_t^r \cdot L^T = \\ = \bar{b} \cdot L^T + (b_t^r \cdot P) \cdot (P^{-1} \cdot L^T) = \tilde{\alpha} + \kappa_t \cdot G \end{aligned} \tag{3.3}$$

where, $\tilde{\alpha} = \bar{b} \cdot L^T$ is the main age–profile, $\bar{b}$ is a k_1–dimensional vector which denotes the column arithmetic mean of matrix B, and b_t^r is a k_1–dimensional vector which denotes the re-scaled vector: $b_t^r = b_t - \bar{b}$, for each calendar year t. The matrix $G = P^{-1} \cdot L^T$, of order $(k_1 \times \text{a})$, corresponds to the age–effect scores, with rows $\beta^{(i)} = e_i^T \cdot L^T$, for each i^{th}–component and values $\beta_x^{(i)} = e_i^T \cdot l_x^T$, where e_i denotes the i^{th}–column of the $(k_1 \times k_1)$–dimensional matrix $P = \{e_i\}$, $i = 1, \ldots, k_1$ and $l_x = (l_{x,0}, l_{x,1}, \ldots, l_{x,k_1-1})$ denotes the x^{th}–row vector of the matrix L. The vector $\kappa_t = b_t^r \cdot P$ is the period–effect scores. Hence, for all calendar years t, the $(n \times k_1)$–dimensional matrix $Y = \{\kappa_t^{(i)}\}$, $i = 1, \ldots, k_1$ is defined, with columns $\kappa^{(i)}$ for each i–component and values $\kappa_t^{(i)} = (\kappa_t^{(1)}, \ldots, \kappa_t^{(k_1)})$.

Thus, an alternative expression for the log–graduated mortality rates, for calendar year t and age x is:

$$\log(\hat{m}_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{k_1} \beta_x^{(i)} \cdot \kappa_t^{(i)} \tag{3.4}$$

Our goal is to reduce the number of age–period components from k_1 to p:

$$\log(\hat{m}_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \varepsilon_{t,x} \tag{3.5}$$

where, the residuals or the disturbance term $\varepsilon_{t,x} \sim N(0, v_x)$ is the error component at age x in year t and denotes the deviation of the model represented by the excluded Sparse Principal Components (SPCs). The corresponding method in order to define

the most important p age–period components following the model structure given by equation (3.5) is described in Section 3.2.1.3.

At this point, one can realize the importance of GLM and SPCA in the context of our model. Firstly, the GLM approach does not involve identifiability problems in parameter estimation, giving a unique solution. Besides, the GLM-polynomial structure produces the matrix B, which inherits all the mortality dynamics incorporated in central mortality rates, $\hat{m}_{t,x}$, reducing the dimension of the problem from a to k_1, and also produces components which are based on the graduated mortality rates. In addition, by applying SPCA to the covariance matrix of the GLM–estimated parameters, we are able to extract most of the information of the data by utilizing only a few components, which explain the maximum amount of variance in the data. Most importantly, for non–stationary time series, as Hatzopoulos & Haberman (2011) have been analyzed and compared, SPCA gives a better clustering and more distinguished mortality dynamics than Principal Component Analysis (PCA). An overview of the proposed age–period construction methodology is given in Algorithm 1.

Algorithm 1: Age–Period component structure

Data: D, E, k_1
Result: M

Input: $D : (n \times \mathrm{a})$-matrix with values $D_{t,x}, \forall t, x$
$E : (n \times \mathrm{a})$-matrix with values $E_{t,x}, \forall t, x$
$k_1 - 1$: optimum degree of polynomials
Output: $M : (n \times \mathrm{a})$-matrix representing $\log(\hat{m}_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)}$

Define parameters
$x \leftarrow x_1, \ldots, x_{\mathrm{a}}$; $\mathrm{a} \leftarrow \mathrm{length}(x)$; // ages
$t \leftarrow t_1, \ldots, t_n$; $n \leftarrow \mathrm{length}(t)$; // calendar years
for $i \leftarrow 0$ **to** $k_1 - 1$ **do**
  $L \leftarrow$ create orthonormal polynomials $\forall x$;
end
$B \leftarrow \mathrm{GLM}(L, D)$ with offset $\log(E)$, $\forall t$;
$A \leftarrow \mathrm{Cov}(B)$;
initial_$s \leftarrow$ is a set of values around the area of the $\mathrm{Var}(b_{t_1,0})$;
for $s \in$ initial_s **do**
  $U \leftarrow \mathrm{SPCA}(A, s)$;
  $P \leftarrow \mathrm{eig}(U)$; // P : eigenvectors of U
  $\bar{b} \leftarrow \mathrm{mean}(B)$;
  $B^r \leftarrow B - \bar{b}$;
  $\tilde{\alpha}_x \leftarrow \bar{b} \cdot L$; // main age-profile
  $Y \leftarrow B^r \cdot P$; // $Y = \{\kappa^{(i)}\}$ i-period effect, $\forall i \in [1, k_1]$
  $G \leftarrow P^{-1} \cdot L^T$; // $G = \{\beta^{(i)}\}$ i-age effect, $\forall i \in [1, k_1]$
  for $i \leftarrow 1$ **to** k_1 **do**
    if ($UVR(i) < 50\%$) **and** ($UVR(i)$ reveals unique age cluster) **then**
      cand_$s \leftarrow s$;
      include this i-component to the model structure;
    else
      reject s;
    end
  end
  // repeat the same procedure for the next initial_s value
end
// cand_$s \leftarrow$ is a vector with acceptable initial_s values using UVR technique
filtered_$s \leftarrow$ from cand_s choose those s-values which achieve the maximum number of significant p components;
opt_$s \leftarrow$ from filtered_s choose the s-value which gives the minimum BIC value
// after UVR criterion we have concluded in the optimum s value and the optimum age-period components (p)
for $i \leftarrow 1$ **to** n **do**
  $M(i, 1 : \mathrm{a}) \leftarrow Y(i, 1 : p) \cdot G(1 : p, 1 : \mathrm{a}) + \tilde{\alpha}$;
end
return M;

3.2.1.2 Age–Cohort effect estimation methodology

The age–cohort components are estimated in a similar way with the age–period components, under the assumption that the cohort effect has dynamics that are independent of the period effect (Renshaw & Haberman (2006)). Therefore, according to equation (3.5), the residuals $\varepsilon_{t,x}$ are modelled in age–cohort effects, by applying the GLM approach. Hence, our data, $D_{t,x}$, $E^c_{t,x}$ and $\hat{m}_{t,x}$, are processed as cohorts. As can be seen intuitively in Figure 3.1 (parallelogram BADF), each diagonal line represents the data

for each cohort $c = t - x = c_1, \ldots, c_{n_c}$, where $n_c = n + (a - 1)$, which are used to estimate the age–cohort effects. For the implementation of the GLM approach to the cohort data, the values outside the range of available calendar years (i.e., Figure 3.1, triangles ACB and DFE), are set equal to zero. The GLM approach is applied, with response variables the number of deaths, for each cohort c independently, while the logarithmic exposures and the log–graduated mortality rates with p age–period effects (equation (3.5)) are handled as an offset:

$$\log(\hat{\hat{m}}_c) = b_c \cdot (L^c)^T \tag{3.6}$$

where, $L^c = \{l^c_{x,k-1}\}$, for $k = 1, \ldots, k_2$, is an $(a \times k_2)$-dimensional matrix, which contains the orthonormal polynomials with degree of $k_2 - 1$ for age x and b_c denotes the vector of the GLM–estimated parameters for each cohort c. Thus, a $(n_c \times k_2)$–dimensional matrix $B^c = \{b_{c,k-1}\}$, $k = 1, \ldots, k_2$, which contains the GLM–estimated parameters for all cohorts $c = c_1, \ldots, c_{n_c}$, is produced.

The optimum degree $k_2 - 1$ of the orthogonal polynomials L^c is determined using the same tests and criteria described in the Section 3.2.1.1, in age–period effect estimation methodology.

By applying eigenvalue decomposition on the matrix of the GLM–estimated parameters B^c using the common PCA (or equivalent SVD) technique, the associated matrix of eigenvectors $P^c = (e^c_1, \ldots, e^c_{k_2})$, of order $(k_2 \times k_2)$, is produced. In this step, we apply PCA to cohort dynamics, as the most important non–stationary mortality dynamics have been incorporated in age-period effect structure, and the remaining cohort–effects constitute stationary processes. Under the proposed SPCA structure, the first few age–period interaction terms explain the vast majority of the mortality variance (see Table A.17) and therefore they capture the most significant non–stationary mortality dynamics.

Hence, the cohort log–graduated central mortality rates, for each cohort c, can be decomposed as:

$$\begin{aligned} \log(\hat{\hat{m}}_c) = b_c \cdot (L^c)^T = \left(\bar{b}^c + b^r_c\right) \cdot (L^c)^T = \bar{b}^c \cdot (L^c)^T + b^r_c \cdot (L^c)^T = \\ = \bar{b}^c \cdot (L^c)^T + b^r_c \cdot P^c \cdot (P^c)^{-1} \cdot (L^c)^T = \tilde{\alpha}^c + \gamma_c \cdot G^c \end{aligned} \tag{3.7}$$

where, $\tilde{\alpha}^c = \bar{b}^c \cdot (L^c)^T$ is the main age–profile, $\bar{b}^c$ is a k_2–dimensional vector which denotes the column arithmetic mean of matrix B^c, and b^r_c is a k_2–dimensional vector

which denotes the re-scaled vector: $b_c^r = b_c - \bar{b}^{\mathrm{c}}$, for each cohort c. The matrix $G^{\mathrm{c}} = (P^{\mathrm{c}})^{-1} \cdot (L^{\mathrm{c}})^T$, of order $(k_2 \times \mathrm{a})$, is the age–effect scores, with rows $\beta^{\mathrm{c}(j)} = (e_j^{\mathrm{c}})^T \cdot (L^{\mathrm{c}})^T$, for each j^{th}–component and values $\beta_x^{\mathrm{c}(j)} = (e_j^{\mathrm{c}})^T \cdot (l_x^{\mathrm{c}})^T$, where e_j^{c} denotes the j^th–column of the matrix P^{c} and $l_x^{\mathrm{c}} = (l_{x,0}^{\mathrm{c}}, l_{x,1}^{\mathrm{c}}, \ldots, l_{x,k_2-1}^{\mathrm{c}})$ denotes the x^{th}–row vector of the matrix L^{c}. The vector $\gamma_c = b_c^r \cdot P^{\mathrm{c}}$ is the cohort–effect scores, thus for all cohorts c, the $(n_c \times k_2)$–dimensional matrix $Y^{\mathrm{c}} = \{\gamma_c^{(j)}\}$, $j = 1, \ldots, k_2$ is defined, with columns $\gamma^{(j)}$ for each j–component and values $\gamma_{\mathrm{c}}^{(j)} = (\gamma_c^{(1)}, \ldots, \gamma_c^{(k_2)})$.

After reducing the number of components from k_2 to q, an alternative expression for the cohort log–graduated mortality rates, for cohorts $c = c_1, \ldots, c_{n_{\mathrm{c}}}$ and ages $x = x_1, \ldots, x_{\mathrm{a}}$ is:

$$\log(\hat{\tilde{m}}_{c,x}) = \tilde{\alpha}_x^{\mathrm{c}} + \sum_{j=1}^{q} \beta_x^{\mathrm{c}(j)} \cdot \gamma_{\mathrm{c}}^{(j)} + \varepsilon_{c,x} \tag{3.8}$$

The corresponding method in order to define the most important q age–cohort components is described in Section 3.2.1.3. An overview of the proposed age–cohort construction methodology is given in Algorithm 2.

3.2.1.3 Sparsity regulation for optimal component selection

This section describes the methodology followed to define the optimal number of p and q components of the equations (3.5) and (3.8) respectively. Initially, it is worth to mention that, the number of p optimum components retained for representing the age–period effects is very important for the identification of the q age–cohort effects. By choosing fewer factors than the optimum p number, then these disregarded period effects will be used as cohort effects, in a non–stationary and more complex manner, whereas, by choosing more age–period factors, then the cohort structure will break down, as these additional age–period components, will describe significant cohort effects (for cases where cohort effects have been observed), otherwise they will describe insignificant mortality dynamics. The optimal model structure is defined through the optimality of the s–value. Different choices of s–value, uncover different SPCs in the mortality data. Thus, the process of defining the optimal model structure is reflected in defining the optimal scalar s, which in turn regulates the sparsity in SPCA. The definition of the s–value is not a straight–forward process as it depends on the experimental data, and in

Algorithm 2: Age–Cohort component structure

Data: D, E, M, k_2
Result: M^{c}

Input: $D : (n_{\mathrm{c}} \times \mathrm{a})$-matrix with values $D_{c,x}, \forall c, x$
$E : (n_{\mathrm{c}} \times \mathrm{a})$-matrix with values $E_{c,x}, \forall c, x$
$M : (n_{\mathrm{c}} \times \mathrm{a})$-matrix with values $M_{c,x}, \forall c, x$
$k_2 - 1$: optimum degree of polynomials
Output: $M^{\mathrm{c}} : (n_{\mathrm{c}} \times \mathrm{a})$-matrix representing $\log(\hat{m}_{c,x}) = \tilde{\alpha}^{\mathrm{c}}_x + \sum_{j=1}^{q} \beta^{\mathrm{c}(j)}_x \cdot \gamma^{(j)}_c$

Define parameters
$x \leftarrow x_1, \ldots, x_{\mathrm{a}}$; $\mathrm{a} \leftarrow \mathrm{length}(x)$; `// ages`
$c_1 \leftarrow t_1 - x_{\mathrm{a}}$; $c_{n_{\mathrm{c}}} \leftarrow t_n - x_1$;
$c \leftarrow c_1, \ldots, c_{n_{\mathrm{c}}}$; $n_{\mathrm{c}} \leftarrow \mathrm{length}(c)$; `// cohorts`
for $i \leftarrow 0$ **to** $k_2 - 1$ **do**
  $L^{\mathrm{c}} \leftarrow$ create orthonormal polynomials $\forall x$;
end
$B^{\mathrm{c}} \leftarrow \mathrm{GLM}(L^{\mathrm{c}}, D)$, offset: $\log(E)$ and the age–period effects, (M), $\forall c$;
$P^{\mathrm{c}} \leftarrow \mathrm{PCA}(B^{\mathrm{c}})$; `//` P^{c} `: eigenvectors of` B^{c}
$\bar{b}^{\mathrm{c}} \leftarrow \mathrm{mean}(B^{\mathrm{c}})$;
$B^{r} \leftarrow B^{\mathrm{c}} - \bar{b}^{\mathrm{c}}$;
$\tilde{\alpha}^{\mathrm{c}} \leftarrow \bar{b}^{\mathrm{c}} \cdot L^{\mathrm{c}}$; `// main age-profile`
$Y^{\mathrm{c}} \leftarrow B^{r} \cdot P^{\mathrm{c}}$; `//` $Y^{\mathrm{c}} = \{\gamma^{(j)}\}$ `j-cohort effect,` $\forall i \in [1, k_2]$
$G^{\mathrm{c}} \leftarrow (P^{\mathrm{c}})^{-1} \cdot (L^{\mathrm{c}})^{T}$; `//` $G^{\mathrm{c}} = \{\beta^{\mathrm{c}(j)}_x\}$ `j-age effect,` $\forall i \in [1, k_2]$
for $i \leftarrow 1$ **to** k_2 **do**
  if $(UVR^{\mathrm{c}}(i) < 50\%)$ **and** ($UVR^{\mathrm{c}}(i)$ reveals unique age cluster) **then**
    include this i-component to the model structure
  end
end
`// after UVR criterion we have concluded in the optimum age-cohort components (`q`)`
for $i \leftarrow 1$ **to** n_{c} **do**
  $M^{\mathrm{c}}(i, 1 : \mathrm{a}) \leftarrow Y^{\mathrm{c}}(i, 1 : q) \cdot G^{\mathrm{c}}(1 : q, 1 : \mathrm{a}) + \tilde{\alpha}^{\mathrm{c}}$;
end
return M^{c};

most cases, requires deep domain knowledge and multiple experiments for each dataset. However, in the context of our model we provide an automated approach which utilises the Unexplained Variance Ratio (UVR) criterion for converging the optimal s–value.

In fact, through the UVR, we first locate a range of initial s–values. Based on our experimental study, this range includes the values of the area around the variance of the GLM–parameter $b_{t_1,0}$. For each trial s–value we define the UVR, for each component and for each age, as a modified version of the one considered in Li et al. (2004). We calculate the empirical variance of the time series of log mortality rates and the variance of the error between the estimated log mortality rates and the crude ones, for each age x and each component i. The ratio of these two variances for successive components

defines the $UVR_x(i)$ value:

$$UVR_x(i) = \begin{cases} \frac{\mathrm{Var}\left(\log(m_{t,x}) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)}\right)}{\mathrm{Var}\left(\log(m_{t,x}) - \alpha_x\right)}, & i = 1 \\ \\ \frac{\mathrm{Var}\left(\log(m_{t,x}) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)} - \beta_x^{(i)}\kappa_t^{(i)}\right)}{\mathrm{Var}\left(\log(m_{t,x}) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)}\right)}, & i = 2, \ldots, p \end{cases} \tag{3.9}$$

Hence, for the first component ($i = 1$), we calculate the ratio between the variance of the crude central mortality rates and the variance of the error between the estimated central mortality rates and the crude ones, including only the first component. These $UVR_x(1)$ values will indicate which ages are explained by the first most important component.

Then, in order to evaluate the contribution of each added component, we calculate each $UVR_x(i)$, for $i > 1$ with respect to the first component. In this way, we evaluate if any added i–component, for $i = 2, \ldots, k_1$, explains further significant information for a different range of ages, with respect to the first component.

An empirical threshold value for significant UVR–values is when $UVR_x(i) < 50\%$. By following, the 50% rule, for significant $UVR_x(i)$ values, we set the restriction that the explained variance has to be greater than the associated unexplained one. However, this is not a constant value, but an experimental threshold, which can be tuned depending on the data and our preferences. Thus, one can select more informative "optimal" components by using a more stringent threshold. If less stringent threshold values are used, the model will define different and less significant components, as expected. This is because the determination of the UVR value affects the optimality of the s value, and hence, different s–values uncover different SPCs in the mortality data. Though, a less stringent threshold value (i.e., greater than 50%) would be suggested in case of poor mortality data. We note that, the $UVR_x(i)$ values are pure numbers and independent on the scale of the variance. Thus, for a given s–value, we evaluate the particular model structure in terms of the $UVR_x(i)$ values, for $i = 2, \ldots, k_1$. If these UVR–values reproduce a distinct and unique significant clustering of ages, for the i–component, and $UVR_x(i) < 50\%$, then we accept this particular s–value, and we also include the i–component to the model structure.

Given the above, if there are significant ages in a neighbourhood of an age, then this

implies the existence of a significant cluster of ages. If any particular s–value gives a unique cluster of ages for at least one component then this s–value is a candidate optimum one. Otherwise, if the s–value does not reveal any cluster or it reveals more than one clusters (i.e., the same component describes different mortality dynamics for diverse range of ages), for all components, then it is rejected. Therefore, it is desirable that the above approach should produce distinct and unique key age groups of UVR neighbouring significant values. Among all the candidate s–values, we choose those that maximise the number of significant SPCs. Therefore, based on this approach, we end up with a filtered set of accepted s–values, which maximised the significant SPCs, as defined by the UVR criterion. Among these filtered s–values, we get the optimum value which is the one that gives the minimum BIC value.

A similar approach is applied to the cohort effects for the derivation of the optimum number of cohort related factors, according to the equation (3.7). Since, the cohort structure is define through a process based on PCA, there is no sparsity factor involved. Thus, our method concludes to the optimum number of age–cohort components by selecting those that provide a $UVR_x^{\mathrm{c}}(j) < 50\%$, for $j = 1, \ldots, k_2$ and reveal a unique age cluster. We calculate the ratio of the variance of the error between the estimated central mortality rates and the crude ones including the p age–period components, and the variance of the error between the estimated mortality rates and the crude ones including the p age–period components and the j^{th} age–cohort component.

Hence, the $UVR_x^{\mathrm{c}}(j)$ value for each age–cohort component is defined as:

$$UVR_x^{\mathrm{c}}(j) = \frac{\mathrm{Var}\big(\log(m_{t,x}) - \alpha_x - \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} - \beta_x^{\mathrm{c}\,(j)} \cdot \gamma_c^{(j)}\big)}{\mathrm{Var}\big(\log(m_{t,x}) - \alpha_x - \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)}\big)}, \qquad j = 1, \ldots, q$$

In this way, we evaluate the contribution of each added age–cohort component by calculating each $UVR_x^{\mathrm{c}}(j)$, for $j = 1, \ldots, q$, with respect to the p age–period components.

3.2.2 Final Model Structure

According to the UVR-based method, after estimating the significant age–period and age–cohort effects the final estimates of the log–graduated central mortality rates in age,

period and cohort effects are defined, following equations (3.5) and (3.8):

$$\log(\tilde{m}_{t,x,c}) = \log(\hat{m}_{t,x}) + \log(\hat{\tilde{m}}_{c,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \tilde{\alpha}_x^{\mathrm{c}} + \sum_{j=1}^{q} \beta_x^{\mathrm{c}(j)} \cdot \gamma_c^{(j)}$$

for the calendars years $t = t_1, \ldots, t_n$, ages $x = x_1, \ldots, x_{\mathrm{a}}$ and cohorts $c = t - x = c_1, \ldots, c_{n_{\mathrm{c}}}$.

Therefore, the estimated log–graduated central mortality rates can be alternatively decomposed in a matrix form as follows:

$$\begin{aligned} \log(\tilde{m}) = \bar{b} \cdot L^T + \bar{b}^{\mathrm{c}} \cdot (L^{\mathrm{c}})^T + \\ + \left(\sum_{i=1}^{p} \kappa_t^{(i)} e_i^T \right) \cdot L^T + \left(\sum_{j=1}^{q} \gamma_c^{(j)} (e_j^{\mathrm{c}})^T \right) \cdot (L^{\mathrm{c}})^T \end{aligned} \tag{3.10}$$

and according to the generic formula (2.1),

$$\begin{aligned} \alpha_x &= \tilde{\alpha}_x + \tilde{\alpha}_x^{\mathrm{c}} = \bar{b} \cdot L^T + \bar{b}^{\mathrm{c}} \cdot (L^{\mathrm{c}})^T, \\ \beta_x^{(i)} &= e_i^T \cdot l_x^T, \qquad & \kappa_t^{(i)} &= b_t^r \cdot e_i \\ \beta_x^{\mathrm{c}(j)} &= (e_j^{\mathrm{c}})^T \cdot (l_x^{\mathrm{c}})^T, \qquad & \gamma_c^{(j)} &= b_c^r \cdot e_j^{\mathrm{c}} \end{aligned}$$

Consequently, the total number of parameters which are needed in order to define the proposed model include: two parameters to define the order of orthogonal polynomials k_1 and k_2, $k_1 + k_2$ parameters to estimate the age profile α_x, $(n + k_1) \cdot p$ parameters for the estimation of the age–period scores, and $(n_{\mathrm{c}} + k_2) \cdot q$ parameters for the estimation of the age–cohort scores. Thus, the total number of the parameters incorporated in the proposed model is: $2 + (k_1 + k_2) + p \cdot (n + k_1) + q \cdot (n_{\mathrm{c}} + k_2)$. The total number of the model parameters is affected, among others, by the number of the identified age–period components (i.e., p). The optimal number of p is in turn affected by the s–value. Notably, the s–value depends on the inherent characteristics of the mortality data, as mentioned in section 3.2.1.3. However, the number of the identified age–period components is by no means proportional to the size of s–value.

Note that, the estimation of the proposed model is done sequentially by first fitting the age–period effects and then the age–cohort effects. However, it has to be stated that swapping the order of the estimation process is not feasible in the context of our particular model structure, due to its inherent characteristics. The utilisation of a Poison

GLM estimation implies that deaths are independent among the ages and hence, GLM estimates can be estimated in a Poisson model. Thus, in this setting, it is not possible to estimate the age–cohort effects first, since deaths are not independent among the ages. Additionally, a complete cohort (of at least 90 calendar years) is needed for the estimation procedure, due to utilisation of the GLM approach. That is, in cases where mortality data have a limited calendar range, like in Greek and Japanese datasets, the main trends and the captured maximum amount of variance would be inadequate. A swapping approach is discussed and implemented in Hatzopoulos & Haberman (2015).

3.3 Evaluation methodology

In this section we describe the experimental methodology followed for evaluating the performance of the mortality models presented in Sections 2.3.1 to 2.3.4, and provide a comparative analysis against the HS model, in order to prove its capacity. Note that, the CBD (Cairns et al. (2006*b*)) model and its extensions (Cairns et al. (2009)) are not considered in the evaluation, as those models are designed for higher ages only, and their fit quality for the whole age ranges is relatively poor (Cairns et al. (2009), Plat (2009)). We intend to provide an extensive experimental testbed in order to evaluate the models from multiple perspectives. More specifically, we choose to conduct a thorough evaluation over several datasets and performance metrics, while we compare the fitting and forecasting performance of the proposed model with respect to the other mortality models.

3.3.1 UVR analysis

Initially, by using the UVR metric described in Section 3.2.1.3, we can visualise the significant amount of variance explained by each of the components of a model, and in which group of ages this variance corresponds to. Therefore, we apply this method, to all models that take part in the evaluation, to derive the age clusters and the corresponding unexplained variance reflected by each component. In this respect, UVR gives a qualitative overview of each model structure.

However, it has to be stated that for the PL model, we utilise a modified version of the UVR metric. The $UVR_x(i)$ values, for each added component, are calculated with

respect to all the previous $(i-1)$–components, and not with respect to the 1st component as defined for the proposed model in Section 3.2.1.3. Thus, for the PL case, we calculate the ratio of the variances of the error, which is between the estimated mortality rates and the crude ones, including i–components (in age–period/cohort effects), and the variances of the error, which is between the estimated rates and the crude ones including $(i-1)$–components. Hence, the difference between the two UVR approaches, is reflected only on PL's 3rd age–period component. In our model, each added component is derived independently with respect to the 1st component. In all mortality experiments, the 1st age–period factor accounts for the vast majority of the total explained variance, as defined later in Section 3.3.3.1 and all the remaining components explain relative deviations from the 1st interaction term. On the other hand, in PL model each added component explains relative deviation from all the previous age–period components in a cumulative manner due to the different construction and estimation employed.

3.3.2 Forecasting

In order to evaluate the ability of the mortality models in terms of prediction we use a backtesting framework with a k–step ahead forecast and we compare the out–of–sample forecasted mortality rates to the actual ones. For providing an objective and unbiased evaluation, we forecast the period and the cohort indexes using three different forecasting models namely, the Random Walk with Drift (RWD) (ARIMA(0,1,0) with drift), the appropriate ARIMA model based on Bayesian Information Criterion (BstAr), and a specific class of Dynamic Linear Regression (DLR) models. In fact, RWD model is a widely adopted forecasting model in the mortality literature (mainly for simplicity reasons), while BstAr is the ARIMA model that can achieve the best BIC performance. For the determination of the BstAr model, the R–code of the software package "forecast" (Hyndman et al. (2019)) was used. DLR models is another class of time series forecasting and for the realisation the Captain Toolbox for Matlab (Young et al. (2009)) was used. That is, the forecasting process consists of the following steps:

1. Fit the models to the mortality experience for years $t_1, \ldots, t_{n-k}$ and for ages $x_1, \ldots, x_{\mathrm{a}}$.

2. Predict the mortality rates k years ahead for years $t_{n-k+1}, \ldots, t_n$.

3. Calculate the error between the out–of–sample forecasted mortality rates and the actual mortality rates for years $t_{n-k+1}, \ldots, t_n$.

4. Evaluate the models using percentage error tests.

Therefore, forecast estimates are needed for the period and cohort dynamics: $\kappa_t^{(i)}$ and $\gamma_c^{(i)}$. Specifically, regarding the projection using RWD, each i–period index is modelled using RWD, whereas each j–cohort index is modelled using BstAr, having checked all relevant ARIMA(p,d,q) models for p, d, q = 0–5. Regarding the forecasting approach based on the BstAr, the latter approach is used to model each of the i–period and j–cohort indexes.

Moreover, in the case of DLR models we utilise a specific class: $Y_t = a + \delta_t \cdot t + \epsilon_t$, where Y_t denotes either the period or cohort dynamics, for each calendar year t, with the slope being a stochastic time variable parameter that follows a smoothed random walk process: $\Delta\delta_t = \phi \cdot \Delta\delta_{t-1} + \zeta_{t-1}$, $0 < \phi \leq 1$ and $\Delta\delta_t = \delta_t - \delta_{t-1}$ denotes the difference operation. The innovations ϵ_t and ζ_t are assumed to be white noise random variables. If $\phi < 1$, Y_t is being modelled as a linear stochastic variable having a slightly tilted s-shape for the short–medium forecasts and also smooth progression to the mortality dynamics. If $\phi = 1$, Y_t is being modelled as a non–linear stochastic variable giving either an accelerating or a decreasing mortality improvement. Experiments with various mortality experiences have shown that the time related non–stationary SPCs can be represented adequately under this particular DLR model structure. Therefore, these two nested DLR structures ($\phi < 1$ and $\phi = 1$) are compared based on the BIC of the mortality model, in order to choose the most appropriate between the two, for each of the period and cohort indexes of each stochastic mortality model. According to Kass & Raftery (1995), one can consider a model selection based on BIC to be roughly equivalent to a model selection based on a Bayes factor. Thus, by applying the two DLR structures to a mortality model, we can have two variations of the same model. Then, we can select the appropriate DLR structure based on the observed BIC difference between these two variations of the model. The difference of the BIC values is defined as, $\Delta\text{BIC} = \text{BIC}(i) - \text{BIC}(b)$, where BIC($b$) denotes the BIC value for the "best" model variation, where the "best" is the one having the lower BIC value and BIC(i) denotes the BIC value for the alternative model variation. In this direction, Kass & Raftery (1995) suggested the following rules of thumb:

- If $\Delta\text{BIC} \leq 2$, there is no clear evidence against or in favour of the two models.
- If $2 < \Delta\text{BIC} \leq 6$, one can say that there is positive evidence against model i, i.e., there is a difference in favour of model b.
- If $6 < \Delta\text{BIC} \leq 10$, there is a strong evidence against model i.
- If $\Delta\text{BIC} > 10$, there is a very strong evidence against model i, and model b is by far the most appropriate.

In the context of this chapter, we examine whether $\Delta\text{BIC} > 2$ in order to judge upon the most appropriate DLR structure for a model.

3.3.3 Evaluation Metrics

Our evaluation incorporates several kinds of quantitative criteria in order to approach the problem of diverse perspectives. For evaluating the fitting performance of each model to historical data, we utilise the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC) and the percentage error tests, Mean Squared Percentage Error (MSPE) and the Mean Absolute Percentage Error (MAPE). Those have been defined in Section 2.4. On top of the above, in order to measure the total amount of the information captured by each model we make use of the Unexplained Variance (UV) and Explanation Ratio (ER) metrics, which are detailed in Sections 3.3.3.1 and 3.3.3.2 respectively.

Moreover, in order to evaluate the models in terms of forecast, we compare the out–of–sample forecasted central mortality rates to the actual ones using the MSPE and the MAPE tests.

Additionally, we aim to provide a final comparative ranking that reflects the overall performance of the models both for the fitting and forecasting modes. To do so, we use the Friedman's statistic test (F–test) Iman & Davenport (1980) and the Bonferroni–Dunn test (BD–test) Dunn (1961).

3.3.3.1 Explanation Ratio (ER)

In order to quantify the total contribution obtained by each individual period or cohort component, for each in–sample dataset, taking into account the whole age range, we

calculate the explained variance of each component including all ages. Particularly, we calculate the sum of the variances of the log crude mortality central rates for all ages, and the sum of the variances of the error between the estimated log mortality rates considering i-components, and the log crude mortality rates, for all ages. Then, one minus the ratio of these two variances defines the ER(i):

$$\mathrm{ER}(i) = 1 - \frac{\sum_x \mathrm{Var}\left(\log(\tilde{m}_{t,x}^{(i)}) - \log(m_{t,x})\right)}{\sum_x \mathrm{Var}\big(\log(m_{t,x})\big)}$$

where, $\tilde{m}_{t,x}^{(i)}$ denotes the estimated mortality rates including i–components for $i = 1, 2, \ldots, p, p+1, \ldots, p+q$, where p and q correspond to the number of period and cohort components respectively. The difference between two successive ER(i) values, i.e., $\Delta\mathrm{ER}(i) = \mathrm{ER}(i) - \mathrm{ER}(i-1)$, for $i = 1, \ldots, p+q$, with $\mathrm{ER}(0) = 0$ gives the magnitude of the contribution by each added component.

3.3.3.2 Unexplained Variance

Unexplained Variance (UV) is used as an additional overall in–sample measure of fit for each particular age x. $\mathrm{UV_x}$ of each age x, is calculated including all age–period–cohort components of a model. $\mathrm{UV_x}$ is calculated using the variance of the time series of log crude mortality rates and the variance of the error between the in–sample estimated log mortality rates and the log crude mortality rates, for each age (Mitchell et al. (2013)). The ratio of these two variances defines the $\mathrm{UV_x}$:

$$\mathrm{UV_x} = \frac{\mathrm{Var}(\mathrm{error}_{t,x})}{\mathrm{Var}(\log(m_{t,x}))}$$

where, $\mathrm{error}_{t,x} = \log(\tilde{m}_{t,x}) - \log(m_{t,x})$ and the variance is taken through time.

It has to be noted that, UV is a different measure than UVR. Specifically, as described in Section 3.2.1.3, UVR is destined to reveal the contribution of each age–period and/or age–cohort component, as well as, to reveal in which age cluster a specific mortality trend of a component belongs to. On the other hand, UV metric provides an overall view of the variance captured by the entirety of a stochastic mortality model, for each age.

3.3.3.3 Comparison of mortality models

To compare the overall fitting and forecast performance of the evaluated mortality models over diverse datasets, we utilise the Friedman's statistic test (F–test) (Friedman (1940)) and the Bonferroni–Dunn test (BD–test) (Dunn (1961)).

F–test uses the null hypothesis H_0: the accuracy is the same among mortality models, against the hypothesis H_1: the accuracy is differed among mortality models. By using F-test, we examine whether there is significant difference in the rank of at least one mortality model. The statistic is:

$$T = \frac{12}{\upsilon \cdot \kappa(\kappa+1)} \cdot \sum_{j=1}^{\kappa} \left(\sum_{i=1}^{\upsilon} R_{ij} \right)^2 - 3 \cdot \upsilon \cdot (\kappa+1)$$

where, υ denotes the number of different datasets, κ denotes the number of different mortality models, and R_{ij} denotes the rank of a relevant Statistical Criterion, for the j–mortality model and i–dataset. As the relevant Statistical Criterion of the F-test, we utilise the Percentage Error tests (MAPE, MSPE) for the forecasting approach, and the Percentage Error tests (MAPE, MSPE) and BIC for the fitting approach.

F-test is a non–parametric analogue of ANOVA with repeated measures, and has approximately a $\mathcal{X}^2$–distribution with $(\kappa-1)$–degrees of freedom, when the null hypothesis is true (Demšar (2006)). Iman & Davenport (1980), showed that Friedman's T is undesirable conservative and derived a better statistic:

$$F = \frac{(\upsilon-1) \cdot T}{\upsilon \cdot (\kappa-1) - T}$$

that has approximately a $\mathcal{F}$–distribution with $(\kappa-1)$ and $(\kappa-1) \cdot (\upsilon-1)$ degrees of freedom.

If the null–hypothesis is rejected, we can proceed with the two tailed Bonferroni–Dunn test (Dunn (1961)), where all mortality models are compared with the "best" model, based on Nemenyi test (Nemenyi (1963)). The performance of two models is significantly different if the corresponding values of average ranks, $\bar{R}_j = \frac{\sum_{i=1}^{\upsilon} R_{ij}}{\upsilon}$, differ by at least the critical difference:

$$CD = q_\alpha \cdot \sqrt{\frac{\kappa \cdot (\kappa+1)}{6 \cdot \upsilon}}$$

where q_a denotes the critical value (see Demšar (2006), Table 5, for q_α values at significant level $\alpha = 5\%$ and $\alpha = 10\%$). In this way, we make $(\kappa - 1)$–comparisons, comparing each model with the "best" model, where "best" model is the one having the best average rank value, $\bar{R}_j$.

3.4 Results

This section details on the datasets, the experimental optimum parameters of HS model and the results of our evaluation and comparative study. Due to the extensive experimental results, and for readability reasons, the reader should refer to the appendices that consolidate the majority of our results.

3.4.1 Mortality data

In this section, we describe the historic mortality data used in our evaluation. We fit five mortality models to data of four different countries. The datasets are provided by the Human-Mortality-Database (HMD) and Eurostat and have the following characteristics:

- Greece (GR), Males, calendar years 1961–2013, individual ages 0–84 (joint dataset from both HMD and Eurostat sources)
- England & Wales (E&W), Males, calendar years 1841–2016, individual ages 0–89 (from HMD)
- France (FR), Males, calendar years 1816–2015, individual ages 0–89 (from HMD)
- Japan (JP), Males, calendar years 1947–2016, individual ages 0–89 (from HMD)

Specifically, in order to evaluate each mortality model in terms of in–sample and out–of–sample performance, we handle the above mentioned mortality data to acquire the datasets described in Table 3.1.

3.4.2 Optimum parameters of the proposed model structure

Table 3.2 presents the parameters used to fit the proposed model structure, for each dataset. More specifically, Table 3.2 contains the number of the optimum parameters of

Country	in–sample	out–of–sample
GR	1961-2013	
	1961-2008	2009-2013
	1961-2003	2004-2013
E&W	1841-2016	
	1841-2006	2007-2016
	1841-1996	1997-2016
	1841-1986	1987-2016
FR	1816-2015	
	1816-2005	2006-2015
	1816-1985	1986-2015
	1816-1965	1966-2015
JP	1947-2016	
	1947-2006	2007-2016
	1947-1996	1997-2016

TABLE 3.1: Datasets used in the evaluation.

the GLM orthonormal polynomials, k_1 (equation (3.1)) and k_2 (equation (3.6)), for the period and cohort effects respectively. Additionally, the optimum s–values under the UVR approach (Section 3.2.1.3), with the corresponding ratio $s/\text{Var}(b_{t_1,0})$–values for each dataset are given, while p and q denotes the optimum number of period (equation (3.5)) and cohort (equation (3.8)) components, according to the approach described in Section 3.2. One can observe that the optimum s–value for each dataset is consistently close to the variance of the GLM's first polynomial parameter $b_{t_1,0}$. This variance corresponds to the overall, age–average, variance of the log–central mortality rates in period effects. The ratio $s/\text{Var}(b_{t1,0})$ gives minimum value 0.78 and maximum value 1.45. Also, we note the high consistency of the optimum parameters among the dataset samples, for each country.

3.4.3 Age-period effects analysis

Figures A.1, A.2, A.4, A.6 and A.10 in Appendix A.1, provide a graphical representation of the 1st age–period component and the corresponding UVR values for all mortality models and datasets. According to the $UVR_x(1)$ values, one can observe that for the most of the models, the 1st component fails to adequately explain the mortality dynamics for the entirety of the age range. Consequently, additional components are needed to capture the rest of the mortality dynamics. Moreover, we can detect similarities among LC, APC and HS structures (Figures A.1, A.4 and A.10). For these models,

Country	years	ages	s	$s/\text{Var}(b_{t_1,0})$	k_1	p	k_2	q
GR	1961-2013	0-84	2.55	0.81	16	4	8	1
	1961-2008	0-84	1.96	0.85	16	3	8	1
	1961-2003	0-84	1.56	0.88	16	3	8	1
E&W	1841-2016	0-89	54.50	1.07	29	5	3	2
	1841-2006	0-89	57.91	1.37	29	5	4	2
	1841-1996	0-89	36.49	1.04	29	5	5	2
	1841-1986	0-89	29.49	1.02	29	5	5	2
FR	1816-2015	0-89	61.42	1.36	33	4	5	2
	1816-2005	0-89	51.22	1.45	33	4	5	2
	1816-1985	0-89	29.11	1.31	33	4	5	2
	1816-1965	0-89	18.71	1.44	33	3	0	0
JP	1947-2016	0-89	33.61	1.16	23	5	7	1
	1947-2006	0-89	19.10	0.78	23	5	4	1
	1947-1996	0-89	22.99	1.12	23	5	4	1

TABLE 3.2: Optimum parameters according to the HS model structure (3.10) for each dataset

the most significant $UVR_x(1)$ values concern the young and the middle ages and there are similarities on the mortality dynamics ($\kappa_t^{(1)}$–values). Furthermore, according to the Explanation Ratio (ER) metric defined in Section 3.3.3.1, as can be seen in Table A.17, their 1st age–period component consistently embrace the majority of the total variance. In contrast, RH and PL models do not present similar characteristics and their 1st component describe different mortality dynamics among different datasets (Figures A.2 and A.6). Specifically, the 1st age–period component of the RH model has significant UVR values at: ages 0–15 for GR, ages 0–18 and 40–60 for E&W, ages 0–20 and 50–80 for FR and ages 0–60 for JP data. Regarding PL model, the 1st age–period component has significant UVR values at ages 0–20 and 60–70 for GR, ages 0–40 for E&W, ages 15–40 for FR and ages 30+ for JP. Also, RH and PL are not able to explain most of the variance in the data as displayed in Table A.17, where for some cases like the E&W and FR datasets, the ER values are negative.

The aforementioned more complicated structures regarding the 1st age–period component of the RH and PL models, i.e., the fact that they explain different mortality dynamics, while they do not capture the maximum amount of variance in the data, are

carried over to the remaining components of the models (either to the remaining age–period or to the age–cohort). Specifically, the 2nd age–period component of PL model shown in Figure A.7, has significant $UVR_x(2)$ values for the ages (50+) for GR, E&W and FR datasets, and for the ages 16–20, for JP dataset. Moreover, the ER values, for E&W and FR experiences, are negative. The 3rd age–period component described in Figure A.8, explains mostly the ages 20–40 for GR, and for the rest countries explains the ages 0–40 (except for ages around 20, and particularly for E&W experience has the biggest ER values).

Based on our results, the HS model reflects several characteristics that have already acknowledged by the mortality literature. Specifically, the first characteristic corresponds to the 2nd HS age–period component (Figure A.11), which explains the relative diversity from the 1st main component for the old ages (mostly for ages 50+). One can observe a significant relative mortality improvement, especially after 1990's (except for JP, where there is a linear trend since the 1970's). Therefore, the 2nd component reflects an acceleration of the rate of improvement for the higher ages, which is aligned to the "ageing of mortality improvement", as has been noted by Horiuchi & Wilmoth (1998), Glei & Horiuchi (2007) and Willets (2004). In terms of explained variance, the ER values range between 1.3% and 11%.

In addition, the 3rd age–period component of HS model (Figure A.12), supports another characteristic which is noted by the existing literature. This component represents relative diversity from the 1st main component for the young–adult ages (mostly for ages 20–35, according to the $UVR_x(3)$ values). Specifically, based on the period component for GR, E&W and FR, one can observe a relative deterioration until 1990's (with one–decade delay for FR, Figure A.12, graph HS.32), whereas later, there is a relative improvement. Particularly for GR dataset, one can observe a even greater improvement after 1990's. Moreover, as reported by Brock & Griffiths (2003), for E&W, the mortality rate for individuals aged from 20 to 35 is increased from the mid–1980's. Crucially, the mortality rates at these ages are quite unstable due to the corresponding small number of deaths. However, as reported in Brock & Griffiths (2003), the causes of the deaths at these ages are mainly due to accidents, suicides and/or drug and alcohol abuse. Even though these deaths are considered "avoidable" and the death rates are relatively low, they contribute considerably to years of life lost. In addition, Brock & Griffiths (2003) note some findings regarding the causes of deaths in this particular age range, which

appear in chronological order. Land transport accidents were the most common cause of deaths during 1960–1980, suicides emerged from the 1990's onwards, and drug–related poisoning mortality tripled from 1979 to 2001. Accordingly for FR, some findings with respect to the causes of deaths at this age group, are mentioned by Meslé (2006). He reports that, one cause that slowed down the improvement in life expectancy (between ages 15 and 30), from 1950's, was the negative trends in violent deaths. In addition, AIDS mortality affected the life expectancy negatively, between ages 25 and 45, for the period 1980–1992, whereas the life expectancy increased because of the decrease in violent deaths in the period 1992–2002. Hence, the 3$^{\text{rd}}$ HS age-period component comes to an agreement with the literature findings regarding the young–adult ages. The corresponding ER values range between 1% and 4%.

An additional characteristic, corresponds to the 4$^{\text{th}}$ age–period component illustrated in Figure A.13. This component represents the relative diversity from the 1$^{\text{st}}$ main component and, according to the $UVR_x(4)$ values, reflects the mortality trend for ages 30–45, for GR, E&W and JP datasets. One can observe a strong deterioration in mortality after 1980's, whereas there is a relative improvement for GR after 1990's, and a steep improvement for JP after 2000's. In addition, according to E&W, we observe a high increase in mortality trend from the mid–1980's. However, as reported by Brock & Griffiths (2003), the causes of deaths at this age group are mainly due to the HIV/AIDS (the impact of AIDS is mostly toward people aged 25–49 (United Nations (2004)) and drugs–alcohol abuse. Concerning E&W, Brock & Griffiths (2003) mentioned that for people aged 35–44, there is a steep increase in mortality rates related to alcohol abuse, from the mid–1980s, having actually twice the value in mortality rates for those aged 40–44. The majority of deaths relates to HIV/AIDS, from 1985 to 2001, appeared in young adults. According to Brock & Griffiths (2003), after introducing more effective treatments in 1996, although the number of newly diagnosed cases had not decreased, mortality rates related to AIDS had decreased dramatically. Furthermore, for JP, Marugame et al. (2005) report that lung cancer is a cause of deaths at ages between 30 and 49, having an increasing trend in mortality after 1958, while decreasing after mid–1970. This can be also observed in the 4$^{\text{th}}$ period component in our results. The related ER values are in most cases less than 1%.

Another significant characteristic is reflected to the 4$^{\text{th}}$ age–period component for the FR data in Figure A.14, which corresponds to age 0, according to the $UVR_x^{(4)}$ values. It

can be seen a relative strong improvement in mortality after 1950's, whereas after 1980's a relative deterioration. According to Meslé (2006), from 1950 to 1980, the contribution of infant mortality to improved life expectancy is high and represents almost half of the improvements. They also refer that this fact is a consequence of the decline in mortality from infectious and respiratory diseases which are mainly congenital infant diseases. The related ER values are around the value of 0.5%

A last characteristic corresponds to the 5^{th} age–period component for E&W and JP mortality experience as shown in Figure A.15. According to the $UVR_x(5)$ values, this component refers to ages 16–21. The 5^{th} period component presents a relative deterioration in mortality until 1980's–1990's, whereas later there is a relative improvement. This range of ages is commonly referred in the actuarial literature as the accident hump effect. Specifically, for E&W, as reported on the Brock & Griffiths (2003), the causes of the deaths at these ages are mainly due to land transport accidents include all motor vehicle and train accidents during the period 1960–1980, and most of these deaths were due to motor vehicle traffic accidents. The corresponding ER values are in most cases less than 0.5%.

3.4.4 Age–cohort effects analysis

Figures A.3, A.5, A.9, A.16 and A.17 provide a graphical representation of the estimated parameters for the 1^{st} age–cohort component of each model. According to the UVR values, APC and PL models explain mostly the very young and old ages jointly. RH's age–cohort component displays the most unstable behaviour and gives significant UVR values: at ages 0–1 and 50+ for GR, at ages 0–1, 20–40 and 50+ for E&W, at ages 20–40 and 80+ for FR and at ages 0–1, 5–10, 20–40 and 60+ for JP experience. The related ER values, for RH model range between 7% and 375%, for APC model between 6.5% and 44% and for PL model between 6% and 93%.

At first, in order to verify the need for incorporating cohort components in the HS model, we examine the deviance residuals of the age–period model structure (equation (3.5)). Figure A.18 illustrates the deviance residuals when those are treated as period effects and as cohort effects. In the former case, the shape of the residuals verifies the appropriateness of the selected age–period components. In the latter case, the noticeable patterns of the graph support the need for incorporating cohort components.

Notably, HS model utilises one cohort for GR and JP datasets and two cohorts for E&W and FR in a consistent manner despite the related low ER values (in most cases they are less than 0.1%). The 1st age–cohort effect illustrated in Figure A.16 explains mostly the ages between 45–75, while the 2nd age–cohort effect in Figure A.17 explains the old ages (75+).

By examining the 1st age–cohort effect, one can observe similar trends between GR and E&W, and between FR and JP. Specifically, for GR and E&W, for the cohorts born from 1900's until 1910's there is a mortality deterioration, from 1920's until 1945's there is a mortality improvement, and from 1945's a mortality deterioration. Therefore, one can observe the so–called "golden cohorts" effect, in which people born within the period 1925 to 1945 experience noteworthy improvements in mortality compared to the nearby decades. These cohort effects have been observed and analysed in several works Renshaw & Haberman (2006), Booth & Tickle (2008) and Cairns et al. (2009, 2011). Especially, Cairns et al. (2011) note that cohorts born around 1930 have experienced strong rates of mortality improvement between ages 40 and 70 in contract to cohorts born 10 years earlier or 10 years later, while the cohort born around 1950 seems to have experienced worse mortality than the immediately preceding cohort. An additional feature revealed by the HS model structure for the same range of ages, refers to people born in 1910's, where it is observed a worse mortality improvement compared to the nearby decades. This feature comes into agreement with the "smoking effect" mentioned in Willets (2004) and Booth & Tickle (2008). Regarding FR and JP, one can observe that for the cohorts born from 1900's until 1920's there is a relative mortality improvement, from 1920's until mid 1930's a relative mortality deterioration, and from mid 1930's until mid 1940's a relative mortality improvement and a mortality deterioration later. These findings are in accordance with Willets (2004), where the mortality trends for JP males have followed a different pattern, showing the existence of two distinctive cohorts: males born in 1910-1920 and those born in 1935-1945. Also, Janssen et al. (2005), reports that the mortality for men declines stagnated among French generations born after 1920.

The 2nd age–cohort effect in Figure A.17 reveals similar patterns for E&W and FR. It can be seen that for the cohorts born before 1900 there is a relative mortality improvement, and starting from 1910 a relative mortality deterioration. This cohort refers to old ages (75+) and describes a second "golden cohort" effect at the end of 19th century for these ages.

It has to be stated that the graphs in Figures A.16 and A.17 have been truncated, excluding some of the initial and the last years, in order to focus exactly at the point of interest. This is because these cohort scores refer mostly to the middle–old or very old ages (according to their UVR values) and therefore the last years have values close to zero, as a result of the estimation method described in Section 3.2.

3.4.5 Comparative analysis

In this section, we present the results of the quantitative analysis used to measure the efficiency of each model for both the fitting and forecasting process.

3.4.5.1 In–sample results

In Appendix A.4, Tables: A.1–A.4 display the in–sample results of the log–likelihood based tests (AIC, BIC) and of the PE tests (MSPE, MAPE), for each country (we display only the datasets with the full range of calendar years). The best scores are highlighted with bold. As can be seen, the HS model outperforms the rest of the models for the majority of the datasets.

With regard to the in–sample performance, the reader can get an overview of the quality of fit of all models based on their average ranking using the F–test (Iman & Davenport (1980)). Moreover, in addition to the F–test, we use the post-hoc BD–test (Dunn (1961)) in order to investigate the significant differences among the models, following the methodology described in Section 3.3.3.3. The results of the comparisons regarding the fitting performance based on the BIC, MSPE and MAPE metrics are given in Table 3.3.

Hence, F–test is performed over 5 mortality models and 14 datasets, for each of the BIC, MSPE and MAPE metrics. F is distributed according to the $\mathcal{F}$–distribution with (4)(52) degrees of freedom. The p–values, of the F–test, bare values of order 10^{-17}, 10^{-27} and 10^{-14}, respectively and therefore, the null hypothesis is rejected for all metrics. Therefore, there is a significant difference in the ranking of at least one mortality model. The CD values of the BD–test are 1.49 and 1.34 at 5% and 10% significant level respectively. According to Dunn (1961), the performance between the HS model (denoted as "best" model according to the average ranking) and each of the LC, APC, and PL models, is significantly different as the corresponding values of average ranks

		MSPE - $R_{i,j}$					MAPE - $R_{i,j}$					BIC - $R_{i,j}$				
		LC	RH	APC	PL	HS	LC	RH	APC	PL	HS	LC	RH	APC	PL	HS
GR	1961-2013	3	1	5	4	2	4	1	5	3	2	4	1	5	2	3
	1961-2008	2	1	4	5	3	4	1	5	3	2	3	1	4	2	5
	1961-2003	2	1	5	4	3	4	1	5	3	2	3	1	4	2	5
E&W	1841-2016	3	2	5	4	1	4	2	5	3	1	4	2	5	3	1
	1841-2006	3	2	5	4	1	3	2	5	4	1	4	2	5	3	1
	1841-1996	3	2	5	4	1	3	2	5	4	1	4	2	5	3	1
	1841-1986	3	2	5	4	1	3	2	5	4	1	4	2	5	3	1
FR	1816-2015	4	2	5	3	1	5	2	4	3	1	4	2	5	3	1
	1816-2005	4	2	5	3	1	4	2	5	3	1	4	2	5	3	1
	1816-1985	5	2	4	3	1	4	2	5	3	1	3	2	5	4	1
	1816-1965	5	2	4	3	1	4	2	5	3	1	4	2	5	3	1
JP	1947-2016	4	2	5	3	1	4	2	5	3	1	4	2	5	3	1
	1947-2006	4	2	5	3	1	4	2	5	3	1	4	1	5	3	2
	1947-1996	3	2	5	4	1	4	2	5	3	1	4	2	5	3	1
$\sum_i R_{ij}$		48	25	67	51	19	54	25	69	45	17	53	24	68	40	25
$\bar{R}_j$		3,4	1.8	4.8	3.6	**1.4**	3.9	1.8	4.9	3.2	**1.2**	3.8	**1.7**	4.9	2.9	1.8
$\left(\sum_i R_{ij}\right)^2$		2,304	625	4,489	2,601	361	2,916	625	4,761	2,025	289	2,809	576	4,624	1,600	625
$\sum_j \left(\sum_i R_{ij}\right)^2$						10,380					10,616					10234
T						44.57					51.31					40.40
F						50.70					142.37					33.67
p–value						$2.5{\cdot}10^{-17}$					$2.4{\cdot}10^{-27}$					$7.3{\cdot}10^{-14}$
$\bar{R}_j - \bar{R}_b$		**2.0**	0.4	**3.4**	**2.3**		**2.6**	0.6	**3.7**	**2.0**		**2.1**		**3.1**	1.1	0.1

TABLE 3.3: In–sample fitting performance and comparison of mortality models, for each dataset, under the MSPE, MAPE and BIC criteria, with CD = 1.49 and CD = 1.34 using the critical values $q_{0.05} = 2.498$ and $q_{0.10} = 2.241$ at significant level $\alpha = 5\%$ and $\alpha = 10\%$ respectively. $\bar{R}_b$ denotes the average ranking of the "best" model.

differ by at least the CD values for the MPSE and MAPE metrics. Between the HS and RH there is no significant difference in their performance. Regarding the BIC criterion, the performance of the RH model is significantly different, only compared to the LC and APC models. In fact, the APC has the lowest performance under all the metrics. Consequently, based on the MSPE and MAPE metrics, the HS model has the best ranking and its overall performance differs significantly from LC, APC and PL models. Based on the BIC criterion, the RH and HS models have almost identical performance and differ significantly from the LC, APC and PL models.

Additionally, in Appendix A.3, Figure A.19, complementary results for the fitting process are presented by the Unexplained Variance (UV) graphs, for each model and for each country, using the dataset samples containing the entirety calendar years, according to the methodology described in Section 3.3.3.2. One can observe that, the HS model significantly outperforms the rest of the models in all countries, except for the GR, where all models are relatively poor in explaining the fluctuations of young adult ages. We should stress that in Table 3.4, depicts the percentage of ages where the HS model has

	LC	RH	APC	PL
GR	96%	49%	72%	60%
E&W	100%	100%	100%	100%
FR	100%	100%	100%	100%
JP	100%	81%	100%	99%

TABLE 3.4: Percentage of ages where HS model achieves better Unexplained Variance values.

better UV with respect to the rest of the models. For example, for GR data, the HS model has better UV values for the 72% of the ages compared to the APC.

3.4.5.2 Out–of–sample results

As far as the forecasting analysis is concerned, we deploy the backtesting framework described in Section 3.3 in order to compute the forecasted mortality rates. The ϕ parameter of the best DLR nested models, as described in Section 3.3.2, according to the Kass & Raftery (1995) rules, and the optimum parameters of the selected ARIMA(p,d,q) models, based on minimum BIC values, for each stochastic mortality model and for each dataset are shown in Table A.16 and Table A.15, respectively.

Tables A.5–A.14 in A.5 provide a detailed an overview for the efficiency of the models by giving the out–of–sample results for the PE tests, for each dataset and each forecast model. By taking a look over the average MSPE and MAPE scores of the three forecasting models, one can observe that the our model, in the vast majority of the datasets, outperforms the rest of the models.

In addition, similar to the in–sample comparison, we compare the models based on their average ranking using the F–test and the post-hoc BD–test. Table 3.5 provides the results of the comparisons based on the average MSPE and average MAPE metrics given in A.5–A.14. Hence, F–test is performed over 5 mortality models and 10 datasets. The value F is distributed according to the $\mathcal{F}$–distribution with (4)(36) degrees of freedom. The p–values, of the F–test, bare values of order 10^{-6} and 10^{-4}, respectively. That is, the null hypothesis is rejected for all metrics, giving a significant difference of at least one model. Regarding the BD–test, the CD values are 1.77 and 1.58 at 5% and 10% significant level respectively. Therefore, the pair–wise comparison between the best

		AVG MSPE - $R_{i,j}$					AVG MAPE - $R_{i,j}$				
		LC	RH	APC	PL	HS	LC	RH	APC	PL	HS
GR	1961-2008	4	2	1	5	3	5	3	4	1	2
	1961-2003	5	1	4	3	2	4	2	5	3	1
E&W	1841-2006	2	4	3	5	1	4	5	2	3	1
	1841-1996	2	3	4	5	1	3	5	2	4	1
	1841-1986	2	3	4	5	1	2	5	3	4	1
FR	1816-2005	3	2	4	5	1	3	2	4	5	1
	1816-1985	3	2	4	5	1	3	4	2	5	1
	1816-1965	3	2	5	4	1	3	2	4	5	1
JP	1947-2006	4	2	5	3	1	4	2	5	3	1
	1947-1996	2	5	4	3	1	3	5	4	2	1
$\sum_i R_{ij}$		30	26	38	43	13	34	35	35	35	11
$\bar{R}_j$		3.0	2.6	3.8	4.3	**1.3**	3.4	3.5	3.5	3.5	**1.1**
$\left(\sum_i R_{ij}\right)^2$		900	676	1,444	1,849	169	1,156	1,225	1,225	1,225	121
$\sum_j \left(\sum_i R_{ij}\right)^2$						5038					4,952
T						21.52					18.08
F						10.48					7.42
p–value						$9.8{\cdot}10^{-6}$					$1.8{\cdot}10^{-4}$
$\bar{R}_j - \bar{R}_b$		**1.7**	1.3	**2.5**	**3**		**2.1**	**2.2**	**2.2**	**2.2**	

TABLE 3.5: Forecast performance and comparison of mortality models, for each dataset, under the MSPE and MAPE criteria, with CD = 1,77 and CD = 1,58 using the critical values $q_{0.05} = 2.498$ and $q_{0.10} = 2.241$ at significant level $\alpha = 5\%$ and $\alpha = 10\%$ respectively. $\bar{R}_b$ denotes the average ranking of the "best" model.

model, which is the HS model, and each of the rest, gives significant CD values based on the MAPE metric, i.e., the performance of our model is significantly different to the other models according to the MAPE criterion. Regarding the MSPE metric, the performance of the proposed model is significantly different, except for the cases of the LC model (only for $\alpha = 5\%$) and the RH model. Also we observe that, PL model has the lowest performance ranking. Consequently, the HS model has the best ranking in both MSPE and MAPE criteria, and based on the MSPE metric the HS model is significant better from LC, APC and PL models (at 10% significant level).

All in all, our results advocate that the proposed model outperforms the other models, especially in forecast terms. Notably, although the performance of the RH is not significantly different from our model based on the fitting performance, the forecasting ability of our model is significantly different under the MAPE metric. One can observe that LC and APC models perform better in forecast rather than in fitting, while the opposite behaviour is observed for PL and RH models.

3.4.6 Comparative analysis using shorter datasets

In this section, we present a comparative analysis using datasets of shorter periods. Overall, by observing the results in subsection 3.4.5, the proposed model outperforms the LC, RH, APC and PL models, both in terms of in–sample and out–of–sample performance for most of the datasets. However, HS model lacks slightly when it comes to the dataset of Greece. It is noticeable that the Greek dataset is the one having the shorter period (starting in 1961), which contrasts to the rather long periods for E&W and France (starting in 1841 and 1816). Over the past two centuries different dynamics have driven the change in mortality. Therefore, one could say that is reasonable the HS model, which deploys multiple age-period and age-cohort effects, to outperform the other more limited models. Motivated by this observation, in order to verify whether the better performance of the HS model is a result of the extremely long fitting periods or not, we conduct a thorough evaluation over shorter datasets, while we compare the fitting performance of the proposed model with respect to the other mortality models.

By following the same rationale as in the previous subsection, we evaluate the fitting performance of each mortality model to historical data of shorter period, utilising the most significant performance metrics and we present the results of the quantitative analysis. Table 3.6 presents the characteristics of each dataset and the parameters used to fit the HS model structure. Note that, all the model are fitted on the same datasets with the aim of providing an unbiased quantitative analysis and to investigate further the behaviour of HS model in shorter datasets.

Country	years	ages	s	$s/\mathrm{Var}(b_{t_1,0})$	k_1	p	k_2	q
GR	1961-2013	0-84	2.55	0.81	16	4	8	1
E&W	1961-2016	0-89	11.41	1.32	38	4	4	1
FR	1961-2015	0-89	11.81	1.17	23	4	5	1
JP	1961-2016	0-89	19.71	1.41	23	5	7	1

TABLE 3.6: Optimum parameters according to the HS model structure (3.10) for datasets of shorter periods

By observing Table 3.6, one could notice that the optimal s–values have been affected compared to those presented in Table 3.2. This is a justified outcome, as the optimality of the s–value, as described in subsection 3.2.1.3, is related to the magnitude and the variance of the data.

In Appendix A.4, Table A.1 and A.8, Tables: A.18–A.20 present the in–sample results of the log–likelihood-based tests (AIC, BIC) and the PE tests (MSPE, MAPE), for each dataset. The best scores are highlighted in bold.

Regarding the PE tests, the HS model outperforms the rest of the models, whereas regarding the Information Criteria, the RH have better performance, for all datasets. In addition, with regard to the fitting performance, the reader can have an overview of the quality of fit of all models based on their average ranking using the F–test (Iman & Davenport (1980)). Moreover, we use the post-hoc BD–test (Dunn (1961)) in order to investigate the significant differences among the models, following the methodology described in Section 3.3.3.3. The results of the comparisons based on the BIC, MSPE and MAPE metrics are given in Table 3.7.

		MSPE - $R_{i,j}$					MAPE - $R_{i,j}$					BIC - $R_{i,j}$				
		LC	RH	APC	PL	HS	LC	RH	APC	PL	HS	LC	RH	APC	PL	HS
GR	1961-2013	3	1	5	4	2	4	1	5	3	2	4	1	5	2	3
E&W	1961-2016	3	2	5	4	1	4	2	5	3	1	5	1	4	2	3
FR	1961-2015	5	2	4	3	1	5	2	4	3	1	5	1	4	3	2
JP	1961-2016	4	2	5	3	1	4	2	5	3	1	4	1	5	3	2
$\sum_i R_{ij}$		15	7	19	14	5	17	7	19	12	5	18	4	18	10	10
$\bar{R}_j$		3.8	1.8	4.8	3.5	**1.3**	4.3	1.8	4.8	3.0	**1.3**	4.5	**1**	4.5	2.5	2.5
$\left(\sum_i R_{ij}\right)^2$		225	49	361	196	25	289	49	361	144	25	324	16	324	100	100
$\sum_j \left(\sum_i R_{ij}\right)^2$						856					868					864
T						13.6					14.8					14.4
F						17					37					27
p–value						$6.9{\cdot}10^{-05}$					$1.2{\cdot}10^{-06}$					$6.4{\cdot}10^{-06}$
$\bar{R}_j - \bar{R}_b$		2.5	0.5	**3.5**	2.3		**3.0**	0.5	**3.5**	1.8		**3.5**		**3.5**	1.5	1.5

TABLE 3.7: Fitting performance and comparison of mortality models, for each shorter dataset, under the MSPE, MAPE and BIC criteria, with CD = 2.79 and CD = 2.51 using the critical values $q_{0.05} = 2.498$ and $q_{0.10} = 2.241$ at significant level $\alpha = 5\%$ and $\alpha = 10\%$ respectively. $\bar{R}_b$ denotes the average ranking of the "best" model.

F–test is applied over 5 mortality models and 4 datasets, for each of the BIC, MSPE and MAPE metrics. F is distributed according to the $\mathcal{F}$–distribution with (4)(12) degrees of freedom. The p–values, of the F–test, bare values of order 10^{-05}, 10^{-06} and 10^{-06}, respectively and therefore, the null hypothesis is rejected for all metrics. Hence, there is a significant difference in the ranking of at least one mortality model. The CD values of the BD–test are 2.79 and 2.51 at 5% and 10% significant level respectively. According to Dunn (1961), the performance only between the HS model (denoted as "best" model according to the average ranking) and the APC model is significantly different for the MSPE. The performance between the HS model and each of the LC and APC models

is significantly different for the MAPE metric, as the corresponding values of average ranks differ by at least the CD values. Among the HS, RH and PL, there is no significant difference in their performance.

Regarding the BIC criterion, the performance of the RH model is significantly different, only compared to the LC and APC models. In fact, the APC has the lowest performance under all the metrics. Consequently, based on the MSPE and MAPE metrics, the HS model has the best ranking and its overall performance differs significantly from LC and APC models. Based on the BIC criterion, the RH model differs significantly only from the LC and APC models.

Additionally, in order to measure the total amount of the information captured by each model we make use of the Explanation Ratio (ER) metric. According to the ER metric defined in Section 3.3.3.1, as can be seen in Table A.21, the HS model consistently captures the most amount of variance compared to the rest of the models.

Consequently, according to the evaluation results for the shorter datasets, one can observe that the HS model consistently outperforms the other models in terms of the MSPE and MAPE metrics, even if there is no significant difference from the performance of the RH and PL models. Additionally, HS captures the most amount of the variance of the data. It can also be observed that RH model performs better from the rest of the models regarding the BIC, whereas there is no significant difference from the performance of the HS and PL models, even though the latter are multiple component models and incorporate more parameters.

Additionally, during our experiments on shorter datasets, the HS model was still able to designate an identified mortality trend to a unique age cluster, as it is the case for long period data. However, we opt not to include the graphs of the parameter estimates and the UVR values in order to keep the evaluation within reasonable limits.

Based on our experimental results, it can be safely argued that the HS model outperforms the rest of the models, but lacks slightly in BIC due to the incorporation of more variables. Moreover, the RH model ranked second, whereas the PL model follows next. The LC and APC models come at the last positions of the ranking. Overall, based on the results of both short and long period datasets, strong evidences are provided to support the beneficial characteristics of the HS model.

3.4.7 Notable differences of Hatzopoulos and Haberman [2011]

As stated in Section 3.2, our proposal adopts two fundamental methods which were first employed in Hatzopoulos & Haberman (2011) (HH), namely SPCA and GLM. Even though our proposal shares this common ground with HH model structure, it presents major improvements and differences.

Specifically, a main relative disadvantage in HH estimation method is that N bootstrap samples, of normally distributed GLM–estimated parameters, are calculated in order to choose the "optimum" s–value in the SPCA. This process is repeated for each candidate sparsity factor leading to a quite complex estimation method. Instead, our work introduces a new method based on the Unexplained Variance Ratio (UVR) metric to pinpoint the "optimum" s–value among the candidates, while according to our experiments the set of the candidate values is located in the area around the variance of the $b_{t_1,0}$ value. Our approach downgrades the computational complexity of the estimation method by N times.

The aforementioned difference is reflected in the models' results as well. Due to the different estimation method, the models capture different age clusters that correspond to the age–period effects. For the same data (E&W, males, 1841–2006), HS derives 5 age–period components (Table 3.2 and Figures A.10–A.15), while Hatzopoulos & Haberman (2011) presented 7. One can observe that the first 5 age–period components of our model are relative similar to the 1^{st}, 2^{nd}, 4^{th}, 6^{th} and 3^{rd} age–period components of the HH, respectively. However, even if these components seem to be similar there are notable peculiarities among them. The 6^{th} period component of the HH model identifies two different age clusters (ages 31–41 and ages 5–7) in contrast to the HS model, which has the ability to reveal unique age clusters for each period component. Regarding the 5^{th} and 7^{th} age–period components of the HH and according to the bootstrap approach, the former refers to ages 10–14 and age 1, while the latter refers to ages 42–48 and age 0. However, as has been discussed in Hatzopoulos & Haberman (2011), the ages 0 and 1 exhibit low dependence with respect to other ages, and therefore these two components combine different mortality dynamics.

Therefore, by using the UVR, the estimation method of the HS model is driven to maximise the captured information in a more effective way. In addition, due to the condition of finding those s–values that deduce to unique age clusters, our estimation

method converges to more informative and efficient stochastic components than the HH model. Hence, HS model is able to designate a unique mortality trend to a specific cluster of ages.

Additionally, a notable difference can be found in the construction of the two models. According to the Hatzopoulos & Haberman (2011) model structure (equation 3a), the cohort effects are estimated by combining both observed and forecasted estimations, after obtaining forecast values in period effects (Figure 1, triangle BCG in Hatzopoulos & Haberman (2011)). This means that the final estimated log–graduated mortality rates are based, not only on the observed, but also on the forecasted mortality rates. In fact, according to age–period–cohort expanded model structure given in equation 3a, the HH model fits the central mortality rates in cohort effects, in parallelogram DCGF in Figure 1. Thus, for $t = t_1, \ldots, t_n$ and $x = x_1, \ldots, x_a$, one can observe that the mortality rates residing in triangle ADF are not being estimated in terms of cohort, but are only estimated by the age–period model structure. Hence, the final estimates omit the cohorts from $t_1 - x_2$ to $t_1 - x_a$. On the contrary, as has been described in Section 3.2.1.2, the estimation method proposed in this work estimates the age–cohort components by using simple GLM polynomial structures without applying any forecast method in period effects. Therefore, the final log–graduated mortality rates are estimated based only on the observed mortality rates defined by parallelogram BCDE in Figure 3.1. The aforementioned characteristics reveal a different model structure between the HH and HS model.

Comparing the age–cohort components of Figures A.16–A.17 and Figure 8 in Hatzopoulos & Haberman (2011), it is obvious that the HS model produces distinct and more interpretable mortality dynamics for particular age ranges. On the downside, the HH model ends up with 3 age–cohort components with quite complicated structure, explaining unrelated range of ages. This happens mainly for two reasons. On the one hand, the estimation method of HH is based on bootstrap for defining the optimal s–value. This kind of selection of the s–value leads to a more indistinct determination of the significant age–period components and, in this way, these defects are carried over to the age–cohort effects. In particular, the inclusion of the 7^{th} age–period component leads the cohort structure of HH to break down, while in the HS model the ages 40–50 are part of the significant ages explained by the 1^{st} age–cohort component (Figure A.16). On the other hand, the fact that the more complicated age–cohort effects of HH are derived out of

both forecasted and observed estimations, can affect the final results in an ambiguous way.

Overall, the proposed method provides a remarkably faster and simpler approach for estimating the mortality rates, and thus more friendly for experimental results. Additionally, our model captures the relevant significant interaction components in a more efficiently and distinct manner.

3.5 Discussion

This Section provides an overall discussion upon the findings of the mortality models' evaluation, while it summarises the contributions of our work with respect to our results. With reference to the Tables 3.3–3.5, and those residing in Appendix A.4 and A.5, we can derive an overall performance ranking for the examined models.

The LC model is the simplest one having the lowest number of parameters, but the absence of additional age–period and age–cohort effects is reflected in its in–sample results. Although the LC model scores one of the lowest in–sample performances (4^{th} ranking among all criteria), it performs much better in forecast terms (2^{nd} and 3^{rd} ranking in MSPE and MAPE criteria respectively).

The APC model, although it is simple and has distinct and consistent components (Figures A.4–A.5), achieves the lowest scores, both in fitting and forecasting (5^{th} ranking for the fitting and 4^{th} or 5^{th} ranking for the forecasting performance). It has to be stated that, the APC model is the only one that separates the age, period and cohort effects independently, utilising a constant age effects in the period and cohort components.

Therefore, although LC and APC models lack of certain desirable characteristics, particularly in terms of fitting performance, the simple structure and the consistent period or cohort trends (Figures A.1 and A.4–A.5) enable them to score better forecast performance.

The RH model performs relatively well for the in–sample data (1^{st} or 2^{nd} best scores), but in terms of forecast achieves relative poor results (2^{nd} and 4^{th} ranking). In fact, RH model achieves better results from LC, APC and PL models for most of the tests. However, RH model can be questioned for its robustness. Figures A.2 and A.3 reveal its

unstable and complicated age–cohort structure, which can justify to an certain extend the worse forecast performance.

Regarding the PL model, although in the in–sample tests it has an average performance (it ranks 3$^{\text{rd}}$), in terms of forecast it has the lowest ranking. Also, according to the UVR values (Figures A.6–A.9), one could say that, for each additional component, the PL model cannot define mortality dynamics which are attributed to a distinct and consistent age range. In addition, the corresponding mortality trends κ_t's and γ_c's scores, are rather complicated and difficult to interpret. This feature affects negatively to a certain extent its forecast performance.

According to the above, it can be noted that the LC and the APC models have the inverse behaviour with respect to the PL and the RH model, regarding the fitting and forecasting performance. This finding is aligned to the conclusion of Cairns et al. (2011) who claim that a good fit to historical data does not guarantee sensible forecasts. Specifically, the LC and APC achieve better performance in forecast terms, mainly due to their simpler structure and to their robust behaviour, while the PL and RH perform worse in forecast, due to lack of robustness, consistency and interpretability of the relevant components.

The RH model is known to have important stability and robustness issues explained by some identifiability problems as reported by Hunt & Villegas (2015). In fact, these issues could explain to an extent the unstable UVR values, negative ER values, poor forecasting performance and non-robust parameter estimates. Notably, Hunt & Villegas (2015) proposed an approach that can resolve some to the issues of the RH model. This approach is implemented by Villegas et al. (2018) using the R package StMoMo. However, for consistency reasons we opt to include in our comparative analysis the initial model as provided by Lifemetrics tool.

The HS model has an excellent performance and ranks first among the compared models, both in terms of fitting and forecasting. Notably, the HS model gives significant differences compared to the remaining models, based on MAPE metric, both in terms of fitting and forecasting performance, while based on MSPE metric, HS model is significant better than LC, APC and PL.

As a general outcome of the above comparisons, it can be noted that the multiple component models, which includes interaction age effects in period and cohort components,

give a more accurate mortality overview concerning both the fitting and the forecasting performance.

We introduce a novel estimation method for constructing the age–period and age–cohort effects, which is driven by the UVR criterion. In particular, a fast and relative simple approach is described in Section 3.2 with the aim of selecting the optimum sparse factor, s–value. In this way, the optimum s–value reveals distinct age–period and age–cohort dynamics. Therefore, our estimation method leverages the SPCA in conjunction with UVR criterion, in order to choose the relative components, as well as to pinpoint the exact age clusters that significantly contribute to particular mortality dynamics.

It has to be stated that, for all the datasets, the proposed model is consistently composed of three significant common age–period effects and one common age–cohort effect. The first common age–period component explains mostly the young and the middle ages (Figure A.10), which captures the vast majority of the explained variance (Table A.17) since those ages historically account for most of the mortality improvements. The second age–period component (Figure A.11) describes the "ageing of mortality improvement", as it reflects an acceleration of the rate of mortality improvement at higher ages. This evident segregation between the young and the old ages designates the mortality dynamics more precisely, and more importantly, results to an improved forecast performance. The third common age–period component (Figure A.12), describes the young–adult ages, where most of the deaths are due to external causes and drugs. Although the mortality rates at these ages are relatively low, they contribute considerably to years of life lost. Furthermore, significant supplementary age–period components are employed based on a dataset's peculiarities. These components describe, either the ages 30–45 in which deaths are mostly connected to the HIV/AIDS and drugs (Figure A.13 for GR, E&W and JP datasets), or the infant diseases (Figure A.14 for FR dataset), or the "accident hump" effect (Figure A.15 for E&W and JP datasets). The common age–cohort component (Figure A.16) refers to middle–old ages, and the findings are in accordance with the relative literature ("golden cohort" and "smoking" effect) as discussed in Section 3.4. In addition, for datasets with long mortality history (EW and FR), our model gives another "golden cohort" effect for the very–old ages, at the end of the 19^{th} century (Figure A.17). In fact, the necessity of incorporating cohort effect is supported by the deviance residual plots in A.2, where the graphs show a clear cohort pattern. Although,

the corresponding ER values are very low, these cohort components facilitates the description of well–known mortality characteristic ("golden cohort", "smoking effects") in a robust manner.

As stated above, due to our model structure and the utilised estimation method, the age-cohort effects of the HS model explain mostly the old ages and produce low ER values (Table A.17), as most of the variance of the mortality data is captured by the age–period components. During our experimental study, we noticed that the age–cohort components do not have any effect on the out–of sample performance, and for this reason, forecasting the cohort effects, γ_c indexes, is unnecessary. Therefore, the associated forecast models of the cohort effect in Tables A.15 and A.16 are omitted, as the log–graduated forecasted mortality rates are estimated under the age-period model structure.

Our results are aligned to well-established findings of the mortality literature and this advocates the beneficial characteristics of the proposed model. The utilisation of the SPCA in association with the UVR-driven estimation method produce interpretable distinct mortality dynamics for particular age ranges. Furthermore, it can be observed from Figures A.10–A.17 that our model is robust, as it describes different mortality datasets with high consistency and accuracy, both in terms of fit and forecast.

We may conclude that, the new estimation method results to a model that satisfies all the desirable characteristics considered by Cairns et al. (2009). In comparison to the Hatzopoulos & Haberman (2011), our model embraces the two fundamental elements of SPCA and GLM, but despite this common ground the newly introduced model brings in several novelties that boost the capabilities of our model in terms of computational complexity, simplicity, efficiency and interpretability.

Chapter 4

Extensions on the Hatzopoulos–Sagianou Multiple-Components Stochastic Mortality Model

In this chapter, we present extensions of the Hatzopoulos–Sagianou (2020) (HS) multiple-component stochastic mortality model which presented in the previous chapter of this doctoral thesis. Our aim is to thoroughly evaluate and stress test the HS model by deploying various link functions, using generalised linear models, and diverse distributions in the model's estimation method. In this chapter, we differentiate the HS approach by modelling the number of deaths using the Binomial model commonly employed in the literature of mortality modelling. Given this, new HS extensions are derived using the off-the-shelf link functions, namely the complementary log–log, logit and probit, while we also reform the model by introducing a new form of link functions with a particular focus on the use of heavy-tailed distributions. The above-mentioned enhancements conclude to a new methodology for the HS model, and we prove that it is more suitable than those used in the literature to model the mortality dynamics. In this regard, this chapter offers an extensive experimental testbed to scrutinise the efficiency, explainability and capacity of the HS extensions using both the off-the-shelf and the newly introduced form

of link functions over datasets with different characteristics. The introduced HS extensions bring an improvement by approximately 16% to the model's *goodness-of-fit,* while they uncover more fine-grained age clusters. In addition, we compare the performance of the HS extensions against other well-known mortality models, both under fitting and forecast modes. The results reflect the advantageous features of the HS extensions to deliver a highly informative structure and enable the attribution of an identified mortality trend to a unique age cluster. The above-mentioned improvements enable mortality analysts to perform an in-depth and more detailed investigation of mortality trends for specific age clusters and can contribute to the attempts of academia and industry to tackle the uncertainties and risks introduced by the increasing life expectancy.

4.1 Introduction

Regardless of the peculiarities of each mortality model implementation, the aim is to analyse mortality by decomposing the mortality rates in the dimensions of age, period and cohort (or year of birth). As described in Section 2.1, a generic formula (see Equation (2.1)) is defined and can represent the majority of stochastic mortality models in line with the principles of generalised linear models (GLMs) Hunt & Blake (2021), Currie (2016), Villegas et al. (2018). The complete documentation of the principles, the theory and application of GLMs can be found in McCullagh & Nelder (1989).

In short, in the context of GLMs, the results of the dependent variables are considered to originate from certain distributions belonging to the exponential family. This is a wide range of probability distributions, such as the Binomial, Poisson, Gamma and Normal distributions. The GLMs are a generalisation of the classical linear models. More specifically, the GLMs generalise linear regression by allowing the linear model to connect to the response variable using a link function. Thus, the latter enables the connection of the linear predictor and the mean of the distribution function. In general, various link functions can be used, but as is the case in the GLM literature, the so-called canonical link functions are the prominent ones used in the mortality literature by stochastic mortality models. The canonical link functions are the log and the logit for the Poisson and Binomial distributions, respectively. Despite the canonical ones, the complementary log–log and probit link functions have also been used for the Binomial distribution.

In the context of GLMs, some requirements must be taken into account in order to select the proper link function. The same applies for mortality models. Early attempts in the mortality modelling field consider as a requirement that the data should be transformed in order to obtain an approximately linear predictor structure Hunt & Blake (2021). To achieve this, one has a wide range of options of link functions to utilise. Notably, there are several studies in the wild that use and experiment on various link functions in order to more accurately model either mortality rates or death probabilities and even to extend already known or introduce new flavours of stochastic mortality models Haberman & Renshaw (1996), Currie (2016). In this line of thought, StMoMo gives the ability to parameterise the models' estimation process. However, it limits the choices only to the use of the canonical link functions, while GLM-based estimation could be performed under various link function options and there is no a priori reason why other links should not work well. Motivated by this fact, our work aims to explore and expand the range of choices of link functions used in the context of mortality modelling. To do so, we conduct our investigation using the HS model and we introduce a new form of link functions that extend the model and its capacity by offering new HS flavours.

In view of the above, this chapter extends the stochastic mortality model of the HS, which was introduced in Chapter 3 using central mortality rates, $m_{t,x}$, under the Poisson distribution and log link function, by formulating it in terms of conditional probabilities of death, $q_{t,x}$, and the Binomial distribution using a wide set of link functions. In fact, the extension of the HS model is twofold. On the one hand, we adapt the HS to be modelled using off-the-shelf, i.e., well-known, link functions in the literature, namely the logit, complementary log–log and probit. This approach is in line with the rationale used in tools such as StMoMo for ratifying models both in the Poisson and the Binomial cases, while it calls forth the HS model to prove its capacity on a new experimental testbed. On the other hand, we introduce a new form of link functions in an effort to further improve the *goodness-of-fit* and the explainability of the model, and we evaluate their applicability in the context of mortality through the HS extensions. A particular focus is given to the use of heavy-tailed distributions. From a modelling perspective, heavy-tailed distributions are important when extreme events must be part of the model. For example, although the probability of shock cases, such as mortality during World War II or the COVID-19 pandemic, is small, the magnitude of the impact is so large that these events are vital in capturing the mortality dynamics in an accurate manner,

i.e., to ensure that the high variability due to the extreme events (e.g., κ_t^3 in Figure B.1) can be explained by the model.

To the best of our knowledge, it is the first time that those link functions are tested in the mortality domain. Therefore, we build a multidimensional testbed, and we showcase that the HS retains its robust characteristics, i.e., (i) the identification of the significant incorporated period and cohort effects, (ii) the identification of distinct stochastic components and (iii) the attribution of an identified mortality trend to a unique age cluster. We assess the capabilities of the HS extensions using the data of males from England and Wales and also Greece provided by the Human-Mortality-Database.

In short, the contributions of the work presented in this chapter are as follows:

- We extend the stochastic mortality model HS formulated in terms of $q_{t,x}$, using generalised linear models and by adopting various link functions. We illustrate through experimental results that the HS model remains robust and consistent under all modelling variations.
- We introduce a new set of link functions, with a particular focus on heavy-tailed distributions, and we evaluate their applicability in the context of mortality through the HS extensions. This approach leads to the definition of a new estimation methodology for the HS model. To the best of our knowledge, it is the first time that those link functions are evaluated in the mortality modelling domain.
- We compare the efficiency of the new model extensions versus the established mortality models in fitting and forecasting modes. For the latter case, we use an out-of-sample approach to assess the prediction ability of each model by using the Random Walk with Drift (RWD) model and optimum ARIMA models based on the Bayesian Information Criterion (BIC) test.
- We highlight the lessons learnt to inform the community about the adoption of the various link functions in the models' estimation methods having witnessed the beneficial impact of this approach to our model's efficacy.

The rest of this chapter is organised as follows: In Section 4.2, we present the newly introduced methodology and the updated estimation method for the HS extensions using the off-the-shelf and the new form of link functions. Section 4.3 outlines the evaluation

approach and the performance results of the HS extensions. Section 4.4 provides an overall discussion on the findings and concludes the chapter. In addition, Appendix B offers technical details and the graphical representations of the generated results.

The reader may refer to Section 2.1 for the notation used and to Section 2.2 for background information for the GLM framework.

4.1.1 A summary of the Hatzopoulos–Sagianou (HS) Model methodology

Before we proceed to the documentation of the necessary adaptations of the model in order to introduce a new form of link functions and test their applicability to the HS in Section 4.2, we offer a synopsis of the estimation method of the HS model. For more details, the reader can refer to Chapter 3.

HS is a dynamic multiple-component model that includes p age–period and q age–cohort non-predefined effects. The HS model has the following form:

$$\log(m_{t,x}) = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{c(j)} \gamma_c^{(j)}$$

The number of age–period and age–cohort effects vary depending on the experimental dataset, i.e., the intrinsic mortality peculiarities of the examined population for a given time frame. The estimation methodology for defining the most important p age–period and q age-–cohort components is briefly described below. The model structuring method is unfolded in four steps, each of which utilises a method which brings to the model its competitive advantages as will be described later.

Step #1: GLM is applied, with the number of deaths being the response variable, $D_{t,x}$, for each calendar year t independently, by using the log link function, i.e., the canonical link function for the Poisson distribution, and we treat the logarithmic exposure variables (i.e., $\log(E_{t,x}^c)$) as an offset. The predictor structure is:

$$\eta_{t,x} = \log\left(\mathbb{E}\left(D_{t,x}\right)\right) = \log(E_{t,x}^c \times m_{t,x}) = \log(E_{t,x}^c) + \log(m_{t,x})$$

Therefore, treating age as an explanatory variable, the linear predictor is produced using an orthonormal polynomial structure in age effects for each calendar year independently. That is, the following predictor structure is defined as:

$$\log(\hat{m}_t) = b_t \cdot L^T \tag{4.1}$$

where, b_t corresponds to the vector of the GLM-estimated parameters for calendar year t. Given this, a random (asymptotically normal) matrix $B = \{b_{t,k-1}\}$, $k = 1, \ldots, k_1$, of order $(n \times k_1)$, includes the GLM-estimated parameters for all calendar years $t = t_1, \ldots, t_n$. $L = \{l_{x,k-1}\}$, $k = 1, \ldots, k_1$, is a (a $\times$ k_1)-dimensional matrix of orthonormal polynomials with degree of $k_1 - 1$ for age x.

Step #2: This step aims to keep in the model only the factors explaining most of the data's information in order to decrease the dimensionality of the problem. To do so, SPCA is applied to the covariance matrix of the GLM-estimated parameters B. Given the covariance matrix, $A = \text{Cov}(B)$, as stated in Luss & d'Aspremont (2006), the (dual) problem of defining a sparse factor, which will allow to capture the maximum amount of data's variance, can be formulated as follows:

$$\min \lambda^{\max}(A + U), \quad \text{s.t. } |U_{ij}| \leq s \tag{4.2}$$

where s is a scalar that defines the sparsity. The definition of the "optimum" s value and, consequently, the most important p age–period (and corresponding q age–cohort components) is achieved using the unexplained variance ratio (UVR) approach.

Step #3: Hatzopoulos & Sagianou (2020) introduced a heuristic methodology based on UVR metric to define the optimal model structure, i.e., to incorporate the most important (i.e., informative) age–period and age–cohort components. This methodology is described in detail in Section 3.2.1.3. It must be stated that the optimal model structure is achieved through the process of defining the optimal s value. In other words, the definition of the optimal model structure coincides with the definition of the optimal scalar s, which in turn controls the sparsity in SPCA. This process is driven by the UVR metric in order to converge to components that maximise the captured variance of the mortality data and to acquire distinct and significant stochastic components, which enable the attribution of an identified mortality trend to a unique age cluster. The

aim is to decrease the number of age–period components from k_1 to p. The excluded components are then treated as residuals.

Overall, the HS estimation methodology takes advantage of the SPCA and the UVR criterion, not only for pinpointing the informative p components but also for uncovering the age clusters that significantly contribute to the mortality trend of each component. This approach results in the following model:

$$\log(\hat{m}_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \varepsilon_{t,x} \tag{4.3}$$

where the residuals $\varepsilon_{t,x}$ at age x in year t denote the deviation of the model as a result of the excluded sparse principal components (SPCs).

The estimation process for the age–cohort components follows a similar approach to that for the age–period ones. However, the cohort structure is defined using principal component analysis. Thus, the UVR method determines the optimal number q of age–cohort components by pinpointing those that, again, reveal unique age clusters.

Step #4: The complete model structure is synthesised. Following the UVR-based method, after having estimated the significant (informative) age–period and age–cohort effects, we conclude to the final estimates of the log-graduated central mortality rates in age, period and cohort effects:

$$\log(\tilde{m}_{t,x,c}) = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{\mathrm{c}(j)} \cdot \gamma_c^{(j)}$$

for the calendar years $t = t_1, \ldots, t_n$, ages $x = x_1, \ldots, x_{\mathrm{a}}$ and cohorts $c = t - x = c_1, \ldots, c_{n_{\mathrm{c}}}$.

4.2 Methodology

In this section we proceed to the documentation of the extended form of the HS model, formulated in terms of $q_{t,x}$ (while the initial version presented in Chapter 3 was in terms of $m_{t,x}$), and we adapt the HS to be modelled using: (i) off-the-shelf link functions in the literature, namely the logit, complementary log–log and probit, and (ii) by introducing a new form of link functions in an effort to further improve the *goodness-of-fit* of the model

and to evaluate their applicability in the context of mortality through the HS extensions. The aforementioned cases are described in Sections 4.2.1 and 4.2.2, respectively.

4.2.1 HS Model Using Off-the-Shelf Link Functions

We extend the HS model through the formulation in terms of $q_{t,x}$ and by adopting the estimation methodology described in Section 4.1.1 making the appropriate configurations. Under this prism, the data include the number of deaths ($D_{t,x}$) and the initial exposure to risk ($E^0_{t,x}$), for $t = t_1, ..., t_n$ and for $x = x_1, ..., x_{\text{a}}$.

We treat $d_{t,x}$ as a random variable, $D_{t,x}$ and the initial exposures $E^0_{t,x}$ as fixed, then $D_{t,x}$ follows a Binomial distribution with mean $E^0_{t,x} \times q_{t,x}$ and let $Q_{t,x} = \frac{D_{t,x}}{E^0_{t,x}}$ be the random variable corresponding to $q_{t,x} = \frac{d_{t,x}}{E^0_{t,x}}$. In the case of the Binomial distribution, the following three link functions can be used:

i **logit:** $\eta = \log\left(\frac{q}{1-q}\right)$

ii **probit:** $\eta = \Phi^{-1}(q)$ where $\Phi(\cdot)$ is the Normal cumulative distribution function.

iii **complementary log–log:** $\eta = \log(-\log(1-q))$

Thus, and according to McCullagh & Nelder (1989) and Currie (2016), we have $0 < q < 1$ and a link should satisfy the condition that it maps the interval $(0, 1)$ onto the whole real line (e.g., $-\infty < \text{logit(q)} < \infty$).

At this point, we follow the four steps mentioned in Section 4.1.1 to model the age–period and age–cohort components. Therefore, a GLM approach is applied for each calendar year t independently, with dependent variable $D_{t,x}$, binomial error, $E^0_{t,x}$ weights, the corresponding link function and linear predictor $\eta = b \cdot \mathbf{X}$. We note that the exposures are introduced into the Poisson model (see Ch.3) as an offset, but into the Binomial model as a weight, and the model matrix, $\mathbf{X}$, is a matrix of orthonormal polynomials, so that:

- $\eta_t = \text{logit}(\mathbb{E}(Q_t)) = \text{logit}\left(\mathbb{E}\left(\frac{D_{t,x}}{E^0_{t,x}}\right)\right) = \log\left(\frac{q_t}{1-q_t}\right) = b_t \cdot L^T$
  or $\eta_t = \text{probit}(\mathbb{E}(Q_t)) = \Phi^{-1}(q_t) = b_t \cdot L^T$
  or $\eta_t = \text{cloglog}(\mathbb{E}(Q_t)) = \log(-\log(1-q_t)) = b_t \cdot L^T$

where b_t corresponds to the vector of the GLM-estimated parameters for calendar year t. Given this, a random (asymptotically normal) matrix $B = \{b_{t,k-1}\}$, $k = 1, \ldots, k_1$, of order $(n \times k_1)$, includes the GLM-estimated parameters for all calendar years $t = t_1, \ldots, t_n$. $L = \{l_{x,k-1}\}$, $k = 1, \ldots, k_1$, is a $(\mathrm{a} \times k_1)$-dimensional matrix of orthonormal polynomials with degree of $k_1 - 1$ for age x. The optimal degree $k_1 - 1$ of the orthonormal polynomials is defined by using a variety of statistical tests.

Then, we continue with the next two steps by applying the SPCA to the associated covariance matrix of the GLM-estimated parameters B and, in tandem with the UVR approach, we come to the following age–period stochastic mortality model:

$$\eta_{t,x} = \mathrm{logit}(q_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \varepsilon_{t,x}$$

In a similar way, we deploy the rest of the link functions, i.e., the probit and cloglog, in order to define the corresponding HS variations for the age–period part of the model.

Finally, the age–cohort components are estimated in a way similar to that of the age–period components, and then the final form of the estimates of the age–period–cohort stochastic mortality model is defined as:

$$\eta_{t,c,x} = \mathrm{logit}(\hat{q}_{t,c,x}) = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{\mathrm{c}(j)} \cdot \gamma_c^{(j)}$$

for the calendar years $t = t_1, \ldots, t_n$, ages $x = x_1, \ldots, x_{\mathrm{a}}$ and cohorts $c = t - x = c_1, \ldots, c_{n_{\mathrm{c}}}$. In a similar way, one can deploy the rest of the off-the-shelf link functions.

4.2.2 Introducing a New Form of Link Functions in the Mortality Modelling

Triggered by the probit case, where the cumulative distribution function of the standard normal distribution is used, we generalise this rationale as the inverse of any continuous cumulative distribution function can be used for the link, as long as a proper transformation is given to satisfy the condition that the CDF's range should be mapped to the whole real line.

As is the case for the probit, the same applies to the newly proposed form of link functions. We treat $d_{t,x}$ as a random variable, $D_{t,x}$ and the initial exposures $E_{t,x}^0$ as

fixed, then $D_{t,x}$ follows a Binomial distribution with mean $E^0_{t,x} \times q_{t,x}$ and let $Q_{t,x} = \frac{D_{t,x}}{E^0_{t,x}}$ be the random variable corresponding to $q_{t,x} = \frac{d_{t,x}}{E^0_{t,x}}$. As noted in Section 4.2.1, we have $0 < q_{t,x} < 1$. Thus, the link function that we choose should satisfy the condition that it maps the interval $(0, 1)$ to the whole real line. Following this rationale, there are the following three categories to be considered and provide a transformation method, depending on the cumulative distribution chosen:

1. The cumulative distribution as link function, maps q, $0 < q < 1$, to $-\infty < F^{-1}(q; \xi, \theta) < \infty$, so that:

$$\eta_t = F^{-1}\left(\mathbb{E}(Q_t); \xi, \theta\right) = F^{-1}\left(\mathbb{E}\left(\frac{D_t}{E^0_t}\right); \xi, \theta\right) = F^{-1}\left(q_t; \xi, \theta\right) = b_t \cdot L^T$$

   or

$$\mathbb{E}\left(\frac{D_t}{E^0_t}\right) = q_t = F(\eta_t; \xi, \theta) \Leftrightarrow q_t = F(b_t \cdot L^T; \xi, \theta)$$

   For instance, some distributions that can be used are: Normal, Logistic, Extreme value, Gumbel, etc.

2. The cumulative distribution as link function, maps q, $0 < q < 1$, to $0 < F^{-1}(q; \xi, \theta) < \infty$, so the logarithmic form of the cumulative distribution is needed to map q, to $-\infty < \log(F^{-1}(q; \xi, \theta)) < \infty$, the natural scale for regression, so that:

$$\eta_t = \log(F^{-1}(\mathbb{E}(Q_t); \xi, \theta)) = \log\left(F^{-1}(q_t; \xi, \theta)\right) = b_t \cdot L^T$$

   or

$$\mathbb{E}\left(\frac{D_t}{E^0_t}\right) = q_t = F(\exp(\eta_t); \xi, \theta) \Leftrightarrow q_t = F(\exp(b_t \cdot L^T); \xi, \theta)$$

   For instance, some distributions that can be used are: Generalised Pareto, Weibull, Fréchet, etc.

3. The cumulative distribution as link function, maps q, $0 < q < 1$, to $0 < F^{-1}(q; \xi, \theta) < 1$, so we need the logit of the cumulative distribution so that maps q, to $-\infty < \text{logit}(F^{-1}(q; \xi, \theta)) < \infty$, the natural scale for regression, so that:

$$\eta_t = \text{logit}(F^{-1}(\mathbb{E}(Q_t);\xi,\theta)) = \text{logit}\left(F^{-1}(q_t;\xi,\theta)\right) = \log\left(\frac{F^{-1}(q_t;\xi,\theta)}{1-F^{-1}(q_t;\xi,\theta)}\right) = b_t \cdot L^T$$

or

$$\mathbb{E}\left(\frac{D_t}{E_t^0}\right) = q_t = F\left(\frac{\exp(\eta_t)}{1+\exp(\eta_t)};\xi,\theta\right) \Leftrightarrow q_t = F\left(\frac{\exp(b_t \cdot L^T)}{1+\exp(b_t \cdot L^T)};\xi,\theta\right)$$

For instance, Beta distribution can be used in this case.

In this work, particular focus is given to the use of heavy-tailed distributions. From a modelling perspective, heavy-tailed distributions are important when extreme events must be part of the model. For example, although the probability of shock cases, such as mortality during World War II or the COVID-19 pandemic, is small, the magnitude of the impact is so large that these events are vital in capturing the mortality dynamics in an accurate manner. In this regard, as will be showcased in Section 4.3, we document the use of the Generalised Pareto and Beta distributions through the corresponding extensions of the HS model.

Table 4.1 summarises the transformations of the HS model along with key characteristics.

TABLE 4.1: Link functions transformations of the HS model.

Dist	**Link Name**	**Link Function $\boldsymbol{\eta = g(\mu) = X\beta}$**	**Mean Function $\boldsymbol{\mu = g^{-1}(X\beta)}$**
Binomial	Logit	$\log\left(\frac{\mu}{1-\mu}\right) = \mathbf{X\beta}$	$\mu = \frac{\exp(\mathbf{X\beta})}{1+\exp(\mathbf{X\beta})}$
	Cloglog	$\log(-\log(1-\mu) = \mathbf{X\beta}$	$\mu = 1-\exp(-\exp(\mathbf{X\beta}))$
	Probit	$\Phi^{-1}(\mu) = \mathbf{X\beta}$	$\mu = \Phi(\mathbf{X\beta})$
	Inverse CDF	$F^{-1}(\mu;\xi,\theta) = \mathbf{X\beta}$, if $-\infty < F^{-1}(x) < \infty$	$\mu = F(\mathbf{X\beta};\xi,\theta)$
		$\log(F^{-1}(\mu;\xi,\theta)) = \mathbf{X\beta}$, if $0 < F^{-1}(x) < \infty$	$\mu = F(\exp(\mathbf{X\beta});\xi,\theta)$
		$\text{logit}(F^{-1}(\mu;\xi,\theta)) = \mathbf{X\beta}$, if $0 < F^{-1}(x) < 1$	$\mu = F\left(\frac{\exp(\mathbf{X\beta})}{1+\exp(\mathbf{X\beta})};\xi,\theta\right)$

Based on the analysis given above, one needs to estimate the parameters of each possible distribution. That is, in the context of the HS model, the estimation methodology outlined in Section 4.1.1 needs to be revised in order to incorporate the estimation

process of the parameters of each possible distribution. This revised method is described in the following subsection.

4.2.2.1 HS Revised Estimation Methodology

The revised estimation process follows the rationale and the methodology steps analysed in Chapter 3 by making appropriate configurations. The ultimate goal is the definition of the age–period and age–cohort effects, but this time, by including the task of estimating the parameters of the adopted distribution.

In this section, we opt to showcase how the estimation process is updated using the 2$^{\text{nd}}$ case of the new form of the link function, namely the logarithm of the inverse cumulative distribution, as documented in Section 4.2.2, i.e., $-\infty < \log(F^{-1}(q;\xi,\theta)) < \infty$, where F is a cumulative distribution (e.g., Generalised Pareto). The same process can also be followed for the rest of the identified cases in Section 4.2.2, but we opt to document one of the cases for simplicity reasons.

In this regard, the estimation of the parameters of the adopted distribution requires the estimation of the (ξ,θ) pairs for both the age–period and the age–cohort part of the model, referring to them as $(\xi^{\text{t}},\theta^{\text{t}})$ and $(\xi^{\text{c}},\theta^{\text{c}})$, respectively.

Hence, regarding the age–period effects, before we follow the four steps of the estimation methodology described in Section 4.1.1, we first need to determine the parameters of the selected distribution and the respective link function.

For fitting the probabilities of deaths, $q_{t,x}$, to the chosen cumulative distribution, we need to pinpoint a particular pair $(\xi^{\text{t}},\theta^{\text{t}})$ of parameter values from the set of all possible pairs, so that the $\hat{q}_{t,x}$ will be estimated as close as possible to the observed data. To build an unambiguous approach, this parameter estimation process must be driven by a measure of distance or discrepancy between the observed $q_{t,x}$ and the fitted $\hat{q}_{t,x}$ (e.g., classical least square) or by maximising the model's log-likelihood. In our case, we opt for the log-likelihood maximisation approach given by:

$$\mathcal{L}(d_{t,x},\hat{d}_{t,x}) = \sum_t \sum_x \left\{ d_{t,x} \log\left(\frac{\hat{d}_{t,x}}{E^0_{t,x}}\right) + (E^0_{t,x} - d_{t,x}) \log\left(\frac{E^0_{t,x} - \hat{d}_{t,x}}{E^0_{t,x}}\right) + \log\binom{E^0_{t,x}}{d_{t,x}} \right\} \tag{4.4}$$

where

$$\hat{d}_{t,x} = E^0_{t,x} \cdot F(\exp(\tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)}); \xi^{\mathrm{t}}, \theta^{\mathrm{t}})) \tag{4.5}$$

and the pair of parameters $(\xi^{\mathrm{t}}, \theta^{\mathrm{t}})$ is the one that maximises the aforementioned measure.

The fitted $\hat{q}_{t,x}$ are estimated by using the GLM approach. Thus, a GLM approach is applied for each calendar year t independently, with dependent variable $D_{t,x}$, and the initial exposures $E^0_{t,x}$ are handled as a weight. By using the appropriate link function, the following predictor structure is defined:

$$\begin{aligned} \eta_t &= \log(F^{-1}(\mathbb{E}(Q_t); \xi^{\mathrm{t}}, \theta^{\mathrm{t}})) = \log\left(F^{-1}\left(\mathbb{E}\left(\frac{D_{t,x}}{E^0_{t,x}}\right); \xi^{\mathrm{t}}, \theta^{\mathrm{t}}\right)\right) \\ &= \log\left(F^{-1}(q_t; \xi^{\mathrm{t}}, \theta^{\mathrm{t}})\right) = b_t \cdot L^T \end{aligned} \tag{4.6}$$

where b_t corresponds to the vector of the GLM-estimated parameters for calendar year t, having chosen the optimum parameter values $(\xi^{\mathrm{t}}, \theta^{\mathrm{t}})$. Given this, a random (asymptotically normal) matrix $B = \{b_{t,k-1}\}$, $k = 1, \ldots, k_1$, of order $(n \times k_1)$, includes the GLM-estimated parameters for all calendar years $t = t_1, \ldots, t_n$. $L = \{l_{x,k-1}\}$, $k = 1, \ldots, k_1$, is a $(\mathrm{a} \times k_1)$-dimensional matrix of orthonormal polynomials of $k_1 - 1$ degree for age x. The optimal degree $k_1 - 1$ of the orthonormal polynomials is defined by using a variety of statistical tests.

According to the steps described in Section 4.1.1, SPCA is applied to the covariance matrix of B to decrease the dimensionality of the problem. Note that we apply the UVR approach, not only to define the "optimum" s value for the sparsity but also to define the most significant p age–period components. In such a manner, the number of age–period components is decreased from k_1 to p. Thus, the age–period form (logarithm of the inverse cumulative distribution in this case) of the model is as follows:

$$\log(F^{-1}(q_t; \xi^{\mathrm{t}}, \theta^{\mathrm{t}})) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)} + \varepsilon_{t,x} \tag{4.7}$$

where the disturbance term $\varepsilon_{t,x} \sim N(0, v_x)$ (or residuals) is the error component at age x in year t and reflects the deviation of the model due to the excluded components.

Regarding the age–cohort components, the estimation process is, in a way, similar to that for the age–period components. Therefore, our data, $D_{t,x}$, $E^0_{t,x}$ and $q_{t,x}$, are processed

as cohorts.

Thus, the GLM approach is applied for each cohort $c = t - c = c_1, ..., c_{nc}$, with dependent variable $D_{c,x}$. Note that the exposures are introduced as a weight, and the p age–period effects (Equation (4.7)) are handled as an offset, so that:

$$\eta_c = \log(F^{-1}(q_c; \xi^{\mathrm{c}}, \theta^{\mathrm{c}})) = b_c \cdot (L^{\mathrm{c}})^T + \text{offset} \tag{4.8}$$

$$\text{offset} = \log(F^{-1}(\hat{q}_t; \xi^{\mathrm{t}}, \theta^{\mathrm{t}})) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)}$$

where b_c corresponds to the vector of the GLM-estimated parameters for each cohort c. Given this, a random (asymptotically normal) matrix $B^{\mathrm{c}} = \{b_{c,k-1}\}$, $k = 1, \ldots, k_2$, of order $(n_{\mathrm{c}} \times k_2)$, includes the GLM-estimated parameters for all cohorts $c = c_1, \ldots, c_{n_{\mathrm{c}}}$. $L^{\mathrm{c}} = \{l^{\mathrm{c}}_{x,k-1}\}$, for $k = 1, \ldots, k_2$, is an (a $\times$ k_2)-dimensional matrix of orthonormal polynomials with $k_2 - 1$ degree for age x. Note that the pair of optimum parameter values $(\xi^{\mathrm{c}}, \theta^{\mathrm{c}})$ for the age–cohort components are defined by following the same approach as described in the age–period case, i.e., by maximising the log-likelihood of the model. To do so, this process is driven again by Equation (4.4), but $d_{t,x}$ and $E^0_{t,x}$ are processed as cohorts and $\hat{d}_{c,x}$ is defined as:

$$\hat{d}_{c,x} = E^0_{c,x} \cdot F\left(\exp\left(\tilde{\alpha}^{\mathrm{c}}_x + \sum_{j=1}^{q} \beta^{\mathrm{c}(j)}_x \cdot \gamma_c^{(j)} + \text{offset}\right); \xi^{\mathrm{c}}, \theta^{\mathrm{c}}\right) \tag{4.9}$$

The optimum degree $k_2 - 1$ of the orthogonal polynomials L^{c} is achieved using the same statistical tests as was the case for the age–period effect estimation methodology.

Finally, we apply eigenvalue decomposition to the matrix of the GLM-estimated parameters B^{c} and, in tandem with the UVR^{c} approach described in Hatzopoulos & Sagianou (2020), which is used to determine the optimum number of age–cohort components, we come to the following age–cohort stochastic mortality model:

$$\eta_{c,x} = \log(F^{-1}(q_{c,x}; \xi^{\mathrm{c}}, \theta^{\mathrm{c}})) = \tilde{\alpha}^{\mathrm{c}}_x + \sum_{j=1}^{q} \beta^{\mathrm{c}(j)}_x \cdot \gamma_c^{(j)} + \varepsilon_{c,x} \tag{4.10}$$

Hence, by combining Equations (4.7) and (4.10), the final form of the age–period–cohort stochastic mortality model is defined as:

$$\eta_{t,c,x} = \log(F^{-1}(\hat{q}_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}}) + \log(F^{-1}(\hat{q}_{c,x};\xi^{\mathrm{c}},\theta^{\mathrm{c}}) = \alpha_x + \sum_{i=1}^{p}\beta_x^{(i)}\cdot\kappa_t^{(i)} + \sum_{j=1}^{q}\beta_x^{\mathrm{c}(j)}\cdot\gamma_c^{(j)} \tag{4.11}$$

for the calendar years $t = t_1,\ldots,t_n$, ages $x = x_1,\ldots,x_{\mathrm{a}}$ and cohorts $c = t - x = c_1,\ldots,c_{n_{\mathrm{c}}}$.

According to the aforementioned method, after estimating the significant age–period and age–cohort effects utilising the SPCA, UVR and GLM approaches, we conclude to the final estimates of the probabilities of death in age, period and cohort effects.

It becomes clear that the UVR plays an important role in the formation of the model. That is, the UVR takes the form given in Equation (4.12), specifically for the case $\log(F^{-1}(q;\xi,\theta))$, where F is the chosen cumulative distribution. We prompt the reader to refer to Section 3.2.1.3 for more information on the use of the UVR metric.

In short, the aim is to measure the variance of the time series of the $\log(F^{-1}(q_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}}))$ and the variance of the error between the estimated probabilities of death and the actual ones (for each age x and each component i). The ratio of these two variances for successive components defines the $UVR_x(i)$ value. Based on the updated estimation methodology described in this section, the UVR metric is given in Equation (4.12).

$$UVR_x(i) = \begin{cases} \dfrac{\mathrm{Var}\left(\log(F^{-1}(q_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}})) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)}\right)}{\mathrm{Var}\left(\log(F^{-1}(q_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}})) - \alpha_x\right)}, & i = 1 \\ \dfrac{\mathrm{Var}\left(\log(F^{-1}(q_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}})) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)} - \beta_x^{(i)}\kappa_t^{(i)}\right)}{\mathrm{Var}\left(\log(F^{-1}(q_{t,x};\xi^{\mathrm{t}},\theta^{\mathrm{t}})) - \alpha_x - \beta_x^{(1)}\kappa_t^{(1)}\right)}, & i = 2,\ldots,p \end{cases} \tag{4.12}$$

In a similar way, one can adapt the HS model to the rest of the link functions documented in Table 4.1.

Implementation-wise, our model and the proposed extensions have been developed in Matlab, which enables a user to define their own link functions. Taking advantage of this functionality, we fit the Binomial model with new link for the probabilities of death

$q_{t,x}$ using the following command:

$$\mathrm{B} = \mathrm{glmfit}(\mathrm{L_x}, \mathrm{q_{t,x}}, \text{"binomial"}, \text{"link"}, \mathrm{new\_F}, \text{"weights"}, \mathrm{E^0_{t,x}}, \text{"constant"}, \text{"off"})$$

Details of the user-defined link new_F are given in Appendix B.1.

4.3 Application

In this section, we elaborate on the experimental approach followed to evaluate the performance of the HS model using the newly proposed methodology presented in Section 4.2. In order to offer a fair and straightforward comparison to the performance of the HS model, we opt for the same evaluation metrics and datasets as in Chapter 3. In addition, we perform a comparison with other well-known mortality models of the literature, namely the Lee & Carter (1992) (LC), Renshaw & Haberman (2006) (RH), Currie (2006) (APC) and Plat (2009) (PL) models. This section also describes the datasets and the experimental parameters of the HS model extensions and analyses the results of the evaluation through a comparative analysis.

Moreover, we evaluate the models in terms of forecast. That is, we provide a back-testing framework to forecast 10 years ahead in order to compare out-of-sample estimations against the actual values. Based on this, we measure the efficiency of the models using the MSPE and the MAPE metrics. For providing an objective and unbiased evaluation, we forecast the κ_t and the γ_c indexes using two different approaches, namely the Random Walk with Drift (RWD) (ARIMA(0,1,0) with drift), and the appropriate ARIMA model based on Bayesian Information Criterion (ARIMA).

It has to be noted that our evaluation considers two experimental setups to evaluate the capacity of the HS extensions on datasets having both long and short periods. More specifically, apart from the E&W dataset that offers an extremely long fitting period, we apply the models to the Greek dataset, which begins in 1961 and contrasts to the rather long periods of the E&W mortality data that start in 1841. The motivation behind this approach lies in the fact that the utilisation of a different link function and distribution could make the difference to the fitting process and deliver a more accurate model fit, as

the link function and the distribution could aid in capturing the intrinsic characteristics of the data in a more precise way.

4.3.1 Evaluation Metrics

We have chosen a set of quantitative criteria in an effort to approach the problem from different angles. Both Information Criteria and Percentage error tests are used to evaluate the fitting performance of the models. Specifically, we use the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), the Mean Squared Percentage Error (MSPE) and the Mean Absolute Percentage Error (MAPE). These metrics have been defined in Section 2.4. Additionally, when it comes to the qualitative analysis of the results, we adopt the unexplained variance ratio (UVR) which is used to reveal the magnitude of the captured information by each age–period and age–cohort component, as has been defined in Section 3.3.3.2.

4.3.2 Evaluation Results

In this section, we present the data and the optimum parameters of the HS model and its extensions used to evaluate the proposed methodology described in Section 4.2. Sections 4.3.2.2 and 4.3.2.3 document the results and proceed with the comparative analysis.

4.3.2.1 Mortality Data and the Optimum Parameters

The historic mortality data used are the same presented in Chapter 3 for the purpose of making our results directly comparable. We have selected datasets of both long and short periods to evaluate the behaviour of all extensions of the HS model in an effort to validate its modelling capacity both in the case of highly informative and non-informative data. The latter are provided by the Human-Mortality-Database and Eurostat, having the following characteristics:

- Greece (GR), males, calendar years 1961–2013, individual ages 0–84.
- England and Wales (E&W), males, calendar years 1841–2016, individual ages 0–89.

Tables 4.2 and 4.3 present the parameters used for fitting each model structure for each of the aforementioned datasets. The tables contain the number of optimum parameters of the GLM polynomials, k_1 (Equation (4.6)) and k_2 (Equation (4.8)), for the age–period and age–cohort effects, respectively. In addition, the tables document the optimum s values, while p and q denote the optimum number of the period (Equation (4.7)) and cohort (Equation (4.10)) components, according to the approach described in Section 4.2. ξ^{t}, θ^{t} and ξ^{c}, θ^{c} are the parameters (see Section 4.2.2.1) of the adopted distribution for the age–period and the age–cohort part of the model, respectively.

TABLE 4.2: Optimum parameters for each HS model extension structure for the E&W dataset.

Country: E&W	**Years: 1841–2016**					**Ages: 0–89 (Years)**			
Model	s	k_1	p	k_2	q	ξ^{t}	θ^{t}	ξ^{c}	θ^{c}
HS_log	54.50	29	5	4	2	-	-	-	-
HS_lgt	55.19	30	5	14	2	-	-	-	-
HS_cll	6.40	30	5	14	2	-	-	-	-
HS_prbt	9.26	30	5	10	2	-	-	-	-
HS_beta	4.13	30	5	8	2	6.00	1.25	4.00	1.25
HS_prt	71.45	23	6	4	1	16.50	1.00	11.50	1.00

TABLE 4.3: Optimum parameters for each HS model extension structure for the GR dataset.

Country: Greece	**Years: 1961–2013**					**Ages: 0–84 (Years)**			
Model	s	k_1	p	k_2	q	ξ^{t}	θ^{t}	ξ^{c}	θ^{c}
HS_log	2.55	16	4	8	1	-	-	-	-
HS_lgt	2.52	16	4	8	1	-	-	-	-
HS_cll	2.58	16	4	8	1	-	-	-	-
HS_prbt	0.21	16	5	8	1	-	-	-	-
HS_beta	2.18	20	5	8	1	1.25	0.50	1.00	0.25
HS_prt	3.37	20	5	8	1	4.00	1.00	7.00	1.00

In our application, we make use of the Generalised Pareto and Beta distributions to offer the HS_prt and the HS_beta model extensions, respectively. The former utilises the probability density function for the Generalised Pareto distribution with shape parameter $\xi \neq 0$, scale parameter θ, as given in Equation (4.13), based on the Matlab implementation given in MATLAB (n.d.*b*).

$$f\left(x|\xi,\theta\right) = \left(\frac{1}{\theta}\right)\left(1+\xi\frac{x}{\theta}\right)^{-1-\frac{1}{\xi}} \tag{4.13}$$

and the respective cumulative distribution function is defined as:

$$F\left(x|\xi,\theta\right) = 1 - \left(1 + \xi\frac{x}{\theta}\right)^{-\frac{1}{\xi}} \tag{4.14}$$

In our implementation, we set the scale parameter $\theta = 1$. Equation (4.13) can take the form of the Lomax distribution as a special case of the Generalised Pareto distribution, with $\xi = \frac{1}{\alpha}$ and $\theta = \frac{\lambda}{\alpha}$. In that case, the probability density function is

$$f\left(x|\alpha,\lambda\right) = \frac{\alpha}{\lambda}\left(1 + \frac{x}{\lambda}\right)^{-1-\alpha} \tag{4.15}$$

where $\alpha = \frac{1}{\xi}$ and $\lambda = \frac{\theta}{\xi}$.

For the Beta extension, we make use of the probability density function of Equation (4.16), with first shape parameter ξ and second shape parameter θ, following the Matlab implementation given in MATLAB (n.d.*a*).

$$f\left(x|\xi,\theta\right) = \frac{1}{B(\xi,\theta)}x^{\xi-1}(1-x)^{\theta-1} \tag{4.16}$$

and the respective cumulative distribution function is defined as:

$$F\left(x|\xi,\theta\right) = \frac{B(x|\xi,\theta)}{B(\xi,\theta)} \tag{4.17}$$

where $B(\xi,\theta)$ is the beta function and $B(x|\xi,\theta)$ is the incomplete gamma function.

4.3.2.2 E&W Data Performance Analysis

E&W Quantitative Analysis Table 4.4 summarises the quantitative results that reflect the efficacy of the HS model and its extensions for the fitting process in terms of the log-likelihood-based tests (AIC, BIC) and PE tests (MSPE, MAPE). The best scores are highlighted in bold, with the HS_prt extension being dominant among them.

TABLE 4.4: Results of the quantitative tests for E&W mortality dataset.

Model	k_1	**npar**	**Log-Likelihood**	**AIC**	**BIC**	**MSPE (%)**	**MAPE (%)**
HS_log	29	1598	−135,970.84	275,137.68	287,394.81	**0.329**	3.696
HS_lgt	30	1634	−125,713.76	254,695.52	267,228.78	0.336	3.694
HS_cll	30	1634	−168,056.21	339,566.43	352,813.03	0.449	4.707
HS_prbt	30	1622	−115,111.76	233,467.52	245,908.73	0.346	3.693
HS_beta	30	**1343**	−129,916.62	262,519.24	272,820.45	0.381	3.945
HS_prt	**23**	1492	**−113,143.64**	**229,271.28**	**240,715.35**	0.346	**3.596**

A key observation that clearly supports the motivation of this chapter is that the HS model under the binomial form and $q_{t,x}$ modelling reached a higher level of performance. More specifically, four out of the five extensions performed better than the original HS_log model, showing better *goodness-of-fit* through the log-likelihood, AIC and BIC metrics and a slight improvement for the MAPE. Notably, even though the BIC is destined to penalise complex structures, the HS_lgt and HS_prbt extensions score a better BIC even when having more parameters than the HS_log, meaning that the additional ones contribute substantially to the improvement of fit. In fact, the better performance of the HS_lgt and HS_prbt will be noted also on the qualitative results later in this section. Among the HS extensions based on the off-the-shelf link functions, the HS_prbt is the one achieving the best performance under all metrics, having also less parameters than the HS_lgt and HS_cll. On the downside, the HS_cll extension had the lowest performance among the collected results. When it comes to the structure of the models of the HS_lgt, HS_cll and HS_prbt extensions, according to Table 4.2, the number of the generated p and q components remains the same as those of the HS_log. However, a slight deviation is noted for the k_1 and k_2 factors, which also justifies the increased number of parameters noted in Table 4.2. Overall, as will be shown in the qualitative analysis, the HS extensions remain consistent when it comes to the identified p and q components and the corresponding age ranges which are revealed.

When it comes to the quantitative results of the HS extensions based on the new form of link functions, i.e., the HS_beta and HS_prt cases, the latter is the one dominating among all, while the former is ranked around the performance achieved for the HS_lgt, HS_cll and HS_prbt cases. According to Table 4.4, the HS_beta incorporates the lowest number of parameters, resulting in a more simple model structure, but this is reflected as an inefficiency in the rest of the qualitative metrics, implying that the HS_beta cannot

capture the E&W data dynamics in an efficient manner. When it comes to the incorporated p and q components, the HS_beta still remains consistent, as will be discussed in the qualitative analysis, but fails to make a difference in terms of the quantitative metrics.

Notably, the case of the HS_prt extension is of special interest as it outperforms the rest of the extensions and the original HS_log for all metrics, apart from a negligible deficiency on the MSPE. More specifically, contrary to the HS_log, the HS_prt has an improvement of 16.24%, 16.67% and 16.79% for the BIC, AIC and log-likelihood metrics, respectively, while the MAPE is improved by 2.7%. According to Table 4.4, the HS_prt comes with a decreased number of parameters and a smaller number of the optimum parameters of the GLM orthonormal polynomials (k_1). Hence, the HS_prt extension concludes to a simpler model which also produces lower error in the fitting process. In addition, according to the p and q parameters in Table 4.2, the HS_prt has identified one age–period component more (i.e., 6) and one age–cohort component less (i.e., 1). Such a difference may sound minor, but as will be documented in the qualitative analysis, it is quite substantial for the explainability of mortality.

Moreover, the results of the quantitative evaluation of the well-known stochastic mortality models are given in Table 4.5. According to the table, and having selected the best model from the extensions in Table 4.4, we notice that the HS_prt not only performs better compared to the HS_log but also achieves the best scores among the rest of the well-known stochastic mortality models.

TABLE 4.5: Results of the quantitative tests for well-known stochastic mortality models against HS_prt for E&W dataset.

Model	**npar**	**Log-Likelihood**	**AIC**	**BIC**	**MSPE (%)**	**MAPE (%)**
LC	354	−1,075,745.35	2,152,198.70	2,154,913.98	7.131	18.473
RH	707	−553,002.06	1,107,418.13	1,112,841.02	5.369	14.896
APC	528	−1,157,738.25	2,316,532.50	2,320,582.42	20.541	23.522
PL	880	−762,205.25	1,526,170.51	1,532,920.37	9.136	15.737
HS_log	1598	−135,970.84	275,137.68	287,394.81	**0.329**	3.696
HS_prt	1492	**−113,143.64**	**229,271.28**	**240,715.35**	0.346	**3.596**

E&W Qualitative Analysis The results of the qualitative analysis aim to complement the numeric ones presented in Table 4.4. To do so, Appendix B.2 offers the

graphical representation of the age–period and age–cohort components for both the off-the-shelf HS extensions and those following the new form of link function. The graphical representation reveals the identified mortality trends and eases the attribution of a trend to a specific age cluster.

Figures B.1 and B.2 provide a graphical representation of the age–period components and the corresponding UVR values for all the HS extensions. One can observe that the HS extensions remain consistent with the HS_log and reveal the same age clusters. However, in some cases, according to the UVR values, the HS extensions under the Binomial form can capture more information from the mortality data. More specifically, according to $UVR_x(1)$, the HS_prbt, HS_beta and HS_prt model structures are able to grasp more information regarding the 60+ ages. In fact, this reflects a slightly different mortality trend in the κ_t^1 factor for the HS_prt. The κ_t^1 for the HS_prbt and HS_beta is factorised in a different way than the other cases, that is why the corresponding graphs seem to be dislocated. Despite this fact, the HS_prbt and HS_beta generate κ_t^1 graphs which are almost identical to the HS_prt case.

$UVR_x(3)$ and $UVR_x(4)$ refer to the 20-25 and 30-40 age clusters, respectively. The lower UVR values achieved by the HS_prbt structure, which are illustrated in Figures B.1 and B.2, confirm the good performance advocated by the quantitative performance reported in Table 4.4.

When it comes to Figure B.2, the HS extensions have revealed a more clear age cluster compared to the HS_log. More specifically, according to $UVR_x(5)$, the HS_log had initially identified a mixed age cluster for the 10–20 and 20–25 age ranges. In this work, all the HS extensions clearly attribute the κ_t^5 mortality trend to the 10–20 age range, while the 20–25 age cluster is clearly addressed by the 3rd age–period component. Overall, the HS_prt structure is of special interest. Again, one can notice that the HS model for the Generalised Pareto case remains consistent as, in general terms, the identified age–period and age–cohort components reveal fine-grained age clusters and ease the attribution of a mortality trend to distinct age ranges. In fact, the most important outcome of the application for the HS_prt is the identification of one additional and distinct age–period component. $UVR_x(6)$ is a new informative component that reveals the mortality trend for the 25–30 age range. As aforementioned, the focal point of $UVR_x(3)$ is the 20–25

age range, while a significant and clear trend for the 10–20 age cluster is revealed in κ_t^5 as advocated by $UVR_x(5)$. In fact, in the case of the HS_log, the 10–20 cluster was included in the captured variance of the 5th age–period component along with the variance of the 20–30 age range. Considering these facts, the HS_prt model generated more fine-grained age clusters, and it eases the attribution of a mortality trend to unique age clusters, contributing significantly to the explainability of the model. This behaviour justifies the improvement of around 16% in the quantitative metrics in contrast to the HS_log.

Moreover, in Figure B.3, one can notice that for the HS_prt, only one age–cohort component is identified. This is a normal outcome, as the age–period modelling process has captured most of the data variance and resulted in one additional age–period component (six in total). As a result, the residuals bear the information for the identification of one age–cohort component. Generally, this modelling behaviour results in a less complex model structure which incorporates less parameters as shown in Table 4.2.

Overall, when it comes to the off-the-shelf HS extensions, the HS_prbt case seems to perform better both in the qualitative and quantitative evaluation results. When using the new form of link function, the HS_prt case is the one that stands out among all cases.

4.3.2.3 Greek Data Performance Analysis

Based on Table 4.3, the Greek dataset covers a shorter period (from 1961) in contrast to the rather long period for the E&W data. One could say that it is reasonable for multiple component models, such as the HS model and the introduced extensions, to achieve good results. That is, in order to verify that the better performance of the HS extensions (as happened for the case of the E&W data) is not solely attributed to the extremely long fitting periods, we also conduct the evaluation using the shorter GR dataset, and we assess the consistency and performance of the Binomial HS extensions with respect to the original Poisson model (HS_log) presented in Hatzopoulos & Sagianou (2020) and in Chapter 3.

In fact, according to the Human-Mortality-Database and as reported in the HMD-Greek-data, the GR dataset contains inconsistencies and is of poor quality. Hence, due to the

limited time period of the GR data and the low quality, we form a more challenging basis for a model to thrive. In addition, we aim to prove through experimental results that the model extensions inherit the beneficial characteristics of the HS model and can perform even better through the use of the suggested transformations of our methodology, being at the same time consistent with the identification of mortality trends for unique age clusters.

GR Quantitative Analysis Following the same approach as in Section 4.3.2.2, Table 4.6 summarises the quantitative results that reflect the efficacy of the HS model and its extensions for the fitting processes. The best scores are highlighted in bold, with the HS_prt being the dominant and the HS_beta being very close to the HS_prt.

TABLE 4.6: Results of the quantitative tests for the Greek mortality dataset.

Model	k_1	**npar**	**Log-Likelihood**	**AIC**	**BIC**	**MSPE (%)**	**MAPE (%)**
HS_log	16	**447**	−20,176.13	41,246.26	44,112.84	4.109	10.115
HS_lgt	16	**447**	−20,069.65	41,033.29	43,899.88	4.080	10.065
HS_cll	16	**447**	−20,133.42	41,160.85	44,027.43	4.104	10.079
HS_prbt	16	516	−19,569.21	40,170.43	43,479.50	4.078	10.074
HS_beta	20	540	−19,358.74	39,797.48	43,260.47	**3.444**	9.515
HS_prt	20	540	**−19,353.29**	**39,786.59**	**43,249.58**	3.449	**9.498**

As advocated by the quantitative results, also for the case of the short GR dataset, the HS model under the binomial form and $q_{t,x}$ modelling achieved better performance scores. More specifically, all four extensions performed better than the original HS_log model, showing better *goodness-of-fit* through the log-likelihood, AIC and BIC metrics and a slight improvement for the MAPE. However, it must be mentioned that the improvement is not as significant as the one achieved for the E&W case. Taking a closer look at the performance of the HS extensions for the off-the-self link functions, the HS_prbt is the one that stands out in contrast to the HS_lgt and HS_cll, which achieve quite similar performances with the initial HS_log. In fact, the HS_prbt has a negligible improvement for the MSPE and the MAPE, but the improvement in the log-likehood, AIC and BIC needs to be highlighted. As can be seen, the HS_prbt is based on 516 parameters, leading one to expect higher BIC and AIC scores due to the more complex structure. Though, despite the higher number of parameters, the HS_prbt structure achieves a better BIC, meaning that the additional parameters contribute substantially

to the improvement of fit. In fact, the better performance of the HS_prbt is more evident in the qualitative results which are analysed later in this section. According to Table 4.3, the number of the generated p and q components of the HS_lgt and HS_cll extensions remain the same as those of the HS_log (i.e., 4) and so does the k_1 factor. However, for the HS_prbt, a new additional age–period component is identified; thus, $p = 5$, and as will be noted in the qualitative analysis, this component contributes to the explainability of the model.

A case of special interest is the application of the HS_prt, for which the new form of link functions led to a notable improvement of the model. More specifically, Table 4.6 reports that the HS_prt achieves a significant improvement for all the evaluation metrics, apart from the MSPE, for which the HS_beta is slightly better. More specifically, contrary to the HS_log, the HS_prt has an improvement of 1.96%, 3.54% and 4.25% for the BIC, AIC and log-likelihood metrics, respectively, while the MSPE is improved by 16.06% and the MAPE by 6.10%. Even though the magnitude of improvement is less than the improvement for the case of the E&W data, one needs to consider the challenging nature of the Greek mortality data, where the room for improvement is rather limited. In addition, the HS_beta achieves almost identical performance to the HS_prt and, without a doubt, it is also a competitive extension for the GR case.

The HS_log model was slightly lacking in performance when compared with the Renshaw–Haberman Renshaw & Haberman (2006) and Plat Plat (2009) models, specifically in the case of the Greek data, as reported by the comparative study given in Hatzopoulos & Sagianou (2020) and as can be seen in Table 4.7. By utilising the proposed methodology, the HS_prt and HS_beta extensions have significantly improved the *goodness-of-fit*. In addition, according to the p and q parameters in Table 4.3, the HS_prt and HS_beta have identified one age–period component more (i.e., 5), as was the case with the HS_prbt, and this contributes substantially to the explainability of mortality.

The results of the quantitative evaluation against the well-known stochastic mortality models LC, RH, APC and PL are given in Table 4.7. According to the table, and having selected the best model from the extensions in Table 4.6, we note that the HS_prt is the model variation that achieves a similar performance to the RH model in terms of the quantitative tests. However, the RH can be questioned for its robustness, as it faces difficulties to converge and presents unstable behaviour (Hunt & Villegas (2015)).

TABLE 4.7: Results of the quantitative tests for well-known stochastic mortality models against HS_prt for GR dataset.

Model	npar	Log-Likelihood	AIC	BIC	MSPE (%)	MAPE (%)
LC	221	−21,213.44	42,868.89	44,286.15	4.287	11.379
RH	441	**−18,633.10**	**38,148.19**	**40,976.30**	3.790	**9.460**
APC	272	−21,317.67	43,179.33	44,923.65	5.405	12.715
PL	378	−19,113.68	38,983.36	41,407.45	5.395	10.371
HS_log	447	−20,176.13	41,246.26	44,112.84	4.109	10.115
HS_prt	540	−19,353.29	39,786.59	43,249.58	**3.449**	9.498

It must be noted that the HS_prt achieved a significant improvement in the case of the E&W data, contrary to the GR. As discussed before, the Generalised Pareto belongs to the heavy-tailed distributions family. That is, it is expected to perform better when extreme values are part of the data. As can be seen from the κ_t^i figures in Appendices B.2 and B.3, the E&W data include several high peaks in contrast to the GR. That is why the HS_prt seems to have a greater performance improvement in the E&W case.

GR Qualitative Analysis The results of the qualitative analysis complement the numeric ones presented in Table 4.6 by explaining the UVR for the identified age clusters. The graphical representations of the identified age–period and age–cohort components of the Greek dataset are given in Appendix B.3. Figures B.4 and B.5 provide a graphical representation of the age–period components and the corresponding UVR values for the HS extensions.

As a general remark, one might note that all the extensions remain consistent, but the HS_prbt, HS_beta and HS_prt identified an extra (5th) age–period component. Overall, slight differences in the UVR graphs and the mortality trends can be noticed among the other extensions.

The HS_prbt is of special interest as its structure leads to the definition of more fine-grained age clusters. More specifically, focusing on Figure B.4, $UVR_x(1)$ reveals that the HS_prbt is able to grasp more information regarding the 60+ ages. $UVR_x(4)$ is a component that reflects the difficulty to identify clear trends on the Greek mortality data. As can be seen, the interpretation of the cluster is quite difficult, as for all model extensions, there is no strong indication of the age cluster which is being highlighted. On the bright side, the HS_prbt is able to identify a quite clear cluster and the corresponding mortality trend for the 30–40 age cluster. In addition, $UVR_x(2)$ shows a clearer

concentration of the explained variance to the ages around 80, in contrast to the rest of the extension which makes an attribution to the 70–80 cluster. In fact, the HS_prbt has identified a clear cluster for the ages of 70 in $UVR_x(5)$ of Figure B.5. Hence, it is evident that the HS_prbt structure achieved a clean cut in the 70–80 cluster, and it eases the attribution of a trend to a specific age cluster. In addition, despite the different structure achieved for the age–period components (i.e., 5 instead of 4), $UVR_x^c(1)$ in Figure B.6 remains consistent and still captures considerable variance of the data.

When it comes to the HS_prt, as advocated by the quantitative results, it achieved a noticeable improvement from the HS_log case, and this is actually reflected in the figures of Appendix B.3. $UVR_x(2)$ for the HS_prt can significantly explain the variance of the 80s age range, while the focal point of $UVR_x(3)$ is on the 20s. The middle ages of 40–50 are addressed by $UVR_x(4)$, and finally, a newly identified cluster for the 70s is highlighted in $UVR_x(5)$. Almost the same behaviour can be seen for the HS_beta extension. The most important outcome of the application for the HS_prt and HS_beta, as was the case for the HS_prbt, is the identification of the 5th age–period component. This was not initially uncovered by the HS_log model. Thus, the addition of the new component justifies the improvements noted for the quantitative performance metrics. The HS_prt model remains consistent as the identified age–period and age–cohort components reveal fine-grained age clusters and ease the attribution of a mortality trend to distinct age ranges, even when operating over the non-informative Greek dataset.

When it comes to the age–cohort components, Figure B.6 shows that for the HS extensions, $UVR_x^c(1)$ remains consistent and still captures a considerable variance of the data for all cases. However, due to the existence of the 5th age–period component for the HS_prbt, HS_beta and HS_prt, the age–cohort differs slightly from that of the HS_log, HS_cll and HS_lgt. Nonetheless, the age–cohort components are still informative, they denote a clear age cluster and contribute to the explainability of the model.

4.3.2.4 Out-of-Sample Results

The results of the quantitative analysis for the forecasting process complement the comparative evaluation. In order to evaluate the HS_prt extension under the out-of-sample

mode, the model needs to be fit to a shorter dataset, excluding the years to be forecasted. This implies that the fitting and estimation processes will be executed again in order to define the necessary model parameters. The parameters of the HS_prt for the E&W and GR datasets for the shorter periods are given in Table 4.8.

TABLE 4.8: Optimum parameters for HS_prt structure for the E&W and GR shorter datasets.

Model: HS_prt										
Country	**Years**	**Ages**	s	k_1	p	k_2	q	ξ^{t}	θ^{t}	ξ^{c}	θ^{c}
E&W	1841–2006	0–89	56.58	29	6	4	1	15	1.00	11.50	1.00
GR	1961–2003	0–84	1.55	20	3	8	1	5.50	1.00	11.50	1.00

According to the results of Tables 4.9 and 4.10, the HS_prt model under the RWD and ARIMA achieves lower error scores and demonstrates a consistent behaviour also under the forecasting mode. More specifically, for the E&W case, the difference in the MSPE and MAPE is considerable, comparing to the other well-known models. This is not only attributed to the high *goodness-of-fit* performance of the model, which provides a solid base to be extrapolated, but also due to the fact that the HS model defines multiple components which reveal the clear mortality trends of unique age clusters. On the contrary, the unreasonably high error values of the RH and PL models come, to a certain extent, as a result of their unstable behaviour during the fitting process and the limitations of their structure and estimation process. The results for the GR dataset in Table 4.10 show overall a more consistent behaviour for all models. The HS_log was slightly lacking in performance in contrast to the PL and RH models. However, the newly derived HS_prt extension was able to outperform the rest of the models both under the MSPE and the MAPE metrics in the RWD and ARIMA approaches.

Overall, it seems that the HS model, due to its structure, i.e., it is a multiple-component model which reveals unique age clusters, and the notable goodness-of-fit performance achieved by utilising the newly introduced methodology, delivers an informative—and close to the actual data—basis that can lead the out-of-sample process to a more precise prediction.

TABLE 4.9: Results of the percentage error tests for predicted mortality rates of 10 years out-of-sample for well-known stochastic mortality models against HS_prt for E&W dataset.

	MSPE (%)		**MAPE (%)**	
	RWD	**ARIMA**	**RWD**	**ARIMA**
HS_prt	**3.206**	**5.550**	**13.219**	**18.852**
HS_log	3.221	7.135	13.271	20.800
RH	271.861	271.861	58.085	58.085
LC	29.508	29.509	49.838	49.837
PL	586.784	371.845	62.748	55.635
APC	86.494	84.316	46.192	46.131

TABLE 4.10: Results of the percentage error tests for predicted mortality rates of 10 years out-of-sample for well-known stochastic mortality models against HS_prt for GR dataset.

	MSPE (%)		**MAPE (%)**	
	RWD	**ARIMA**	**RWD**	**ARIMA**
HS_prt	8.446	**8.549**	**14.961**	**15.067**
HS_log	8.691	10.390	15.332	15.857
RH	**8.361**	8.581	15.440	15.275
LC	10.398	10.333	20.320	20.964
PL	8.860	11.163	18.402	15.148
APC	10.054	10.077	21.364	21.599

4.4 Discussion and Conclusions

This section offers a discussion over our experimental results and highlights the lessons learnt by the adoption of the various link functions in a model's estimation methods, having witnessed the beneficial impact of the proposed methodology to our model's efficacy.

Our aim was to extend the HS model and investigate its behaviour when formulated in terms of $q_{t,x}$, using generalised linear models and by adopting various link functions. Our motivation originates from similar endeavours in the literature Haberman & Renshaw (1996), Currie (2016) and is aligned with the capabilities offered by well-known tools,

such as StMoMo Villegas et al. (2018)). To the best of our knowledge, it is the first time that a systematic approach is documented for the adoption of this new form of link functions, $F^{-1}(x;\xi,\theta)$ for the mortality modelling domain, along with the necessary transformations to satisfy the condition that the CDF's range should be mapped to the whole real line. We argue that the ability to use various link functions can lead to the better performance of the mortality models, as showcased through the use of the HS model in the context of this chapter.

An additional point to be highlighted is the integration of heavy-tailed distributions (Generalised Pareto in our case). The motivation is two-fold. On the one side, the proposed methodology of Section 4.2 enables the mortality analyst to apply such kinds of distributions in a mortality model in case the intrinsic characteristics of the mortality data fit better to a heavy-tailed distribution profile, i.e., when extreme events must be part of the modelling. In this context, the proposed methodology eases an analyst to adapt a model depending on the data, the time period and the case to be analysed, without being limited only to the off-the-shelf link functions and distributions. Based on our results for the κ_t graphs given in the appendices, one can observe that the long period of the E&W data includes several mortality trend peaks (extreme values) as a result of the mortality during critical events of the 20th century. However, the Greek mortality data and the respective κ_t graphs do not imply the presence of extreme events and outliers. Thus, despite the fact that the improvement on the use of the new form of link functions and different distributions is evident based on the quantitative and qualitative results, the improvement is greater for the E&W for which we finally applied the Generalised Pareto heavy-tailed distribution to better fit to the nature of the data. Even though the rationale behind the use of the heavy-tailed distribution in the mortality realm is valid, we need to note that this outcome needs to be further evaluated. In fact, we aim to continue in this line of research and to explore in our future endeavours the behaviour of such distributions in mortality modelling.

In addition, it has to be noted that the adaptation of a mortality model to the new form of link functions implies an additional step in the model's estimation process, as explained in detail in Section 4.2.2, in order to estimate the ξ and θ parameters. In terms of the codebase, this requires the addition of an extra, but simple routine, which can tackle this problem. Overall, the complexity of the process is slightly increased but we argue that the potential benefit outweighs the added complexity.

The above-mentioned highlights are advocated by our experimental results, while the efficacy of the HS model extensions support the motivation of this work. In fact, the transition to the use of the Binomial distribution has improved the performance of the HS model under the vast majority of the adopted link functions. Notably, the performance improvement for the reference model of the HS_log was even greater in the case of the HS_prt extension for the E&W data, where we noted an improvement of 16.24%, 16.67% and 16.79% for the BIC, AIC and log-likelihood metrics, respectively, while the MAPE improved by 2.7%. In the case of the short GR dataset, the HS_prt extension was the one that outperformed the rest, achieving a similar performance with the HS_beta, proving again that the selection of the correct distribution and link function can further boost the performance of a mortality model even in cases where we need to operate over a poor and non-informative dataset. During our experiments, we applied several distributions, including the Generalised Extreme value, Gumbel, Fréchet, Weibull and Burr, and we experimented with other various datasets (e.g., French data). In order to keep the length of this chapter within reasonable limits, we chose to present the best results. However, our methodology enables an analyst to apply virtually any distribution in the modelling process. In addition, we offered a comparative analysis among the HS extensions and other well-established mortality models of the literature, both under the fitting and forecasting modes. The results showed that the HS_prt achieved the best *goodness-of-fit* performance, while under the forecasting mode, the new HS_prt extension improved the position of the HS and outperformed the rest of the models in terms of prediction accuracy.

In addition, we need to note that the HS improvement through the introduced extensions is not solely reflected in the quantitative results, but most importantly, it is reflected in the qualitative ones. The HS model was able to improve its performance in the identification of more fine-grained age–period and age–cohort components, and thus prove that the proposed methodology retains the model's efficacy and consistency, and it can boost even more its explainability through the attribution of a mortality trend to unique age clusters. This feature is one of the unique characteristics of the HS model in contrast to other models of the literature.

Overall, our application advocates that different mortality data imply the need for a different model transformation in order to increase the *goodness-of-fit*, capture the data dynamics and uncover interesting characteristics in the mortality trend of individual age

clusters. Our results show that even a top-notch mortality model, such as the HS, has room for improvement, as the adaptation with an appropriate link function can lead to even better qualitative and quantitative performance.

Chapter 5

HSTool: A Matlab toolbox for the Hatzopoulos–Sagianou Multiple-Component Stochastic Mortality Model

In this chapter we offer a software implementation of the Hatzopoulos–Sagianou multiple-component stochastic mortality model in the form of a Matlab toolbox called: *HSTool*. The *HSTool* offers the age-period-cohort stochastic mortality model presented in Hatzopoulos & Sagianou (2020), and delivers a practical approach of its -unique of its kind- estimation method. The HS is based on the use of Sparse Principal Component Analysis and Generalised Linear Models, while it is driven by the Unexplained Variance Ratio metric to maximise the captured variance of the mortality data and to regulate the sparsity of the model with the aim of acquiring the optimal number of distinct and significant stochastic components. In the context of this chapter we enhance the original model by introducing codebase-related improvements utilising multi-criteria decision making methods to converge to "optimal" values for critical model's parameters, offering increased tool automation and "plug 'n' play" user experience. The *HSTool* provides the means for fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projections. In this chapter, we offer a complete handbook and we

illustrate some of the capabilities of the toolbox by applying the model to the England and Wales population so that to perform a demonstration on its usage.

5.1 Introduction

Unfolding the qualities of mortality models just on the paper would not have a substantial contribution to the mortality modelling domain. That is, software implementations and packages have been developed and made available to the community for making the use of the models a trivial process for mortality analyst and actuaries. For instance, and as reported in Villegas et al. (2018), the **demography** package Hyndman et al. (2015) capitalises on the Lee & Carter (1992) model and its variants presented in Lee & Miller (2001), Booth et al. (2002) and Hyndman & Shahid Ullah (2007)). The **ilc** package Butt et al. (2014) offers the Renshaw and Haberman model and the Lee–Carter model under a Poisson regression framework. The **LifeMetrics** R functions Coughlan et al. (2007) consider the Cairns–Blake–Dowd (CBD) models and their extensions introduced in Cairns et al. (2009), the Lee–Carter model (using Poisson maximum likelihood), the age–period–cohort model Osmond (1985) Currie (2006) and the Renshaw and Haberman model. **StMoMo** presented in Villegas et al. (2018) and has become a well-established mortality modelling R package which provides the tools for fitting stochastic mortality models, assessing their goodness-of-fit and performing mortality projections. In fact, **StMoMo** encompasses the vast majority of stochastic mortality projection models proposed to date.

In this line of thought, this chapter aims to provide a Matlab implementation of the Hatzopoulos & Sagianou (2020) (HS) multiple-component stochastic mortality model in order to contribute to the domain from a more practical standpoint and add one more tool to the arsenal of mortality analysts. Hence, through this chapter we aim to replicate the methodological innovations that were initially published in the peer-reviewed work of Hatzopoulos & Sagianou (2020) and provide a tool for fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projections. The toolbox reproduces the estimation methodology of the HS model, while it introduces improvements in the code implementation logic that aids the HS model to converge to the optimal model structure, i.e., to identify the most informative age-period and age-cohort components, easier. The qualities of the parameter estimation methodology of

HS, its performance and efficacy have been evaluated over several datasets in Hatzopoulos & Sagianou (2020), and a thorough comparison has been performed, based on diverse qualitative and quantitative evaluation metrics, advocating its beneficial characteristics and surpassing other well-known mortality models, in terms of fitting and forecasting performance. Thus, in order to increase the impact of our work and further contribute to the domain, this chapter aims to the documentation of the codebase of HS and the provision of a technical guide to steer other researchers and professionals on the use of the model in their testbeds, addressing the need of the community for full-fledged tools instead of solely theoretical approaches.

Overall, the following points summarise the contributions of this chapter.

- The offering of a Matlab toolbox, namely *HSTool*, of the HS model presented in Hatzopoulos & Sagianou (2020) supported by the necessary material for the replication of testbeds and use of model's commands. This chapter can be seen as a complete handbook of the *HSTool*.
- Documentation of new codebase-related improvements that contribute to the model's stability and ability to provide a calibrated operation for "plug 'n' play" user experience.
- Introduction of multi-criteria decision making (MCDM) methods in the model's workflow to achieve increased tool automation and convergence to "optimal" values for critical model's parameters.
- Documentation of different model's functions variations to enable more advanced users to carry out experiments by testing different parameterisations given different datasets so that to explore the capabilities of the model and mortality data peculiarities.
- Documentation of the code commands for replicating the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projections.

The rest of this chapter is organised as follows: Section 5.2 documents the code structure of HS, the introduced improvements in its codebase and the details of each available function of the *HSTool*. For each function, Section 5.2 documents the methodology

behind it, the inputs/outputs and indicative execution examples. Finally, Section 5.3 provides a short discussion and concludes. The reader may refer to Section 2.1 to recall the data and notation used and to Section 2.2 for background information for the GLM framework. In addition, the reader can refer to Section 4.1.1, where an outline of the methodology of the HS model is provided, summarising its estimation methodology and its beneficial characteristics.

5.2 HSTool - A Matlab toolbox for HS model

This section provides an overview of the code structure of the HS model, focusing on the core functions that one need to call in order to create a complete mortality analysis testbed. We support the documentation with pseudocodes in order to intuitively steer the user on the use of the tool. Algorithm 3 shows a complete mortality analysis testbed, while the individual commands are further elaborated in the next sections. In addition, in Section 5.2.2 we elaborate on improvements on the codebase of the HS toolbox since its debut back in 2020 in Hatzopoulos & Sagianou (2020).

5.2.1 HS mortality modelling testbed structure

As will be documented in Sections 5.2.3 to 5.2.9, the *HSTool* provides a set of core functions to the user in order to replicate the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projection and simulation. Thus, Algorithm 3 intuitively shows an indicative sequence of commands to be called for this purpose and for supporting even inexperienced users throughout the execution of a complete testbed.

Initially, the data loading takes place. The `hsdataload()` function is executed. The user needs to define the system path where the mortality dataset resides, as well as indexes that indicate the range of years and ages (rows and columns respectively) to be loaded depending the scope of the mortality analysis. Once the data are loaded, they are used as input to the `hsfit()`. The latter command undertakes the complete model construction, including both the estimation of the age-period and age-cohort components, and outputs a set of estimated parameters.The input to the `hsfit()` is the loaded data and the MCDM method that will be used for the definition of a set of the

optimal values of the model and the final optimal structure. This is further explained in Section 5.2.4.

After fitting the model, the `hsevaluate()` undertakes the calculation of the metrics for evaluating the *goodness-of-fit* including all possible metrics. The *HSTool* enables also the execution of mortality forecasting and simulation functions using the `hsforecast()` and `hssimulate()` functions respectively. The former takes as inputs the time series reflecting the mortality rates to be extrapolated along with the corresponding method to be applied (ARIMA or DLR). This function outputs the forecasted time series and the necessary evaluation values. The `hssimulate()` executes a bootstrapping methodology for performing a simulation by extrapolating the mortality rates derived from the `hsforecast()` function. The last step is the execution of the `hsbootstrap()` function in order to analyse the parameter uncertainty.

The documentation of the individual functions of the aforementioned execution flow is given in the following sections. For each function, we document the inputs/outputs, we offer the theoretical background behind each function, while execution examples are provided.

Algorithm 3: Indicative testbed of HS

Input: $file$: Mortality dataset in xls format
$Indexes$: Indexes denoting the range of years and ages to be fitted

Output: $Model$: $(n \times \text{a})$-matrix of $\log(\tilde{m}_{t,x}) = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)} \kappa_t^{(i)} + \sum_{j=1}^{q} \beta_x^{c(j)} \gamma_c^{(j)}$
$Evaluation\ metrics$: Quantitative and qualitative evaluation results
$Forecasted\ mort.\ rates$: Mortality rates extrapolated for h-years ahead.
$Mortality\ paths$: Simulated future mortality paths.
$Param.\ uncertainty$: Parameter uncertainty using bootstrap procedure.

$data \leftarrow$ **hsdataload**$(xlsfile, xlssheet, dataIndexes)$;
$model \leftarrow$ **hsfit**$(data, MCDMmethod)$;
$eval\_metrics \leftarrow$ **hsevaluation**$(model)$;
$forecasted\_rates \leftarrow$ **hsforecast**$(model)$;
$mortality\_paths \leftarrow$ **hssimulate**$(model)$;
$parameter\_uncertainty \leftarrow$ **hsbootstrap**$(model)$;

5.2.2 HS codebase improvements

In this section we elaborate on the improvements which are introduced to the HS model in the context of this chapter. Apart from the implementation and the documentation of the toolbox per se, this chapter aims to enhance the modelling process based on the following additions:

- **Integration of MCDM methods for defining the "optimal" values for critical parameters:**

  In the previous version of HS in Hatzopoulos & Sagianou (2020), the final decision of the appropriate values for some critical parameters was taken by the user based on the results of the conducted analysis. That is, the user had to manually scrutinise the results and conclude to the "optimal" values based on her/his domain knowledge and experience. In order to address this limitation, we adopt MCDM methods in the HS workflow to automate the decision making on that specific parameters and ensure that -if not the optimal- a near "optimal" solution will be selected. MCDM is used for the definition of the followings:

  - Optimal values $k1$ and $k2$ of the orthonormal polynomials.
  - Optimal selection of the "optimal" age-period components for the final model structure.

  More details on the criteria and the MCDM methods are given in Section 5.2.4.

- **Dynamic searching approach of the UVR threshold instead of the definition of an empirical threshold:**

  In Hatzopoulos & Sagianou (2020), an empirical threshold of $UVR(i) < 50\%$ was defined in order to decide whether a age-period or age-cohort component was informative enough to be included in the model structure. In this version of the HS codebase, the selection of the appropriate components is based on a heuristic method of graph analysis for pinpointing the local minima on the $UVR(i)$ figures that reveal the optimal UVR thresholds and the purity of the respective age range reflected in the i candidate age-period or age-cohort component. Thus, the method converges to the optimal components that maximise the captured information of the mortality data revealing clear mortality trends that can be attributed to

specific and unique (per component) age clusters. The previous approach was still able to identify clear age clusters but, it would have accepted any component i having a $UVR(i) < 50\%$. The current approach, forces the search approach to capture a UVR graph starting from the lowest possible value (global minimum), i.e., it can be as more informative as possible, and reach even a quite loose upper level (e.g., $UVR(i) < 60\%$) in an effort to cover a wide search space for the identification of the appropriate components. More specifically, the new approach requires that the identified thresholds of the age cluster reflected in the $UVR(i)$ shall have at least a δ amount of UVR values. Figure 5.1 presents an example of a UVR graph (right) that illustrates a clear age range (ages 30 to 40) having a very low UVR global minimum (0.31). This purity of the UVR curve is reflected in the mortality trend of the respective κ_t (left). The specific process is placed in lines 8-10 of Algorithm 4 (for the estimation of the age-period components) and in lines 5-8 in Algorithm 5.

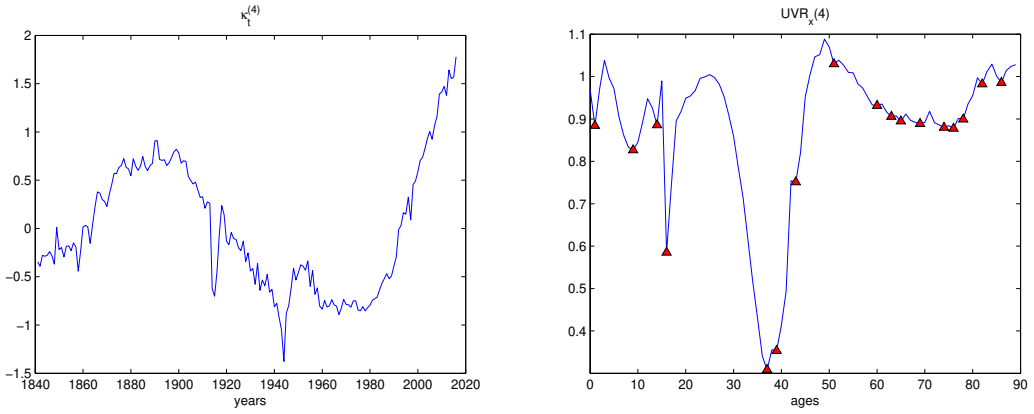

FIGURE 5.1: Plots of $\kappa_t^{(4)}$ and $UVR_x(4)$ for E&W dataset

5.2.3 Mortality data loading

Description The mortality data aimed to be analysed are given as input to the model prior to the execution of the fitting process (See Section 5.2.5). The latter will decompose the mortality rates to the dimensions of age, period and cohort effects. Thus, the *HSTool* handles data as those provided by the Human-Mortality-Database in order to be aligned with the standard practices. The general structure of the `hsdataload()` function is given below.

```
[dtx, etx, mtx, logmtx, datavrbls] = hsdataload (xlsfile, xlssheet,
datafx, datalx, dataft, datalt, fitfx, fitlx, fitft, fitlt);
```

Inputs The input parameters of the `hsdataload()` function are given bellow.

Parameter	Description
`xlsfile`	Points to the xls file in the file system holding the mortality data. The data have the shape illustrated intuitively in Table 5.1.
`xlssheet`	Denotes the excel sheet to be loaded. Normally this factor refers to the male or female data.
`datafx`	First age x in the dataset.
`datalx`	Last age x in the dataset.
`dataft`	First year t in the dataset.
`datalt`	Last year t in the dataset.
`fitfx`	First age x to be selected for model fitting.
`fitlx`	Last age x to be selected for model fitting.
`fitft`	First year t to be selected for model fitting.
`fitlt`	Last year t to be selected for model fitting.

Outputs The output parameters of the `hsdataload()` function are given bellow.

Parameter	Description
`dtx`	Matrix of deaths data.
`etx`	Matrix of observed exposures.
`mtx`	Matrix of crude mortality rates.
`logmtx`	Matrix of logarithmic crude mortality rates.
`datavrbls`	Structure which encapsulates variables and metadata generated during the data loading process.
`datavrbls.ages`	Vector of ages selected to be fitted.
`datavrbls.years`	Vector of the years selected to be fitted.
`datavrbls.cohorts`	Vector of the cohorts to be fitted.
`datavrbls.fitfx`	First age x selected for model fitting.
`datavrbls.fitlx`	Last age x selected for model fitting.
`datavrbls.fitft`	First year t selected for model fitting.
`datavrbls.fitlt`	Last year t selected for model fitting.

Documentation As aforementioned the `hsdataload()` is a core function of the *HSTool* as it is used for loading the mortality data in the analysis testbed. As can be inferred from input variables, the user needs to define the excel file in which the mortality data reside. The data shape is illustrated intuitively in Table 5.1. The dataset columns consist of the *Years*, *Ages*, *Number of Deaths* and the *Exposure to Risk*. In the illustrated example the *Years* are in the range of 1841-2016, the *Ages* from 0 to 110+. In addition, in the `hsdataload()` the user defines the age and year ranges of the dataset

YEARS	AGES	#DEATHS	EX. TO RISK
1841	0	41,525	245,435.36
1841	1	14,014	221,712.36
⋮	⋮	⋮	⋮
1841	110	0	0
1842	0	44,115	239,745.69
1842	1	14,771	221,800.34
⋮	⋮	⋮	⋮
1842	110+	0	0
2016	0	1506	357,650.19
⋮	⋮	⋮	⋮
2016	110+	0.7	0.17

TABLE 5.1: Mortality data shape as provided by Human-Mortality-Database

being loaded, as well as the age and year ranges which will be used in the model's fitting process. Thus, the fitting ages and years must be a subset of the dataset's range.

Given the inputs, the `hsdataload()` undertakes the data curation for the generation of the dtx, etx, mtx, logmtx and datavrbls considering the age and year ranges instructed for the fitting process. The datavrbls encapsulates several supportive variables and metadata generated during the data loading process, as documented above. Those variables can be useful for the user and can be used as input to other functions provided by the *HSTool*.

Example An indicative example of the `hsdataload()` execution is given bellow. The England and Wales dataset for males will be used as the main dataset for all the execution examples throughout this work. The dataset has the following characteristics:

- England & Wales (E&W), Males, calendar years 1841–2016, individual ages 0–89

```
>> [dtx, etx, mtx, logmtx, datavrbls] = hsdataload('E&W_data_(1841 - 2016)',
'males', 0, 110, 1841, 2016, 0, 89, 1841, 2016);

The mortality data have been loaded!

>> dtx
dtx =
   1.0e+04 *
  Columns 1 through 6
    4.1525    1.4014    0.7531    0.5038    0.3627    0.2768
    4.4115    1.4771    0.7592    0.4859    0.3426    0.2602
    4.4550    1.4315    0.7111    0.4716    0.3350    0.2508
```

```
>> etx
etx =
   1.0e+05 *
  Columns 1 through 6
    2.4544    2.2171    2.0367    1.9336    1.9000    1.9386
    2.3975    2.2180    2.1175    1.9816    1.8988    1.8766
    2.4245    2.1674    2.1300    2.0669    1.9553    1.8843

>> mtx
mtx =
  Columns 1 through 6
    0.1692    0.0632    0.0370    0.0261    0.0191    0.0143
    0.1840    0.0666    0.0359    0.0245    0.0180    0.0139
    0.1837    0.0660    0.0334    0.0228    0.0171    0.0133

>> logmtx
logmtx =
  Columns 1 through 6
   -1.7767   -2.7613   -3.2975   -3.6475   -3.9586   -4.2490
   -1.6928   -2.7091   -3.3283   -3.7082   -4.0150   -4.2784
   -1.6942   -2.7174   -3.3997   -3.7803   -4.0667   -4.3192

>> datavrbls
datavrbls =
       ages: [1x90 double]
      years: [1x176 double]
    cohorts: [1x265 double]
      fitfx: 0
      fitlx: 89
      fitft: 1841
      fitlt: 2016
```

5.2.4 Multi-criteria Decision Making

Description This section elaborates on the Multi-criteria Decision Making (MCDM) approaches introduced in the *HStool*. More specifically, the adoption of such methods in the HS codebase extends the methodology presented initially in Hatzopoulos & Sagianou (2020) and contribute to the model's stability and ability to provide a calibrated operation for "plug 'n' play" user experience, increased tool automation and convergence to "optimal" values for critical model's parameters. It must be noted that the MCDM functions are not top-level functions of the toolbox, but they aim to enhance the operation of the `hsfit()` function (See Section 5.2.5) in the background. However, the contribution of these functions to our offering is crucial and, thus, it is vital to document the methodology followed. Note that, as aforementioned in Section 5.2.2, the addition

of MCDM is one of the enhancements in the tool's operational flow. That is, this section is dedicated on how the MCDM methods contribute in the HS model. The general structure of the MCDM function is given below.

```
[index_max] = MCDM (mcdmmethod, decMatrix, w1,w2,w3,w4,w5);
```

Inputs The input parameters of the MCDM functions are given bellow.

Parameter	Description
`mcdmmethod`	The multi-criteria decision making method (SAW or TOPSIS) to be used for the definition of the optimal k-degree of the orthonormal polynomials and the final decision on the optimal p and q components.
`decMatrix`	The decision matrix, i.e., the matrix containing the options (rows) to be evaluated based on the defined criteria (columns) (See Table 5.2).
`w1`	Weight of the MCDM method for Avg.Hypothesis_t criterion (See Table 5.2).
`w2`	Weight of the MCDM method for Avg.Hypothesis_x criterion (See Table 5.2).
`w3`	Weight of the MCDM method for AICc criterion (See Table 5.2).
`w4`	Weight of the MCDM method for AIC criterion (See Table 5.2).
`w5`	Weight of the MCDM method for BIC criterion (See Table 5.2).

Outputs The output parameters of the `MCDM()` functions are given bellow.

Parameter	Description
`index_max`	Value indicating the index of the decision matrix (`decMatrix`) which points to the optimal option for the decision making problem.

Documentation *HStool* implements two well-known MCDM methods namely, the *Simple Additive Weighting* (**SAW**) (Fishburn (1967)) and *Technique for Order of Preference by Similarity to Ideal Solution* (**TOPSIS**) (Hwang & Yoon (1981)), as methods used to solve multi-attribute decision problems. In this way, we aim to provide targeted enhancements to the execution workflow of the HS model in order to solve in an efficient manner some decision making puzzles that was left upon the analyst's shoulders in Hatzopoulos & Sagianou (2020).

A **Simple Additive Weighting (SAW):** SAW can be also found in the literature under the names Weighted Sum Model (WSM) or Weighted Linear Combination (WLC). It was the first method introduced in the MCDM domain and remains the

most widely used one due to its simplicity and effectiveness. In problems where m options exist and n criteria drive the decision, the SAW scores the options using the following expression:

$$Score_{saw} = \max_i \sum_{j=1}^{n} \alpha_{ij} w_j, \qquad for \qquad i = 1, 2, 3, \ldots, m \tag{5.1}$$

$Score_{saw}$ is the best (max) among the m options, α_{ij} is the normalised value of the i-th alternative in terms of the j-th criterion, and w_j is the weight of importance of the j-th criterion. In single-dimensional cases, where all the units are the same, the use of SAW is straightforward. In multi-attribute decision making a normalisation procedure is used for transforming the different measurement units of attributes to a comparable unit (Triantaphyllou (2000)). Thus, α_{ij} needs to be normalised. Among various normalisation techniques that can be found in the literature, in this work we use the following ones for transforming the beneficial and cost criteria.

- Benefit criteria: $\alpha_{ij} = \frac{r_{ij}}{r_{max}}$
- Cost criteria: $\alpha_{ij} = 1 - \frac{r_{ij}}{r_{max}}$

where r_{ij} is the actual value of the i-th alternative in terms of the j-th criterion and r_{max} is the max value among the values of the j-th criterion. The w_j are weights assigned to regulate the decision process. The analyst may chose to treat all criteria equally or to boost some others by assigning a higher w_j value.

B **Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS):** The basic concept of this method is that the selected alternative should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution in some geometrical sense. The TOPSIS method assumes that each criterion has a tendency of monotonically increasing or decreasing utility. Therefore, it is easy to define the ideal and negative-ideal solutions. The Euclidean distance approach was proposed to evaluate the relative closeness of the alternatives to the ideal solution. Thus, the preference order of the alternatives can be derived by a series of comparisons of these relative distances.

The TOPSIS method considers a decision matrix (DM) having m alternatives in its rows which are evaluated in terms of n criteria. x_{ij} denotes the performance measure of the i-th alternative in terms of the j-th criterion. TOPSIS takes specific

steps until the final ranking of alternatives: (i) the construction of a normalised DM, (ii) Construction of the Weighted Normalised DM, (iii) Determination of the ideal and negative-ideal solutions, (iv) Calculation of the separation measure, (v) Relative Closeness to the Ideal Solution, and (vi) Rank the Preference Order.

The normalisation converts the various criteria dimensions into non-dimensional criteria. An element r_{ij} of the normalised DM is calculated as follows using the notion of the Euclidean distance:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{kj}^2}}$$

Given the normalised r_{ij}, the Weighted Normalised DM V is generated by multiplying each r_{ij} with a user-defined weight ($W = (w_1, w_2, \ldots, w_n)$, (where: $\sum w_j = 1$)). The Weighted Normalised matrix V has the following form:

$$V = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \ldots & w_n r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \ldots & w_n r_{2n} \\ . & & & . \\ . & & & . \\ . & & & . \\ w_1 r_{m1} & w_2 r_{m2} & \ldots & w_n r_{mn} \end{bmatrix}$$

Then, the determination of the ideal (A^*) and negative-ideal (A^-) solutions follows, based on the following expressions:

$$A^* = \{(\max_i \nu_{ij} | j \in J), (\min_i \nu_{ij} | j \in J'),\ i = 1, 2, 3, \ldots, m\} = \{\nu_{1^*}, \nu_{2^*}, \ldots, \nu_{n^*}\}$$

$$A^- = \{(\min_i \nu_{ij} | j \in J), (\max_i \nu_{ij} | j \in J'),\ i = 1, 2, 3, \ldots, m\} = \{\nu_{1^-}, \nu_{2^-}, \ldots, \nu_{n^-}\}$$

where: $J = \{j = 1, 2, 3, \ldots, n$ and j is associated with the benefit criteria$\}$,
$J' = \{j = 1, 2, 3, \ldots, n$ and j is associated with the cost/loss criteria$\}$.

Given the above, the separation measure can be calculated as follows:

$$S_{i^*} = \sqrt{\sum_{j=1}^{n} (\nu_{ij} - \nu_{j^*})^2}, \qquad for \qquad i = 1, 2, 3, \ldots, m \tag{5.2}$$

$$S_{i^-} = \sqrt{\sum_{j=1}^{n}(\nu_{ij} - \nu_{j^-})^2}, \qquad for \qquad i = 1, 2, 3, \ldots, m \tag{5.3}$$

where, S_{i^*} is the distance (in the Euclidean sense) of each alternative from the ideal solution, and S_{i^-} is the distance (in the Euclidean sense) of each alternative from the negative-ideal solution.

Given equations 5.2 and 5.3, the relative closeness of an alternative A_i to the ideal solution can be calculated as follows:

$$C_{i^*} = \frac{S_{i^-}}{S_{i^*} - S_{i^-}} \tag{5.4}$$

where $1 \geq C_{i^*} \geq 0$, and $i = 1, 2, 3, \ldots, m$. Apparently, $C_{i^*} = 1$, if $A_i = A^*$, and $C_{i^-} = 0$, if $A_i = A^-$.

Finally, the ranking of all alternatives will reveal the best one that has the shortest distance to the ideal solution C_{i^*}.

Having documented the theoretical background, we elaborate on the placement of the methods in the *HSTool* workflow. More specifically, the MCDM methods contribute in solving the following two problems:

- The identification of the "optimal" degree of $k1$ and $k2$ of the orthonormal polynomials (See Algorithm 4, Lines 2-3 and Algorithm 5, Lines 2-3).
- The identification of the "optimal" set of p components, which capture most of the variance of the mortality data and will be incorporated in the model structure. (See Algorithm 4, Line 12)

In both cases, multiple criteria are taken into consideration in an effort to decide which are the best solutions, given a wide set of information criteria and statistical tests. Table 5.2 provides an example of the decision matrix that illustrates how the different options and the criteria look like. As can be seen, several criteria are used based on the use of Information Criteria and statistical hypothesis tests. More specifically, the information criteria used are AIC, its modification AIC, and the BIC. Their calculation is a straightforward process and is further discussed in Section 5.2.6.

Avg.Hypothesis_t (%)	Avg.Hypothesis_x (%)	AICc	AIC	BIC
39.33	66.33	1,4263.47	14,252.35	15,239.95
35.5	64	15,600.76	15,589.64	16,577.24
⋮	⋮	⋮	⋮	
57.5	73.5	10,229.27	10,205.72	11,635.81

TABLE 5.2: Instance of decision matrix of the multiple criteria decision making methods

The Avg.Hypothesis_t and Avg.Hypothesis_x indicate the average of the percentages of the results of a set of statistical hypothesis tests which are calculated over the residuals for the dimensions of t (years) and x (ages), respectively. More specifically, the following six statistical tests are used:

- Ljung–Box (Ljung & Box (1978))
- Sign (Dixon & Mood (1946))
- Runs (Bradley & Bradley (1968))
- Kolmogorov-Smirnov (Massey (1951))
- Engle's ARCH (Engle (1982))
- Jarque-Bera (Jarque & Bera (1987))

Given this, the average of the percentages of the test decision for the non-rejected null hypothesis, for each statistical test, is calculated. Each test, evaluates the residuals in order to check whether the results of the fitting process meets a desired quality (e.g., the Runs test returns a decision for the null hypothesis that the values in a data vector come in random order). Thus, the average over all tests provide an overview on the quality of the fitting process over the acquired residuals. That is, the Avg.Hypothesis_t and Avg.Hypothesis_x take part in the MCDM to steer the decision making towards the "optimal" result.

5.2.5 Model Fitting

Description The *HSTool* offers the `hsfit()` function for fitting the stochastic mortality model to the loaded data. This function undertakes the execution of the parameters estimation method, following the steps described in Section 4.1.1. Two function variations are given to enable more advanced users to carry out experiments by testing different parameterisations so that to explore the capabilities of the model and the mortality data peculiarities. That is, the first function bellow unfolds all the available input

parameters that a user may wish to modify, while the second offers a parsimonious function invocation in which the necessary parameters are set to default values. Those values have been defined by us, based on our domain knowledge and HS modelling experience in order to offer a calibrated operation. The general structure of the `hsfit()` function is given below.

```
[lgmtxhat, mtxhat, ax, betax, kappat, betac, gammac, fitvrbls] =
hsfit (dtx, etx, logmtx, datavrbls, init_s, last_s, stp_s, maxiter, info,
gapchange, mcdmmethod, w1, w2, w3, w4, w5);
```

```
[lgmtxhat, mtxhat, ax, betax, kappat, betac, gammac, fitvrbls] =
hsfit (dtx, etx, logmtx, datavrbls);
```

Inputs The input parameters of the `hsfit()` function are given bellow.

Parameter	Description
`dtx`	Matrix of deaths data.
`etx`	Matrix of observed exposures.
`logmtx`	Matrix of logarithmic crude mortality rates.
`datavrbls`	Structure which encapsulates variables and metadata generated during the data loading process.
`init_s`	Initial value of sparsity factor S.
`last_s`	Last value of sparsity factor S.
`stp_s`	Step of changing the S iterative.
`maxiter`	Maximum number of iterations to be executed in DSPCA function which performs the Sparse Principal Component Analysis (See Section 5.2.10).
`info`	Parameter to tune the DSPCA reporting verbosity.
`gapchange`	Required change in gap from first gap in DSPCA (default: 1e-4).
`mcdmmethod`	The multi-criteria decision making method (SAW or TOPSIS, see Section 5.2.4) to be used for the definition of the optimal k-degree of the orthonormal polynomials and the final decision on the optimal p and q components.
`w1`	Weight of the MCDM method for Avg.Hypothesis_t criterion (See Section 5.2.4, Table 5.2).
`w2`	Weight of the MCDM method for Avg.Hypothesis_t criterion (See Section 5.2.4, Table 5.2).
`w3`	Weight of the MCDM method for AICc criterion (See Section 5.2.4, Table 5.2).
`w4`	Weight of the MCDM method for AIC criterion (See Section 5.2.4, Table 5.2).
`w5`	Weight of the MCDM method for BIC criterion (See Section 5.2.4, Table 5.2).

Outputs The output parameters of the `hsfit()` function are given bellow.

Parameter	Description
`lgmtxhat`	Matrix of the logarithmic estimated mortality rates.
`mtxhat`	Matrix of estimated mortality rates.
`ax`	Vector with the estimated values of age effect which represents the main age profile of mortality α_x.
`betax`	Matrix with the estimated values of each age-related effect $\beta_x^{(i)}$ for each period component, where $i = 1, 2, \ldots, p$.
`kappat`	Matrix with the estimated values of each period-related effect $\kappa_t^{(i)}$, where $i = 1, 2, \ldots, p$.
`betac`	Matrix with the estimated values of each age-related effect $\beta_x^{c(j)}$ for each cohort component, where $j = 1, 2, \ldots, q$.
`gammac`	Matrix with the estimated values of each cohort related effect $\gamma_c^{(j)}$, where $j = 1, 2, \ldots, q$, and $c = t - x$.
`fitvrbls`	Structure which encapsulates variables generated during the HS fitting process.
`fitvrbls.s`	Value of the sparsity factor of SPCA.
`fitvrbls.k1`	Degree of the orthonormal polynomial used for the estimation of the age-period components.
`fitvrbls.k2`	Degree of the orthonormal polynomial used for the estimation of the age-cohort components.
`fitvrbls.p`	Number of age-period components incorporated in the model.
`fitvrbls.q`	Number of age-cohort components incorporated in the model.
`fitvrbls.uv_ratio`	Matrix with the UVR values of each age-period component, $UVR_x(i)$, where $i = 1, 2, \ldots, p$.
`fitvrbls.uvc_ratio`	Matrix with the UVR values of each age-cohort component, $UVR_x^c(j)$, where $j = 1, 2, \ldots, q$.
`fitvrbls.ttl_lglklh`	Total log-likelihood of the model.
`fitvrbls.nobs`	Number of observations.
`fitvrbls.npar`	The number of parameters being estimated in HS.

Documentation The estimation process of the HS model has been thoroughly documented in Hatzopoulos & Sagianou (2020) while in Section 4.1.1 we provided an overview.

The goal of the `hsfit()` function is to define the p age-period and q age-cohort components to be included in the model structure. The `hsfit()` function takes as input the number of deaths, the exposures and the logarithmic crude mortality rates. Those inputs are readily created using the `hsdataload()` function as exemplified in Section 5.2.3. Algorithms 4 and 5 summarise the code-related procedures executed in the context of

the `hsfit()` function. The improvements made in the HS codebase are abstractly highlighted with blue color in the algorithms. The documentation does not aim to thoroughly describe the estimation method, but to highlight the points where the method has been enhanced with the additions reported in Section 5.2.2.

As described in Step#1 in Section 4.1.1, the HS model applies GLM in order to generate the GLM-estimated parameters $B = \{b_{t,k-1}\}$, $k = 1, \ldots, k_1$. To do so, we need first to define the optimal degree k_1 of the orthonormal polynomials, as those generated using the Gram–Schmidt process. In Hatzopoulos & Sagianou (2020) the user had to decide the degree k_1 based on the BIC values of the GLM fitted model. To enhance the process, we adopted the MCDM methods presented in Section 5.2.4 in order to increase the tool's automation level and take an informed decision on the "optimal" k_1 degree. Thus, the degree is determined based on the information criteria AIC, AICc and BIC and the average result of the hypothesis testing of a set of statistical tests. Note that, this improvement is applied also for the case of the k_2 degree for the age-cohort estimation.

Another enhancement in the model's estimation process is the addition of a dynamic approach for the definition of the UVR threshold, instead of the use of an empirical one. As explained in Section 5.2.2, the process is now based on the analysis of the UVR curve, finding the lowest possible UVR value (global minimum) of the curve, and then searching even a quite loose upper level (e.g., $UVR(i) < 60\%$) in an effort to cover a wide search space for the identification of the appropriate components. This addition leads to a more efficient implementation contrary to the Hatzopoulos & Sagianou (2020), making sure the most of the data variance will be captured by the model.

The last enhancement, but the most important one, is the addition of MCDM for deciding the final and optimal model structure. More specifically, as denoted in step#3 in Section 4.1.1, the definition of the optimal model structure coincides with the definition of the optimal scalar s. That is, the estimation method of HS examines iteratively a range of s values, and each one can lead potentially to a different model structure (for the p age-period components). Those are treated as candidate solutions. That is, MCDM is used in this work for taking the decision on the "optimal" set of age components to be incorporated in the final model structure. In Hatzopoulos & Sagianou (2020) the user had to scrutinise manually the results, having BIC as an indicator, to conclude to the optimal components. This process has a full degree of automation, as the optimal model

structure is now being determined based on the information criteria AIC, AICc and BIC and the average result of the hypothesis testing of a set of statistical tests. Note that, this improvement is applied only for the case of the p age-period components, as the estimation of the q age-cohort components is straightforward due to the use of PCA.

Example Using again the E&W data as the basis, the inputs for the `hsfit()` function have been created using the `hsdataload()` function as exemplified in Section 5.2.3. Then, `hsfit()` is called as follows:

```
>> [lgmtxhat, mtxhat, ax, betax, kappat, betac, gammac, fitvrbls] =
hsfit (dtx, etx, logmtx, datavrbls);
```

The fitting process started....

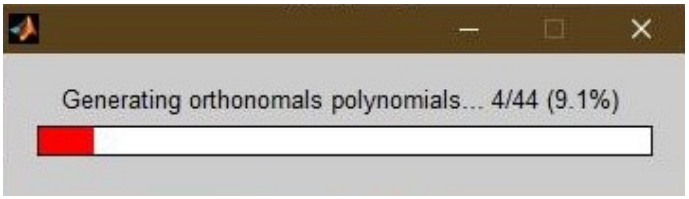

FIGURE 5.2: Generation of orthonormal polynomials

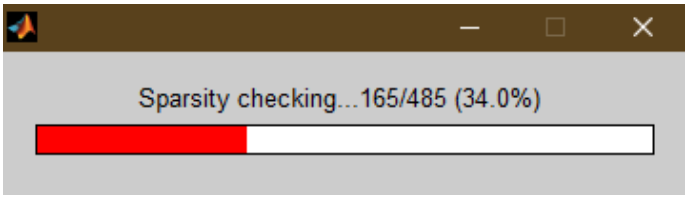

FIGURE 5.3: Checking sparsity factor s

The fitting process has been completed!

```
>> fitvrbls
fitvrbls =
             s: 54.2900
            k1: 34
            k2: 8
             p: 5
             q: 2
      uv_ratio: [90x5 double]
     uvc_ratio: [90x2 double]
    ttl_lglklh: -1.3665e+05
          nobs: 15840
          npar: 1640

>> plot(datavrbls.ages, betax(4,:))
title('\beta^{(4)}_x')
```

```
xlabel('ages')

>> plot(datavrbls.years, kappat(:,4))
title('\kappa^{(4)}_t')
xlabel('years')

>> plot(datavrbls.ages, fitvrbls.uv_ratio(:,4))
title('UVR_x(4)')
xlabel('ages')

>> plot(datavrbls.ages, betac(1,:))
title('\beta^{c(1)}_x')
xlabel('ages')

>> plot(datavrbls.cohorts, gammac(:,1))
title('\gamma^{(1)}_c')
xlabel('cohorts')

>> plot(datavrbls.ages, fitvrbls.uvc_ratio(:,1))
title('UVR^c_x(1)')
xlabel('ages')
```

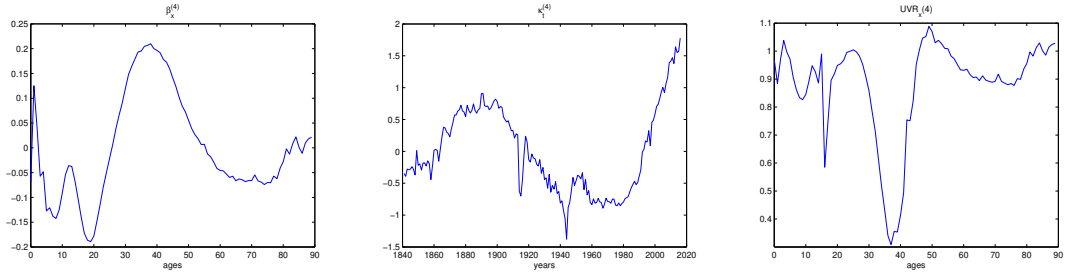

FIGURE 5.4: Plots of $\beta_x^{(4)}$, $\kappa_t^{(4)}$ and $UVR_x(4)$

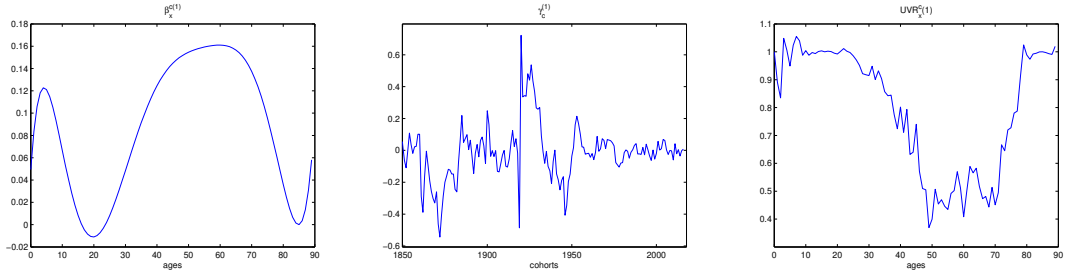

FIGURE 5.5: Plots of $\beta_x^{c(1)}$, $\gamma_c^{(1)}$ and $UVR_x^c(1)$

5.2.6 Goodness-of-fit analysis

Description The evaluation of the *goodness-of-fit* of the HS model on a given dataset is based on a set of quantitative and qualitative criteria in order to approach the analysis from different angles. To do so, the *HSTool* provides the `hsevaluate()` function to provide the evaluation results for all the selected quantitative metrics and the

`hsresiduals()` in order to calculate and plot the residuals of the fitted model. When it comes to the qualitative results, we use the native Matlab command `plot()` to visualise the outputs, as those generated from the `hsfit()` function. More specifically, as part of the `fitvrbls` object, we can visualise the $UVR(i)$ and $UVR^c(j)$ values, which are used to illustrate the purity of a model's component, as well as the age cluster which is attributed to the identified mortality trend revealed by each corresponding $\kappa_t^{(i)}$ and $\gamma_c^{(j)}$.

Therefore, the following criteria are supported in *HSTool*:

- Quantitative criteria:
  - Information Criteria: Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc).
  - Percentage error tests: Mean Squared Percentage Error (MSPE) and the Mean Absolute Percentage Error (MAPE).
  - Explanation Ratio (ER): Used for quantifying the total contribution obtained by each individual period or cohort component.
- Qualitative Critεria:
  - Unexplained Variance Ratio (UVR): Used for revealing the magnitude of the captured information by each age–period and age–cohort component.
  - Residuals: Identification of regular patterns for checking whether a model can describe all the features of the data appropriately.

The general structure of the `hsevaluate()` and `hsresidual()` functions are given below.

```
[aic, aicc, bic, mape, mspe, er] = hsevaluate(mtxhat, mtx, logmtx, ax,
kappat, betax, gammac, betac, fitvrbls);

[sdrestx, dev] = hsresiduals(dtx, etx, lgmtxhat, datavrbls,
fitvrbls, 'visualise');
```

Inputs The input parameters of the `hsevaluate()` and `hsresiduals()` functions are given bellow.

Parameter	**Description**
`hsevaluate()`	
`mtxhat`	Estimated mortality rates.
`mtx`	Matrix of crude mortality rates.
`logmtx`	Matrix of the logarithmic estimated mortality rates.
`ax`	Vector with the estimated values of age effect which represents the main age profile of mortality α_x.
`kappat`	Matrix with the estimated values of each period-related .
`betax`	Matrix with the estimated values of each age-related effect $\beta_x^{(i)}$ for each period component, where $i = 1, 2, \ldots, p$.
`gammac`	Matrix with the estimated values of each cohort related effect $\gamma_c^{(j)}$, where $j = 1, 2, \ldots, q$, and $c = t - x$.
`betac`	Matrix with the estimated values of each age-related effect $\beta_x^{c(j)}$ for each cohort component, where $j = 1, 2, \ldots, q$.
`fitvrbls`	Structure which encapsulates variables generated during the HS fitting process.
`hsresiduals()`	
`dtx`	Matrix of deaths data.
`etx`	Matrix of observed exposures.
`lgmtxhat`	Matrix of the logarithmic estimated mortality rates.
`datavrbls`	Structure which encapsulates variables generated during the data loading process.
`fitvrbls`	Structure which encapsulates variables generated during the HS fitting process.
`visualise`	Option to instruct whether the function will generate a plot ("on") or not ("off").

Outputs The output parameters of the `hsevaluate()` and `hsresiduals()` functions are given bellow.

Parameter	**Description**
`hsevaluate()`	
`aic`	Akaike Information Criterion.
`aicc`	Corrected Akaike Criterion.
`bic`	Bayesian Information Criterion.
`mape`	Mean Absolute Percentage Error.
`mspe`	Mean Squared Percentage Error.
`er`	Explained Variance.
`hsresiduals()`	
`sdrestx`	Matrix of standardised deviance residuals.
`dev`	Matrix of deviance residuals.

Documentation This paragraph documents the theoretic background and scope of the various metrics used to evaluate the *goodness-of-fit* of the model.

Information Criteria is a common practice in the domain for the evaluation of mortality models. AIC, AICc and BIC, are defined as follows:

$$\text{AIC} = -2l + 2\text{npar}, \text{ and its correction } \text{AICc} = \text{AIC} + \frac{2\text{npar}(\text{npar}+1)}{N - \text{npar} - 1}$$

and

$$\text{BIC} = -2l + \text{npar} \log N$$

where N is the number of observations (`fitvrbls.nobs`), l is the log-likelihood (`fitvrbls.ttl_lglklh`) and npar is the number of parameters (`fitvrbls.npar`) being estimated in HS, and is calculated based on the following formula:

$$\text{npar} = 2 + (k_1 + k_2) + p * (n + k_1) + q * (n_c + k_2)$$

where n is the index of last calendar year (i.e., $t = t_1, \ldots, t_n$) and n_c is the index of the last cohort (i.e., $c = c_1, \ldots, c_{n_c}$).

Percentage error tests are used to measure the difference between the estimator and what is estimated, so that to measure the fitting quality of an estimator. For all ages from x_1 to x_a and for all years t_1 to t_n, these measures are defined as follows:

$$\text{MSPE} = \frac{1}{t \cdot x} \sum_t \sum_x \left(\frac{\hat{m}_{t,x} - m_{t,x}}{m_{t,x}} \right)^2 \text{ and } \text{MAPE} = \frac{1}{t \cdot x} \sum_t \sum_x \frac{|\hat{m}_{t,x} - m_{t,x}|}{m_{t,x}}$$

where, $\hat{m}_{t,x}$ is the estimated *central mortality rate* and $m_{t,x}$ is the observed *crude central mortality rate*.

Explanation Ratio (ER) is used in order to quantify the total contribution obtained by each individual period or cohort component. Considering the whole age range defined in the `hsdataload()`, we calculate the explained variance of each component including all ages. Particularly, we calculate the sum of the variances of the log crude mortality central rates for all ages, and the sum of the variances of the error between the estimated

log mortality rates considering the i-component, and the log crude mortality rates, for all ages. Then, one minus the ratio of these two variances defines the ER(i):

$$\mathrm{ER}(i) = 1 - \frac{\sum_x \mathrm{Var}\left(\log(\tilde{m}^{(i)}_{t,x}) - \log(m_{t,x})\right)}{\sum_x \mathrm{Var}\big(\log(m_{t,x})\big)} \tag{5.5}$$

where, $\tilde{m}^{(i)}_{t,x}$ denotes the estimated mortality rates including the i–component for $i = 1, 2, \ldots, p, p+1, \ldots, p+q$, where p and q correspond to the number of period and cohort components respectively. Each ER(i) values, $i = 1, \ldots, p+q$, gives the magnitude of the contribution of the i component.

Unexplained Variance Ratio is used in order to capture and visualise the significant amount of variance captured by each of the model's components, and to illustrate the age clusters where this variance corresponds to, easing the attribution of a mortality trend to a specific age cluster. Hence, UVR gives a qualitative overview of each model structure. For more details on the definition of the UVR metric we prompt the interested reader to refer to Hatzopoulos & Sagianou (2020). Note that the UVR figures and the corresponding Matlab commands to visualise them were documented in the examples of Section 5.2.5.

Residuals Another common measure of the *goodness-of-fit* of a mortality model is the residuals of the fitted model. By detecting regular patterns in the visualised residuals, one can identify if the model can describe all the features of the data appropriately. According to Villegas et al. (2018), with a Poisson random component, it is appropriate to look at the scaled deviance residuals defined as:

$$\mathrm{dev}_{t,x} = 2\left[d_{t,x}\log\left(\frac{d_{t,x}}{e_{t,x} * \hat{m}_{t,x}}\right) - \Big(d_{t,x} - \big(e_{t,x} * \hat{m}_{t,x}\big)\Big)\right]$$

$$r_{t,x} = \mathrm{sign}\Big(d_{t,x} - \big(e_{t,x} * \hat{m}_{t,x}\big)\Big)\sqrt{\frac{\mathrm{dev}_{t,x}}{\hat{\phi}}} \quad \text{and} \quad \hat{\phi} = \frac{\sum_t \sum_x \mathrm{dev}_{t,x}}{N - \mathrm{npar}}$$

Example Using again the E&W data as the basis, the `hsevaluate()` and the `hsresiduals()` functions are called as indicated below. In addition, execution examples using the native Matlab `plot()` function are given for visualising the qualitative results. Note that, the `hsdataload()` and the `hsfit()` need to be executed beforehand.

```
>>[aic, aicc, bic, mape, mspe, er] = hsevaluate(mtxhat, mtx, logmtx, ax,
kappat, betax, axc, gammac, betac, fitvrbls);

>> aic
aic =
   2.7657e+05
>> aicc
aicc =
   2.7695e+05
>> bic
bic =
   2.8915e+05
>> mape
mape =
    0.0378
>> mspe
mspe =
    0.0035
>> er
er =
    0.9272    0.0352    0.0255    0.0060    0.0015    0.0006    0.0001

>> [sdrestx_final, dev] = hsresiduals(dtx, etx, lgmtxhat, datavrbls,
fitvrbls,'on');
```

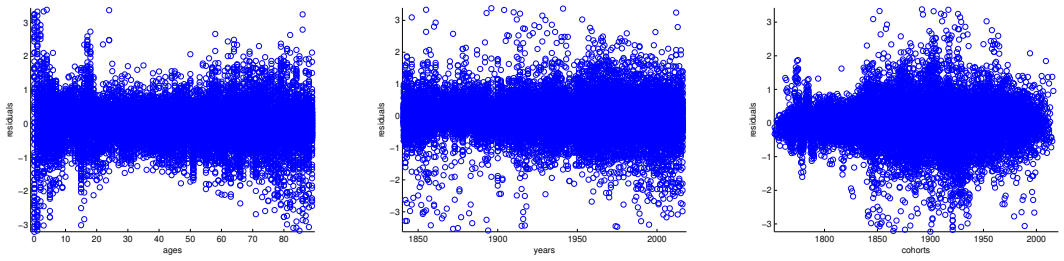

FIGURE 5.6: Scatter plots of deviance residuals for the HS model fitted to the E&W data for ages 0-89 and the period 1841-2016.

5.2.7 Mortality Forecasting

Description Forecasting mortality rates is a typical approach in the modelling testbeds as a logical step after the decomposition of the mortality rates in the dimensions of age, period and cohort. The mortality dynamics are reflected by the $\kappa_t^{(i)}$ terms which are treated as time series that can be extended years ahead. Thus, *HSTool* provides the relevant function `hsforecast()` in order extrapolate the mortality rates including the estimation of the respective confidence intervals. To do so, *HSTool* utilises Dynamic Linear Regression (DLR) or ARIMA models. DLR models is class of time series forecasting and for its realisation the Captain Toolbox for Matlab (Young et al. (2009))

is used. In the ARIMA case, the toolbox deploys a routine for the determination of the best ARIMA model considering the best BIC performance acquired among all the (p,d,q) combinations. The `hsforecast()` function receives as input, among others, the estimated values of a period index, as has been created from the `hsfit()` function, and forecasts the index for h–steps (years) ahead.

The general structure of the `hsforecast()` function is given below:

```
[lgmtx_for, mtx_for, kappat_for,uci_for, lci_for, forecastvrbls] =
hsforecast(h,'method_for', kappat, betax, ax, datavrbls, alpha, 'visualise');
```

Inputs The input parameters of the `hsforecast()` function is given bellow.

Parameter	Description
`h`	Years ahead to be forecasted.
`method_for`	Method to be used for forecasting. It can be "DLR" or "ARIMA".
`kappat`	Matrix with the estimated values of each period index $\kappa_t^{(i)}$, where $i = 1, 2, \ldots, p$.
`betax`	Matrix with the estimated values of each age-related effect $\beta_x^{(i)}$ for each period component, where $i = 1, 2, \ldots, p$.
`ax`	Vector with the estimated values of age effect which represents the main age profile of mortality α_x.
`datavrbls`	Structure which encapsulates variables generated during the data loading process.
`alpha`	Significance level 0.05 (default) — scalar value in the range 0 to 1.
`visualise`	Option to instruct whether the function will generate a plot ("on") or not ("off").

Outputs The output parameters of the `hsforecast()` function is given bellow.

Parameter	Description
`lgmtx_for`	Matrix of the forecasted logarithmic mortality rates.
`mtx_for`	Matrix of forecasted mortality rates.
`kappat_for`	Matrix of the forecasted values of each period index $\kappa_t^{(i)}$.
`uci_for`	Matrix of the upper bounds of the confidence interval.
`lci_for`	Matrix of the lower bounds of the confidence interval.
`kappat_for`	Matrix of the forecasted values of each period index $\kappa_t^{(i)}$.
`forecastvrbls`	Structure which encapsulates variables generated during the forecasting process.
`forecastvrbls.dlr_se`	Standards error of the fit using DLR process.
`forecastvrbls.dlr_model`	The ϕ–parameters of the DLR model.
`forecastvrbls.arima_mse`	Mean Square Error from the ARIMA process.
`forecastvrbls.arima_model`	The best ARIMA model (p,d,q).

Documentation As aforementioned *HSTool* utilises Dynamic Linear Regression (DLR) or Best ARIMA model for mortality forecasting, while it generates the respective confidence intervals.

- **Dynamic Linear Regression (DLR):** Specifically for the DLR models, the HSTool utilises a specific class:

$$Y_t = a + \delta_t \cdot t + \epsilon_t, \tag{5.6}$$

where Y_t denotes the period dynamics, for each calendar year t, with the slope being a stochastic time variable parameter that follows a smoothed random walk process: $\Delta\delta_t = \phi \cdot \Delta\delta_{t-1} + \zeta_{t-1}$, $0 < \phi \leq 1$ and $\Delta\delta_t = \delta_t - \delta_{t-1}$ denotes the difference operation. The innovations ϵ_t and ζ_t are assumed to be white noise random variables. If $\phi < 1$, Y_t is being modelled as a linear stochastic variable having a slightly tilted s-shape for the short–medium forecasts and also smooth progression to the mortality dynamics. If $\phi = 1$, Y_t is being modelled as a non–linear stochastic variable giving either an accelerating or a decreasing mortality improvement. Experiments with various mortality experiences have shown that the time related non–stationary SPCs can be represented adequately under this particular DLR model structure. Therefore, these two nested DLR structures ($\phi < 1$ and $\phi = 1$) are compared based on the BIC of the mortality model, in order to choose the most appropriate between the two for each period index. According to Kass & Raftery (1995), one can consider a model selection based on BIC to be roughly equivalent to a model selection based on a Bayes factor. Thus, by applying the two DLR structures we can have two variations of the same model. Then, we can select the appropriate DLR structure based on the observed BIC difference between these two variations of the model. The difference of the BIC values is defined as, $\Delta\text{BIC} = \text{BIC}(i) - \text{BIC}(b)$, where $\text{BIC}(b)$ denotes the BIC value for the "best" model variation, where the "best" is the one having the lower BIC value and $\text{BIC}(i)$ denotes the BIC value for the alternative model variation. In this direction, Kass & Raftery (1995) suggested the following rules of thumb:

  - If $\Delta\text{BIC} \leq 2$, there is no clear evidence against or in favour of the two models.
  - If $2 < \Delta\text{BIC} \leq 6$, one can say that there is positive evidence against model i, i.e., there is a difference in favour of model b.

- If $6 < \Delta\text{BIC} \leq 10$, there is a strong evidence against model i.
- If $\Delta\text{BIC} > 10$, there is a very strong evidence against model i, and model b is by far the most appropriate.

In the context of our work, we examine whether $\Delta\text{BIC} > 2$ in order to judge upon the most appropriate DLR structure.

- **ARIMA:** Regarding the projection using Best ARIMA, each i–period index is modeled considering all the relevant ARIMA(p,d,q) models for p, d, q = 0–3. Thus, the tool applies a routine for the creation of the models and the evaluation based on BIC performance for each (p,d,q) combination.

  To do so, we assume that $\kappa_t^{(i)}$ period indexes follow a general univariate ARIMA(p_i, q_i, d_i) with drift, so that:

$$\Delta^{d_i}\kappa_t^{(i)} = \delta_0^{(i)} + \phi_1^{(i)}\Delta^{d_i}\kappa_{t-1}^{(i)} + \cdots + \phi_{p_i}^{(i)}\Delta^{d_i}\kappa_{t-p_i}^{(i)} + \xi_t^{(i)} + \delta_1^{(i)}\xi_{t-1}^{(i)} + \cdots + \delta_{q_i}^{(i)}\xi_{t-q_i}^{(i)}, \quad (5.7)$$

  where Δ is the difference operator, $\delta_0^{(i)}$ is the drift parameter, $\phi_1^{(i)} \ldots \phi_{p_i}^{(i)}$ are the autoregressive coefficients with $\phi_{p_i} \neq 0$, $\delta_1^{(i)}, \ldots, \delta_{q_i}^{(i)}$, are the moving average coefficients with $\delta_{q_i}^{(i)} \neq 0$ and $\xi_t^{(i)}$ is a Gaussian white noise process with variance $\sigma^{2(i)}\xi$.

The time series model in Equation 5.6 and 5.7 can be used to obtain projected (simulated) values of a period index $\kappa_{t_n+h}^{(i)}$ to derive forecasted (simulated) values of the predictor

$$\log(\tilde{m}_{t+h,x}) = \alpha_x + \sum_{i=1}^{p} \beta_x^{(i)}\kappa_{t+h}^{(i)} \quad (5.8)$$

Example The forecasting example shown below is based on the E&W data used before in the `hsfit()`. We demonstrate the use for the case of the ARIMA model for 50 years ahead projection. In addition, the native Matlab `plot()` function is used for the visualisation of the foretasted $\kappa_t^{(i)}$ period indexes and the mortality rates.

```
>> [lgmtx_for, mtx_for, kappat_for,uci_for, lci_for, forecastvrbls] =
hsforecast(50, 'ARIMA', kappat, betax, ax, datavrbls, 0.05, 'on');
```

```
forecastvrbls =
    arima_model: '(0,1,0)(3,2,0)(0,2,0)(0,1,0)(0,1,0)'
         dlr_se: []
```

```
      dlr_model: []
      arima_mse: [50x5 double]

>>kappat_for(:,1)'
ans =
  Columns 1 through 6
   14.8880   15.0229   15.1579   15.2929   15.4279   15.5628
  .....
  Columns 49 through 50
   21.3667   21.5016

>> lci_for(:,1)'
ans =
  Columns 1 through 6
   13.2910   12.7645   12.3919   12.0990   11.8569   11.6511
  .....
  Columns 49 through 50
   10.1879   10.2094

>> uci_for(:,1)'
ans =
  Columns 1 through 6
   16.4849   17.2814   17.9239   18.4868   18.9988   19.4746
  .....
  Columns 49 through 50
   32.5454   32.7939
```

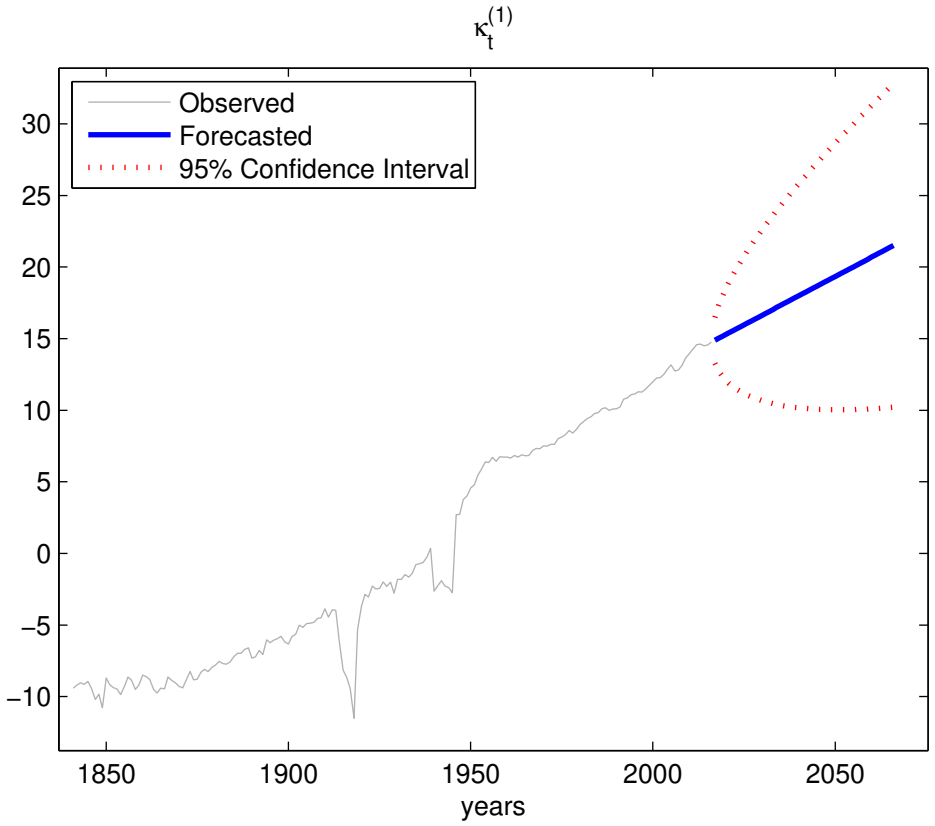

FIGURE 5.7: Forecast of the period index $\kappa_t^{(1)}$ with random walk with drift

5.2.8 Mortality Simulation

Description *HSTool* provides the `hssimulate()` function in order to generate future trajectories of the identified age-period indexes. In order to simulate the indexes the `hssimulate()` function wraps the native `estimate()`[1] and `simulate()`[2] function of Matlab. The function takes as input, among others, the period index, as well as the number of simulated trajectories and years ahead. The general structure of the `hssimulate()` function is given below:

```
[kappat_sim, uci_sim, lci_sim, mn_sim, simvrbls] = hssimulate(h, N,
kappat, datavrbls, alpha, 'visualise');
```

Inputs The input parameters of the `hssimulate()` function is given bellow.

Parameter	Description
`h`	Years ahead to be forecasted.
`N`	Number of trajectories to be generated.
`kappat`	Matrix with the estimated values of each period index $\kappa_t^{(i)}$, where $i = 1, 2, \ldots, p$.
`datavrbls`	Structure which encapsulates variables generated during the data loading process.
`alpha`	Significance level 0.05 (default) — scalar value in the range 0 to 1.
`visualise`	Option to instruct whether the function will generate a plot ("on") or not ("off").

Outputs The output parameters of the `hssimulate()` function is given bellow.

Parameter	Description
`kappat_sim`	Matrix with the simulated values for period index $\kappa_t^{(i)}$.
`uci_sim`	Vector of the upper bounds of the confidence interval.
`lci_sim`	Vector of the lower bounds of the confidence interval.
`mn_sim`	Vector of the mean values for period index $\kappa_t^{(i)}$.
`simvrbls`	Structure which encapsulates variables generated during the simulation process.
`simvrbls.arima_res`	Infers residuals of a univariate ARIMA model fit to data.
`simvrbls.arima_model`	The best ARIMA model (p,d,q).

[1] Mathworks - Fit autoregressive integrated moving average (ARIMA) model to data. Online: https://www.mathworks.com/help/econ/arima.estimate.html

[2] Mathworks - Simulate Stationary Processes. Online: https://www.mathworks.com/help/econ/simulate-stationary-arma-processes.html

Example The simulation example shown below is based on the E&W data used before in the `hsfit()`. We demonstrate the generation of 25 trajectories for 50 years ahead for $\kappa_t^{(1)}$.

```
>> [kappat_sim, uci_sim, lci_sim, mn_sim, simvrbls] = hssimulate(50, 25,
kappat(:,1), datavrbls, 0.05, 'on');


simvrbls =
    arima_res: [176x1 double]
    arima_model: '(0,1,0)'

>> kappat_sim
kappat_sim =
  Columns 1 through 6
   14.3229   15.6368   15.8368   13.6646   15.2944   15.5287
   14.1278   16.5217   15.2274   14.7226   15.5444   15.5909
   14.0460   15.7399   15.1495   14.7562   16.1140   15.4038
   15.3616   16.6649   16.9554   15.5282   16.6281   16.7641
   15.6645   16.7013   17.8361   15.7665   15.6550   16.7597

>> uci_sim'
ans =
  Columns 1 through 6
   17.2536   18.8460   18.7000   19.0065   19.0034   19.5953
  .....
  Columns 49 through 50
   38.3487   39.8068

>> mn_sim'
ans =
  Columns 1 through 6
   15.0843   15.3555   15.6745   16.0838   16.2527   16.4384
  .....
  Columns 49 through 50
   23.0347   23.2068

>> lci_sim'
ans =
  Columns 1 through 6
   13.2860   13.6880   14.0624   13.1090   13.5165   13.1158
  .....
  Columns 49 through 50
   10.8271   11.7517
```

5.2.9 Bootstrapping

Description *HSTool* provides the `hsbootstrap()` function in order to analyse the parameter uncertainty. As advocated also in Villegas et al. (2018), parameter uncertainty can be attributed to diverse sources of risk, including the uncertainty due to forecasting

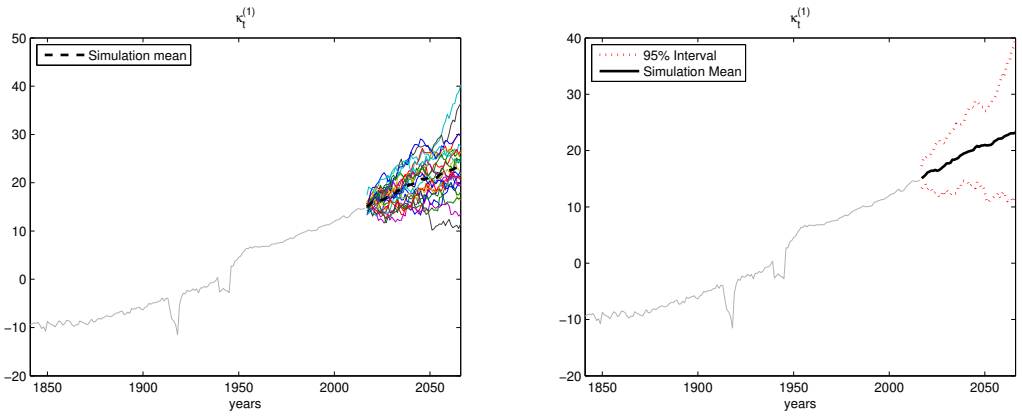

FIGURE 5.8: Simulated period index $\kappa_t^{(1)}$ (left), 95% percentile confidence interval (Right)

errors and the uncertainty from the estimation process of the model's parameters. That is, the bootstrap procedure is used in *HSTool* for measuring the parameter uncertainty. The general structure of the `hsbootstrap()` function is given below:

```
[ax_boot, betax_boot, kappat_boot, uci_boot, lci_boot] = hsbootstrap(samples,
mtxhat, etx, fitvrbls, datavrbls, alpha, 'visualise');
```

Inputs The input parameters of the `hsboostrap()` function is given bellow.

Parameter	Description
`samples`	Number of bootstrap samples.
`mtxhat`	Matrix of estimated mortality rates.
`etx`	Matrix of observed exposures.
`fitvrbls`	Structure which encapsulates variables generated during the HS fitting process.
`datavrbls`	Structure which encapsulates variables and metadata generated during the data loading process.
`alpha`	Significance level 0.05 (default) — scalar value in the range 0 to 1.
`visualise`	Option to instruct whether the function will generate a plot ("on") or not ("off").

Outputs The output parameters of the `hsboostrap()` function is given bellow.

Parameter	Description
`ax_boot`	Matrix (a × N) of the N-bootstraped values of age effect of the age profile of mortality α_x.
`betax_boot`	Matrix (3-dimensional $k1 \times$ a × N) with the N-bootstraped values of each age-related effect $\beta_x^{(i)}$ for each period component.
`kappat_boot`	Matrix (3-dimensional $n \times k1 \times$ N) with the N-bootstraped values values for period index $\kappa_t^{(i)}$.
`uci_boot`	Matrix of the upper bounds of the confidence interval.
`lci_boot`	Matrix of the lower bounds of the confidence interval.

Documentation *HSTool* adopts the semi-parametric bootstrap method proposed by Brouhns et al. (2005) to take into account the sampling errors of the GLM process in the estimated SPCs, generating also the associated confidence intervals.

N bootstrap samples $d_{t,x}^{(i)}, i = 1, 2, \ldots, \mathrm{N}$ are generated as realisations from the Poisson distribution with mean $\hat{d}_{t,x}$. Each sample is used to re-estimate the model so that to get N parameter estimations of $B^{(i)}$ which contains the GLM–estimated parameters as produced using the GLM approach. Then, we get N parameter estimations $\alpha_x^{(i)}, \beta_x^{(1),(i)}, \ldots, \beta_x^{(p),(i)}, \kappa_t^{(1),(i)}, \ldots, \kappa_t^{(p),(i)}$, $i = 1, \ldots, \mathrm{N}$, by executing the HS estimation process N times. As suggested in Renshaw & Haberman (2008), we make use of the fitted number of $\hat{d}_{t,x}$ instead of the observed $d_{t,x}$. Once the HS is bootstrapped we can simulate it forward to obtain simulated trajectories which account for both the forecast error in the period indexes and the error in the model fitting. Having the N model realisations, the 95% percentile confidence and prediction intervals $CI = [p_{0.025}, p_{0.975}]$ can be generated. According to Hatzopoulos & Haberman (2009) the bootstrap confidence interval avoids the normal assumption and is more reliable than the standard normal interval.

Example The bootstrapping example shown below is based on the E&W data used before in the `hsfit()`. We demonstrate the generation of 5000 samples.

```
>> [ax_boot, betax_boot, kappat_boot, uci_boot, lci_boot] =
hsbootstrap(5000, mtxhat, etx, fitvrbls, datavrbls, 0.05, 'on');
```

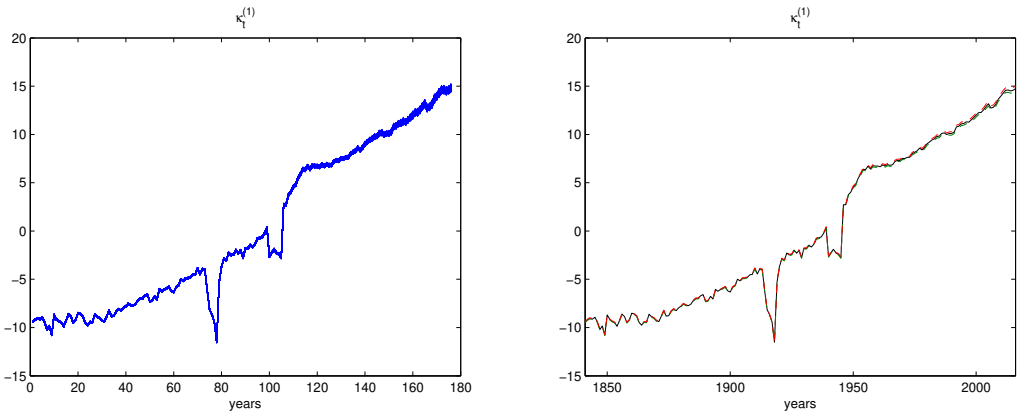

FIGURE 5.9: Bootstrapped $\kappa_t^{(1)}$ (left) and 95% percentile confidence intervals (Right)

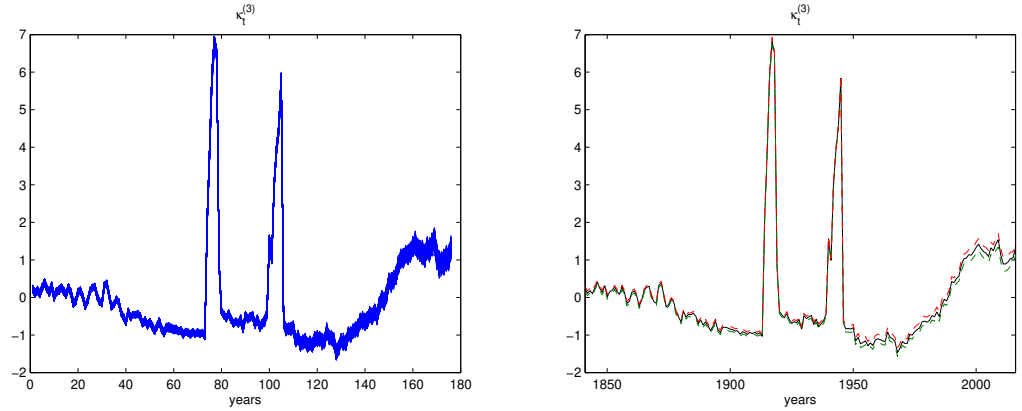

FIGURE 5.10: Bootstrapped $\kappa_t^{(3)}$ (left) and 95% percentile confidence intervals (Right)

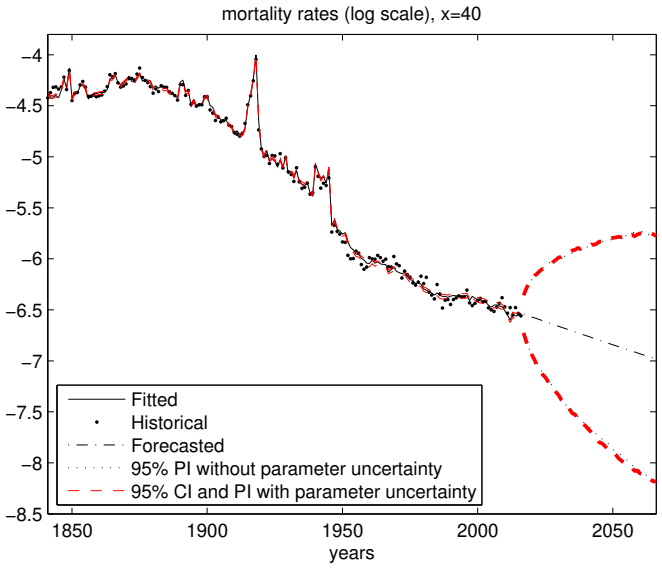

FIGURE 5.11: 95% prediction intervals for mortality rates $m_{t,x}$ at age $x = 40$ fitted to the E&W data for ages 0–89 and the period 1841–2016.

5.2.10 HSTool Dependencies

The *HSTool* has the following dependencies:

- DSPCA Toolbox: This is a Matlab toolbox for Sparse Principal Component Analysis. This toolbox needs to be part of a system's path in order to be able to perform the SPCA operations for reducing the dimensionality of the problem of identifying the p age-period components. The tool has been introduced in Luss & d'Aspremont (2006). A complete user guide on the installation and use can be found in `https://www.di.ens.fr/~aspremon/PDF/DSPCAUserGuide.pdf`
- CAPTAIN Toolbox: It provides access to novel, mainly recursive, algorithms for various important aspects of identification, estimation, non-stationary time series analysis and signal processing, adaptive forecasting and automatic control system design. *HSTool* uses CAPTAIN for the realisation of the DLR functionalities used in the context of the forecasting the estimated mortality rates. CAPTAIN was introduced in Young et al. (2009) and complete documentation can be found in `http://captaintoolbox.co.uk/Captain_Toolbox.html`

5.3 Conclusion

In this chapter we introduced the *HSTool*, a Matlab toolbox for the Hatzopoulos & Sagianou (2020) multiple-component stochastic mortality model. The chapter offers a complete technical documentation of the functionalities of the tool supported by the necessary documentation of the background methodologies and examples of the functionalities execution. In this context, we documented the *HSTool* core functionalities for realising the *Mortality Data Loading*, *Model fitting*, *Goodness-of-fit analysis*, *Multi-criteria decision making* and *Forecasting*, *Simulation* and *Bootstraping* functionalities. The aforementioned functionalities are those that enable a user to create a comprehensive mortality modelling testbed and fully evaluate the HS model under diverse datasets and conditions.

Apart from the realisation of the HS model in the form of a Matlab toolbox, this chapter has introduced a number of improvements in the operational behaviour of the model

compared to the initial approach presented in Hatzopoulos & Sagianou (2020) back in 2020.

More specifically, we introduced the use the multi-criteria decision making methods (SAW and TOPSIS) to support the decision making and model's convergence to "optimal" values for critical model's parameters, based on the use of diverse performance criteria. Thus, on the one hand, these methods used for identifying the optimal $k1$ and $k2$ degrees of the orthonormal polynomials produced by the model's graduation process based on the use of GLM, and on the other hand, for the identification of the optimal p age–period components that comprise the final model structure. The introduction of such enhancements came to augment the model's operation and alleviate the user from manually scrutinising the model's results and setting empirical thresholds to define the optimal parameters to be finally used. Hence, *HSTool* offers a high level of automation and enhanced user experience by still enabling testing under different parameterisations to explore the capabilities of the model and the peculiarities of mortality data.

In addition, another enhancement was the introduction of a dynamic searching approach of the UVR thresholds of the age–period and age–cohort components. More specifically, in Hatzopoulos & Sagianou (2020), an empirical threshold value for significant UVR values was set to $UVR(i) < 50\%$. However, the determination of the UVR value affects the optimality of the s value and, hence, different s values uncover different components in the mortality data under different stringent or loosen UVR threshold values. That is, *HSTool* delivers a new approach for the analysis of the local minima of the UVR graphs that guarantee that the model converges to the optimal components that maximise the captured information.

Overall, this chapter aspired to unfold the qualities of HS model not just on the paper, but to make substantial contribution is the mortality modelling domain by providing the complete tool to the community. In this direction, this work has provided the necessary material steering the users on the use of model's commands for replicating the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projection.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

This chapter summarises the contribution of this doctoral thesis and provides research directions for future work. In the frame of mortality modelling, this thesis created a steppingstone by analysing the behaviour of existing mortality models in the context of a comparative evaluate testbed, and used it to go beyond the prominent and well-established solutions. This was done by introducing new methods and tools that can provide accurate modelling and foster research in academia. Reaching to the end of this doctoral thesis, it becomes clear that there was room for improvement in the mortality domain, as the rigid structures of the well-established models of the literature cannot offer a fine-grained and in-depth analysis of the mortality trends.

More specifically, as described in Chapter 2, there is a wide set of well-established mortality models, but as advocated in Chapter 3, those still have limitations and pose open research challenges.

That is, Chapter 3 performed a comparative analysis of the aforementioned models in terms of model fitting and mortality rates prediction in order to benchmark their performance. Given this, Chapter 3 introduced the HS model that adopted a dynamic structure with multiple age–period and age–cohort interaction terms, estimated using

Generalised Linear Models (GLMs) and Sparse Principal Component Analysis (SPCA). The model gains a highly informative structure in an efficient way thanks to the use of the Unexplained Variance Ratio (UVR) metric, and it is able to designate an identified mortality trend to a unique age cluster, tackling a major limitation on the interpretability of modelling results in the domain. Notably, the HS model is able to achieve high scores over diverse qualitative and quantitative evaluation metrics and outperforms the rest of the models in the majority of the experiments, while the model comes into agreement with well-established findings in the mortality literature.

As a next step, Chapter 4 introduced a set of HS extensions, as we identified that there was room for improvement to the initial HS model. That is, we formulated the HS model in terms of $q_{t,x}$ and we made use of various link functions and diverse distributions in the model's estimation method. We gave particular focus on the use of heavy-tailed distributions, and we evaluated their applicability in the context of mortality through the HS extensions.

Chapter 5 aimed to have a practical impact for this thesis through offering the *HSTool*, a Matlab package for the HS model, along with the necessary material for the replication of testbeds and the use of the model's commands. In addition, this chapter extended the initial HS codebase by introducing multi-criteria decision making methods in the model's workflow to achieve increased tool automation and convergence to "optimal" values for critical model's parameters. That is, this chapter completed a series of contributions of this doctoral thesis by enabling the replication of the full cycle of a mortality model, i.e., fitting the stochastic mortality model, assessing its goodness-of-fit and performing mortality projections.

Table 6.1 provides a mapping between the objectives (see Section 1.1), the key contributions of this thesis and the corresponding published works to peer-reviewed scientific journals. In other words, this mapping describes how our research endeavours fulfilled the initial goals set for contributing in the mortality modelling domain. Hence, this table summarises our path towards the completion of this PhD thesis.

Obj.	Chapt.	Contribution	Publication
2	3	Introduction and evaluation of a new multiple component stochastic mortality model, namely the Hatzopoulos-Sagianou (HS) model	Hatzopoulos & Sagianou (2020)
2, 3	3	Introduction of Unexplained Variance Ratio (UVR) metric in order to pinpoint the most important age–period and age–cohort components and the optimal model structure.	Hatzopoulos & Sagianou (2020)
1	3	Extensive evaluation of the new model and comparative analysis of the obtained results against well-known existing mortality models.	Hatzopoulos & Sagianou (2020)
3	4	Extensions on the Hatzopoulos–Sagianou multiple-component stochastic mortality model based on the use of different link functions and probability distributions.	Sagianou & Hatzopoulos (2022)
2, 3	4	Introduction of a new set of link functions, with particular focus on heavy-tailed distributions, and applicability evaluation in the context of mortality through the HS extensions.	Sagianou & Hatzopoulos (2022)
3	4	Lessons learnt to inform the community on the adoption of the various link functions in a models' estimation methods.	Sagianou & Hatzopoulos (2022)
1, 2	3, 4	Design of a backtesting testbed (out-of-sample comparison) in order to evaluate the ability of each model in terms of prediction	Hatzopoulos & Sagianou (2020), Sagianou & Hatzopoulos (2022)
4	5	A software implementation and technical documentation of the HS model in the form of a MATLAB package under the name *HSTool*	Hatzopoulos & Sagianou (n.d.)
3, 4	5	Introduction of multi-criteria decision making methods in the model's workflow to achieve increased tool automation and convergence to "optimal" values for critical model's parameters	Hatzopoulos & Sagianou (n.d.)

TABLE 6.1: Overall PhD Thesis Contribution.

6.2 Summarising the advantageous characteristics of HS

As was mentioned in the introduction, several mortality models have been introduced in the literature, and each time a new model aims to address open challenges and gaps. The HS has the same purpose. Having highlighted the limitations of existing models based on the comparative study presented in Chapter 3, we summarise here the advantageous characteristics of the HS.

- Introduces a dynamic structure with multiple age-period and age-cohort components - Legacy models have a static structure and consider only 1 to 3 components.

To name a few, Lee & Carter (1992) model offers 1 age-period component. Renshaw & Haberman (2006) offer 1 age-period and 1 age-cohort, and Plat (2009) offers 3 period and 1 cohort components, with pre-specified age effect terms.

- Its estimation methodology reveals significant and distinct age clusters by identifying the optimal number of incorporated period and cohort effects, and enables the interpretability of the identified mortality trends. - Legacy models have a static structure and they do not provide distinct age clusters to enable the attribution of an identified trend to a specific age range. In this sense, their components are not clear, and the extracted mortality trend is attributed to the whole age range or a mix of age ranges.

- Its estimation methodology is driven by the Sparse Principal Component Analysis and the Unexplained Variance Ratio (UVR) to identify the most informative components in order to maximise the captured variance of the mortality data. - Legacy models aim solely to maximise the log-likelihood of the model.

- Due to its extensible structure (i.e., its components are not predefined), different age-period and age-cohort components can be identified on different mortality datasets, as different datasets (and countries) can have different mortality behaviour. - Legacy models, due to their static structure, always offer the same number of components regardless of the dataset provided.

- Can operate over the whole age range and, thus, it can identify all the significant age clusters having high information variance (using the UVR metric). Thus, it can offer a holistic view for mortality behaviour over a dataset. On the other hand, there are legacy models that, either cannot operate by design over the whole age range (e.g., CBD models), or they do not include the necessary components in order to uncover the mortality trends for distinct and fine-grained age clusters.

- Can be used for in-depth analysis of mortality on specific age clusters. Legacy models offer a generic representation of mortality dynamics without specifying which ages are reflected in a mortality trend.

- Based on the extended methodology presented in Sagianou & Hatzopoulos (2022), the HS enables the adoption of virtually any distribution as link functions in the estimation process, in order to improve the *goodness-of-fit* and better adapt to the

intrinsic information of the mortality data analysed. Legacy models make use only of a standardised set of link functions (usually log and logit).

- The HS model has been proven to outperform well-known models of the literature over diverse qualitative and quantitative evaluation metrics (MAPE, MSPE, AIC, BIC, Explained variance), as can be seen in Hatzopoulos & Sagianou (2020) and in Sagianou & Hatzopoulos (2022).

All the aforementioned qualities are advocated by the quantitative and qualitative experimental results in Hatzopoulos & Sagianou (2020) and Sagianou & Hatzopoulos (2022). In addition, in Hatzopoulos & Sagianou (n.d.) we aim to make the HS available to the community and offer a tool to realise the above-mentioned qualities and introduce improvements in the model's algorithm for increased automation and operational stability.

6.3 Meeting the mortality modelling design criteria

As described in Section 2.5, a list of criteria has been defined in the mortality literature as a set of qualities that need to be met by a mortality model or to be considered as design requirements that researchers need to follow in their endeavour of introducing a new model. Table 6.2 summarises this analysis on the satisfaction of the criteria by the models used in the comparative study of this Phd thesis, while the last column refers to the HS model. In fact, part of this table was initially presented in Plat (2009) and, in the context of this work, we extend it by adding the results regarding the HS model, and notably, by including more criteria in the analysis.

More specifically, criteria 16 to 19 have been included as desirable characteristics that coincide with the advantages of the HS model structure as summarised in the previous section. As can be seen, none of the models included in the comparative study meets the criteria that the HS introduces.

Criteria	Renshaw & Haberman (2006)	Currie (2006)	Lee & Carter (1992)	Plat (2009)	**Hatzopoulos & Sagianou (2020)**
1. Positive mortality rates	+	+	+	+	+
2. Consistency with historical data	+	+	+/-	+	+
3. Long-term biological reasonableness	+	+	+	+	+
4. Estimation robustness	+/-	+	+	+	+
5. Forecasts biological reasonable	+	+	+/-	+	+
6. Straightforward to implement	+	+	+	+	+
7. Parsimony	+/-	+/-	+	+/-	+
8. Possibility generating sample paths	+	+	+	+	+
9. Allowance parameter uncertainty	+	+	+	+	+
10. Incorporation of cohort effects	+	+	-	-	+
11. Non-trivial correlation structure	+/-	+/-	-	+	+
12. Applicable for a full age range	+/-	+/-	+/-	-	+
13. Transparent - not a "black box"	+	+	+	+	+
14. Good fit to the historical data	+/-	-	-	+	+
15. Rank well against other models	+	-	-	+	+
16. Dynamic structure	-	-	-	-	+
17. Adaptability to population idiosyncrasies	-	-	-	-	+
18. Attribution of trend to clear age clusters	-	-	-	-	+
19. Maximisation of captured information	-	-	-	-	+

TABLE 6.2: Comparison of mortality models - Satisfaction of criteria

In the following points we explain how the HS model meets each of the criteria:

1. *Positive mortality rates:* Positive mortality values are derived from our model structure since all the link functions in our modelling are being used with Poisson or Binomial GLM structure.

2. *Consistency with historical data:* Section 3.4 provided a thorough discussion on the model results and documented how the HS comes into agreement with well-established findings in the mortality literature. This ability of the HS is attributed to the dynamic linear regression modelling and to the polynomial expansion approach for modelling age patterns of mortality. The latter offers great flexibility in the graduation process since most continuous functions can be approximated by a polynomial to any degree of accuracy in the form of a Taylor power series.

3. *Long-term biological reasonableness:* The HS is aligned with historical data, as well as with mortality dynamics that have been widely recognised in the mortality literature. Its ability to identify clear mortality trends for distinct age clusters aid to the identification of certain dynamics, such as the "accident hump", the "golden

cohort effect", the "smoking effect" and the "ageing of mortality improvement" (See Section 3.4). This ability comes as a result of the use of SPCA and the UVR metric.

4. *Estimation robustness:* The diverse evaluation testbeds built in the context of Chapters 3 and 4 verify this criterion. The HS was stress-tested using various datasets both for short and long modelling periods.

5. *Forecasts biological reasonable:* The forecasting results are biologically reasonable. The HS was tested under diverse out-of-sample setups achieving better results than the rest of the well-established mortality models.

6. *Straightforward to implement:* The methodology of the HS has been fully documented in Chapter 3, while pseudocodes that describe the model's operational flow are presented in Algorithms 1 and 2. The HS has been implemented in Matlab offered as a toolbox, namely *HSTool*. In addition, Chapter 5 offered a complete handbook of the *HSTool*.

7. *Parsimony:* The incorporation of age period and cohort effects occurs in a dynamic manner. Thus, depending on the mortality dataset being modelled, a different number of parameters is estimated. One could argue that the incorporation of multiple parameters (e.g., for E&W 5 age-period and 2 age-cohort) is opposed to the parsimony criterion. However, as advocated by our results, all the incorporated age and cohort terms are significant and they contribute to the explainability of the mortality trends. In addition, even under the BIC metric, the HS achieved the best performance among all the models, even if its structure incorporated the highest number of parameters. This implies that the added components are indeed necessary to be incorporated in the model structure.

8. *Possibility generating sample paths:* HS is able to generate sample paths utilising bootstrapping and generate percentile confidence intervals, as can be seen in Sections 5.2.8 and 5.2.9.

9. *Allowance parameter uncertainty:* The generation of sample paths is an important criterion to witness parameter uncertainty. By using semi-parametric bootstrap,

parameter uncertainty can be incorporated in the HS model, as proposed in Section 5.2.9. HS takes into account the sampling errors of the GLM process in the estimated components, generating also the associated confidence intervals.

10. *Incorporation of cohort effects:* The age-period-cohort structure of HS model incorporates q stochastic cohort effects. The number of the components varies depending on the mortality dynamics of the population being modelled.

11. *Non-trivial correlation structure:* The HS model structure incorporates p age-period and q age-cohort components, implying a non-trivial correlation structure.

12. *Applicable for a full age range:* The model is applicable to the whole age range, as can be seen in the evaluation testbeds of Chapters 3 and 4.

13. *Transparent - not a "black box":* The methodology of the HS has been clearly documented in Chapter 3, while its operational flow is presented in Algorithms 1 and 2. In that sense, the model is transparent and users can understand its internal workings.

14. *Good fit to the historical data:* The HS model achieved exceptional performance under several quantitative and qualitative performance criteria, as reported in Chapters 3 and 4.

15. *Rank well against other models:* Chapters 3 offered a comparative analysis on the performance of the HS against other well-known models in the literature. The HS outperforms the rest of the models in the majority of the experiments.

16. *Dynamic structure:* As advocated by its structure, the HS model incorporates a no-predefined number of p age-period and q age-cohort components. The p and q, and the components per se, are estimated dynamically during the execution of the fitting process of the model. Thus, HS ensures that the final model structure will incorporate the necessary number of components for offering a fine-grained decomposition of the mortality rates. Its dynamicity contrasts with the rigid model structures of other models.

17. *Adaptability to population idiosyncrasies:* The varying number of p and q components make the HS model to adapt to the mortality dynamics of each population

(or dataset) being analysed. Considering that HS has the ability to perform accurately, it becomes clear that, adaptability is a key criterion for performing an in-depth analysis of mortality. HS does not consider pre-defined factors that instruct a definite behaviour to the model, but its estimation method helps the model to adapt. When data change, we cannot expect that rigid models can always perform adequately and adapt to the data mortality dynamics.

18. *Attribution of trend to clear age clusters:* The principal components identified by the SPCA method are scrutinised by the estimation process using the UVR metric. The latter is able to quantify the data variance of a component and identify a unique age cluster to which the identified mortality trend is attributed to. This characteristic contributes to the interpretability of the modelling results. Legacy models do not provide distinct age clusters and cannot enable the attribution of an identified trend to a specific age range.

19. *Maximisation of captured information:* The estimation process of HS is driven towards acquiring as much as possible of the variance of the mortality data. This is achieved through the SPCA, which provides a better clustering of significant age-period and age-cohort factors, supported by the UVR metric which instructs the model to incorporate in the structure the most informative p and q components (See Table A.7).

6.4 Future Research Directions

This PhD thesis has mainly contributed to the field of mortality modelling by introducing a new model and a set of supportive methodologies for addressing key limitations of this research field. Undoubtedly, the quest for novel methods and tools should be continuous. It becomes clear that mortality modelling is a steppingstone for other business and research directions where the advancements in mortality modelling can have a positive impact. To this end, this subsection elaborates on possible next steps and future research directions.

- *HSTool extensions* - A key outcome of this PhD thesis is an open source version of the *HSTool* as a Matlab toolbox. This version of the tool realises the model version

presented in Hatzopoulos & Sagianou (2020) including the enhancements related to the optimisation based on the use of multi-criteria decision making methods. After establishing the *HSTool* and acquiring the feedback of the community, several enhancements can follow based on the HS extensions published in Sagianou & Hatzopoulos (2022) in order to make available the HS model formulated in terms of $q_{t,x}$ and to make use of various link functions and diverse distributions in the model's estimation method.

- *Applications of the HS for optimal allocation of Solvency Capital Requirement* - In the context of Solvency II, the longevity risk is a concern that drives the allocation of Solvency Capital Requirement. Mortality models have been a link in the chain of operations for the determination of SCR. To this end, a future direction will be the evaluation of the HS model, and a comparative study against other models, on the effect that the more accurate fitting and forecasting performance of the HS model may have on the allocation of SCR.

- *Use of Artificial Neural Networks for accurate predictions of mortality rates* - In the context of this PhD thesis, the ARIMA and DLR where used as the main methods to extrapolate mortality rates and perform out-of-sample forecasts. Neural Networks and Deep Neural Networks attract significant attention nowadays and provide new opportunities for modelling mortality (Deprez et al. (2017), Levantesi & Pizzorusso (2019)). In this direction, such methods can be used to directly process mortality rates to produce forecasts with increased accuracy. This approach is worthy of investigation when considering its integration in the modelling life-cycle. Such estimations of future mortality rates can contribute to pricing of insurance products and in the longevity risk management.

Appendix A

Evaluation results of the experimental testbeds of Chapter 3

A.1 Parameters estimations and Unexplained Variance Ratio

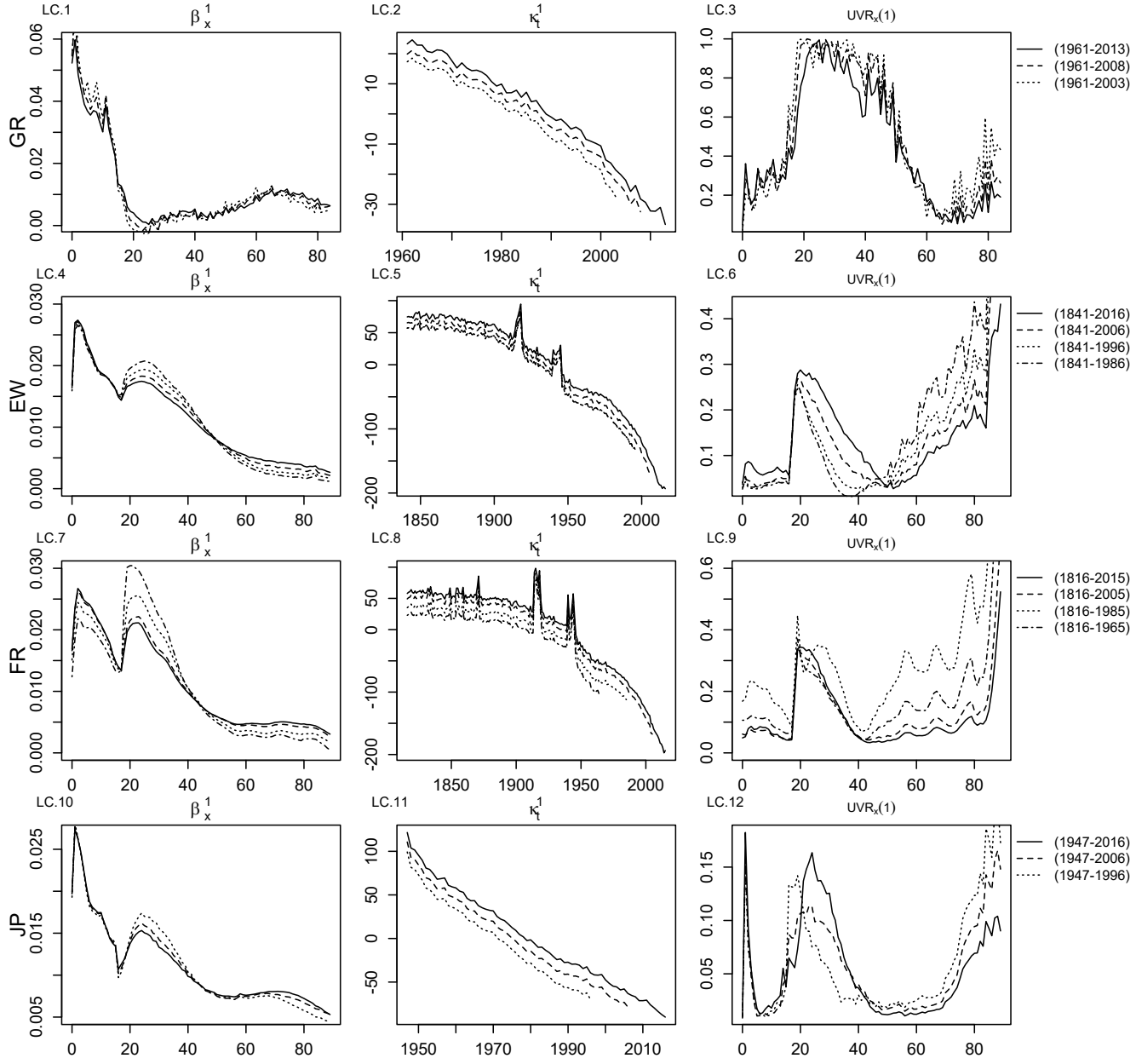

FIGURE A.1: LC model: first age–period component for all countries (GR, EW, FR, JP).

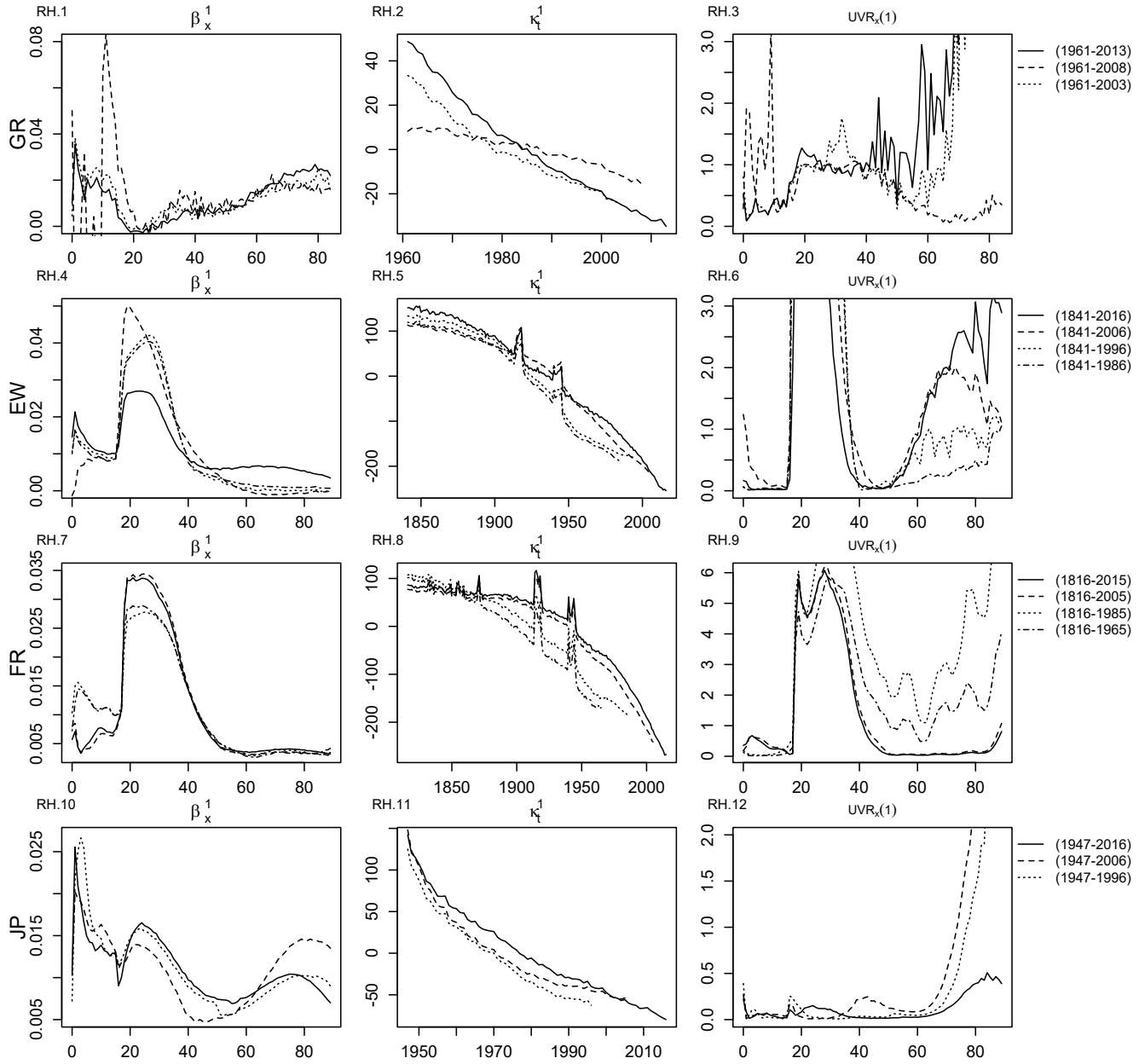

FIGURE A.2: RH model: first age–period component for all countries (GR, EW, FR, JP).

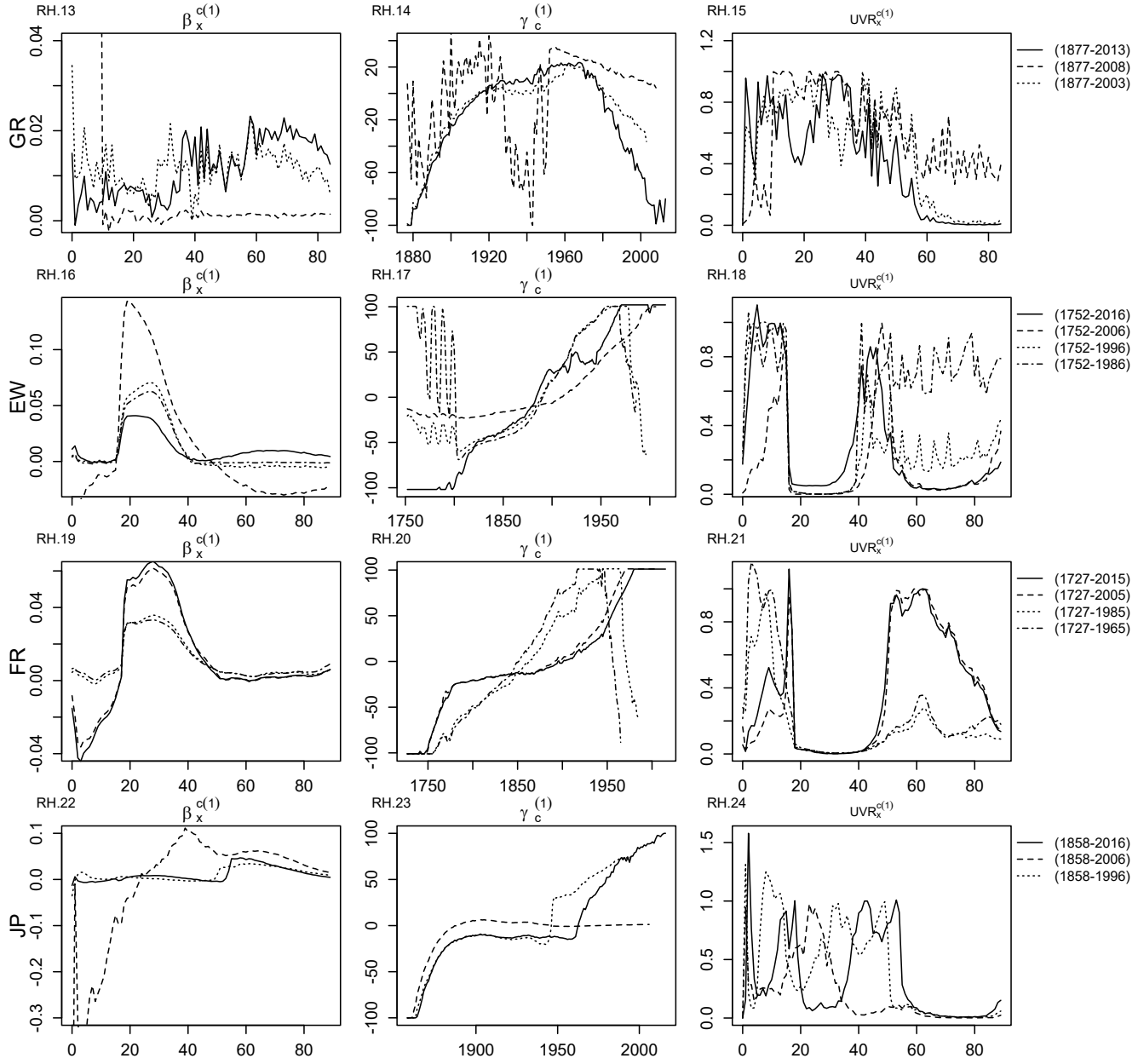

FIGURE A.3: RH model: age–cohort component for all countries (GR, EW, FR, JP).

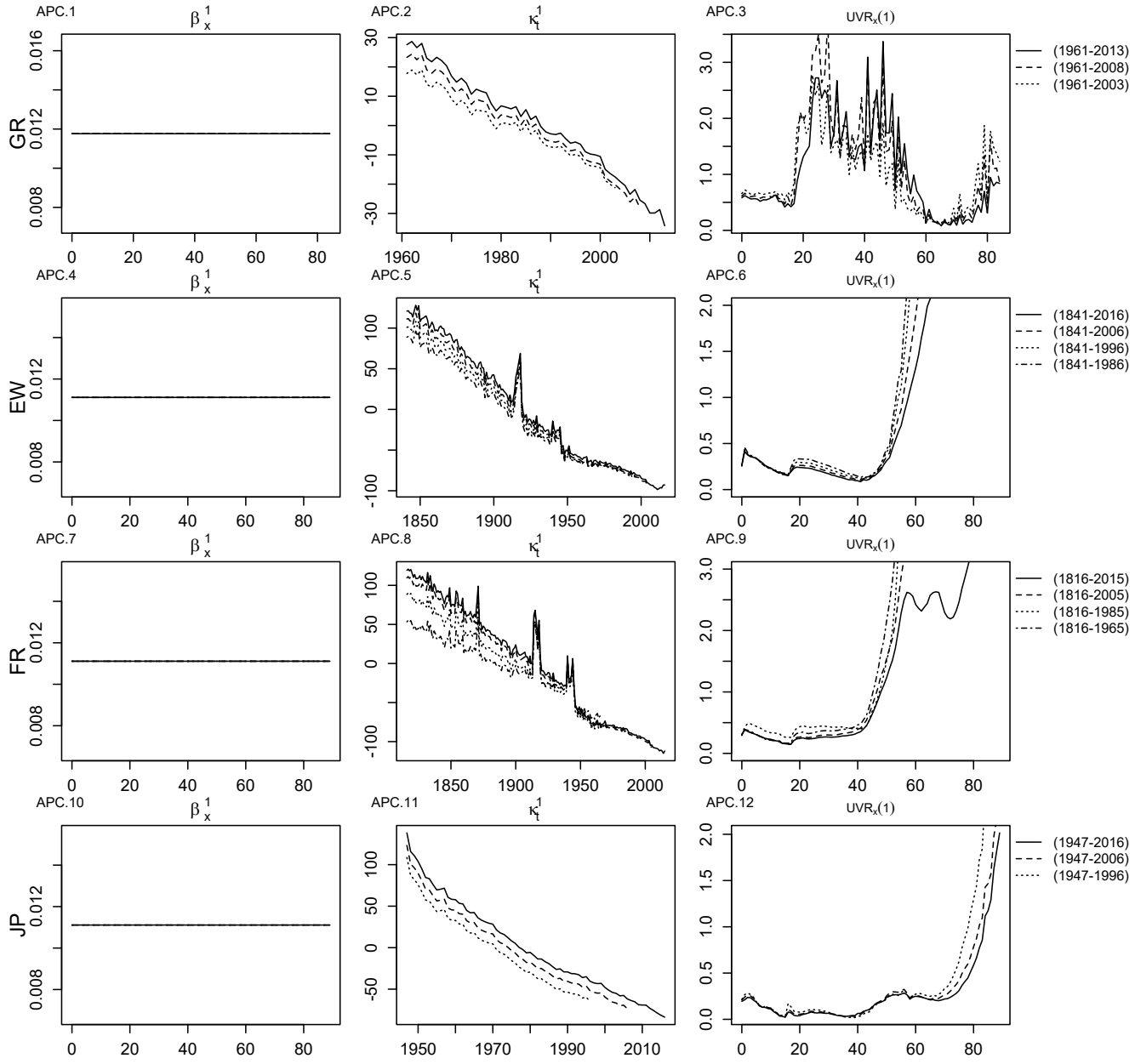

FIGURE A.4: APC model: first age–period component for all countries (GR, EW, FR, JP).

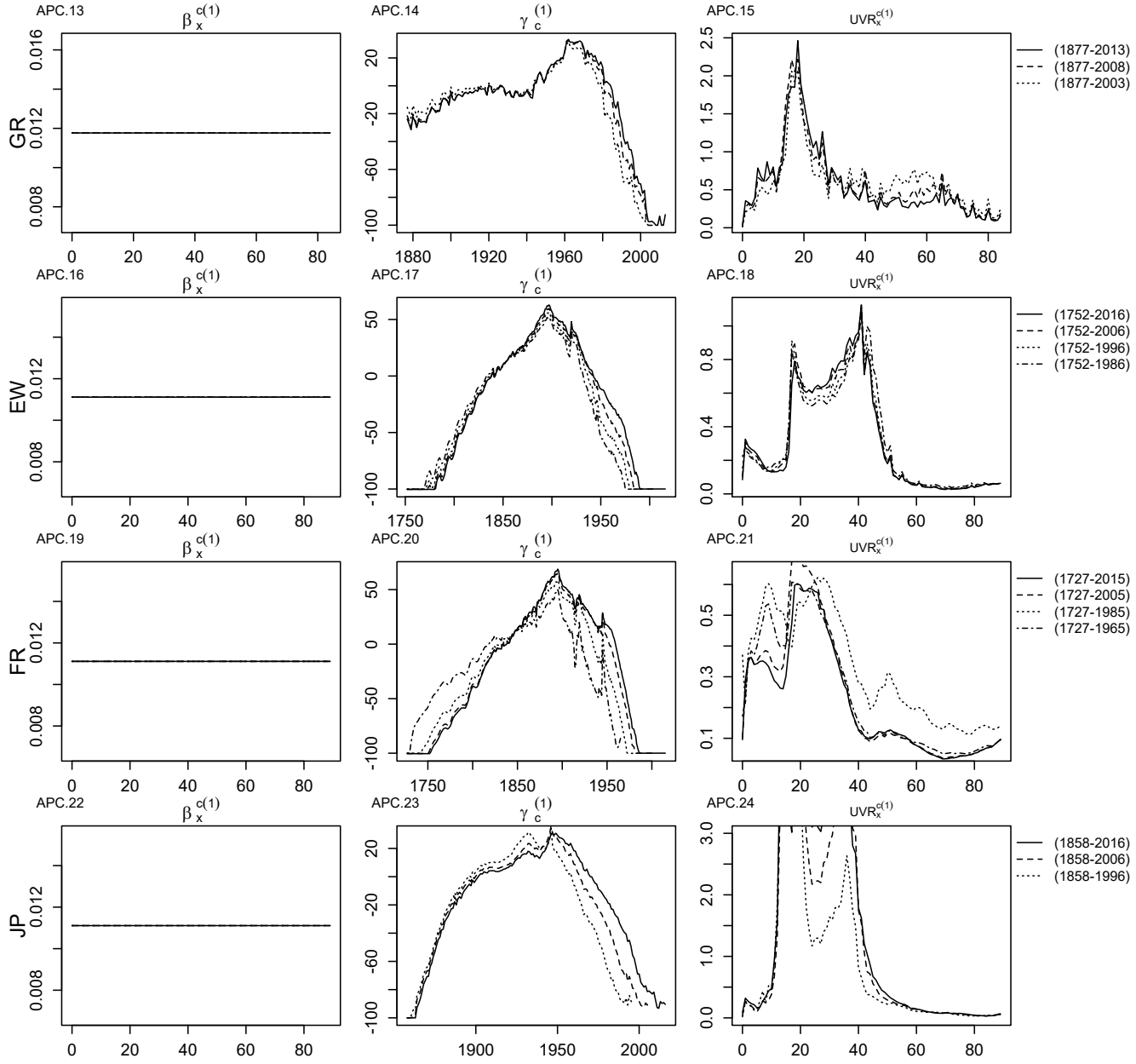

FIGURE A.5: APC model: age–cohort component for all countries (GR, EW, FR, JP).

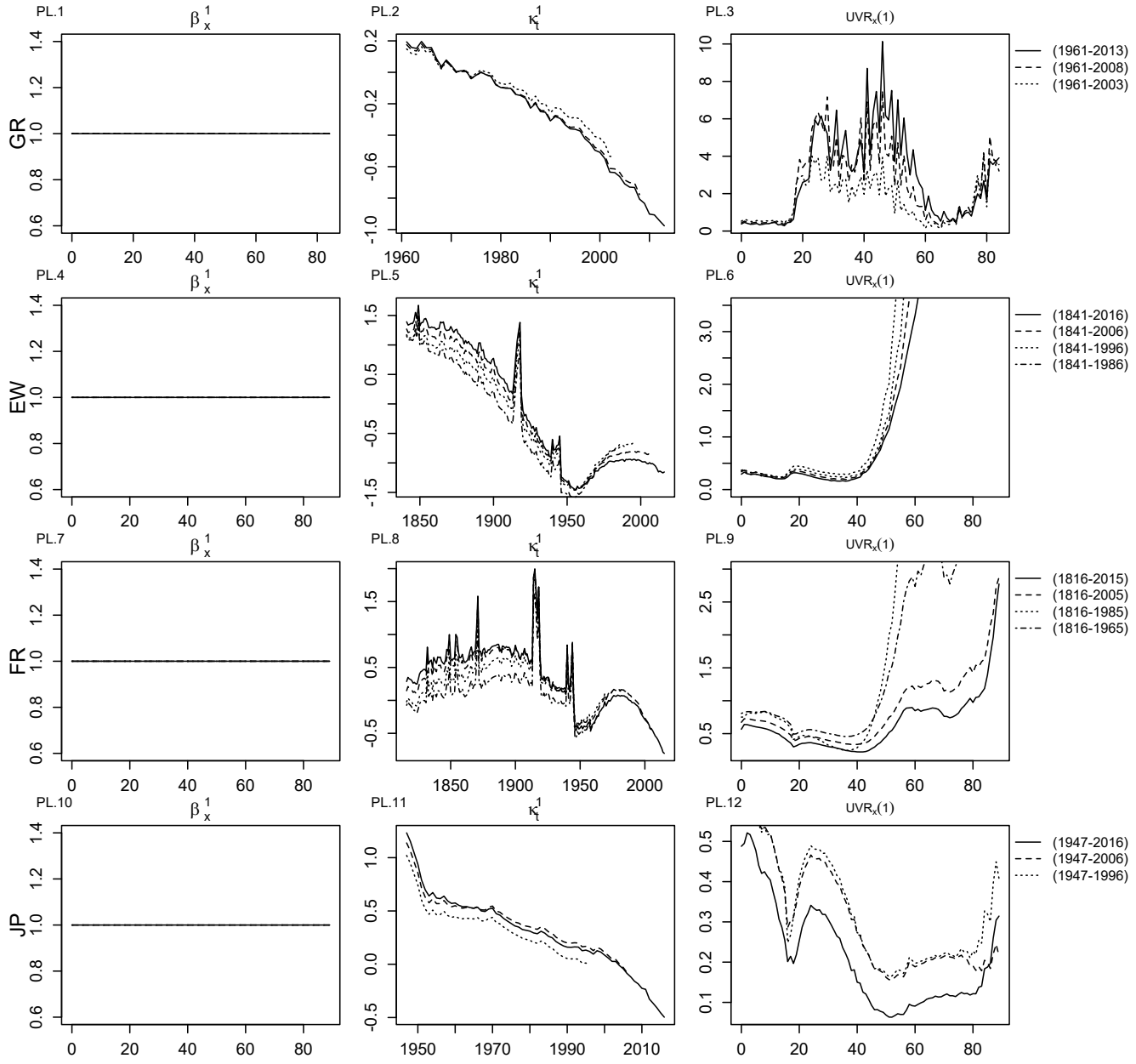

FIGURE A.6: PL model: first age–period component for all countries (GR, EW, FR, JP).

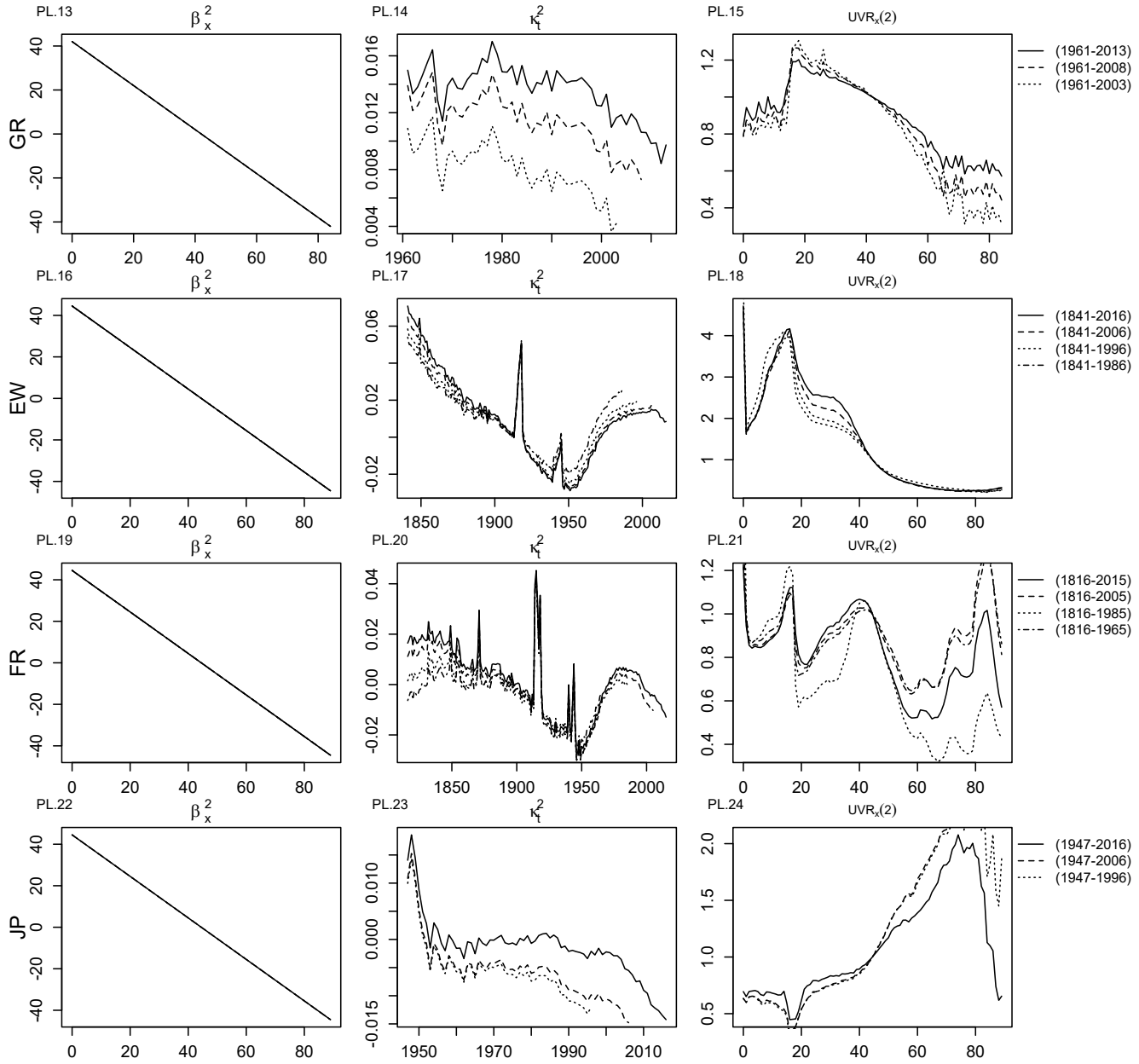

FIGURE A.7: PL model: second age–period component for all countries (GR, EW, FR, JP).

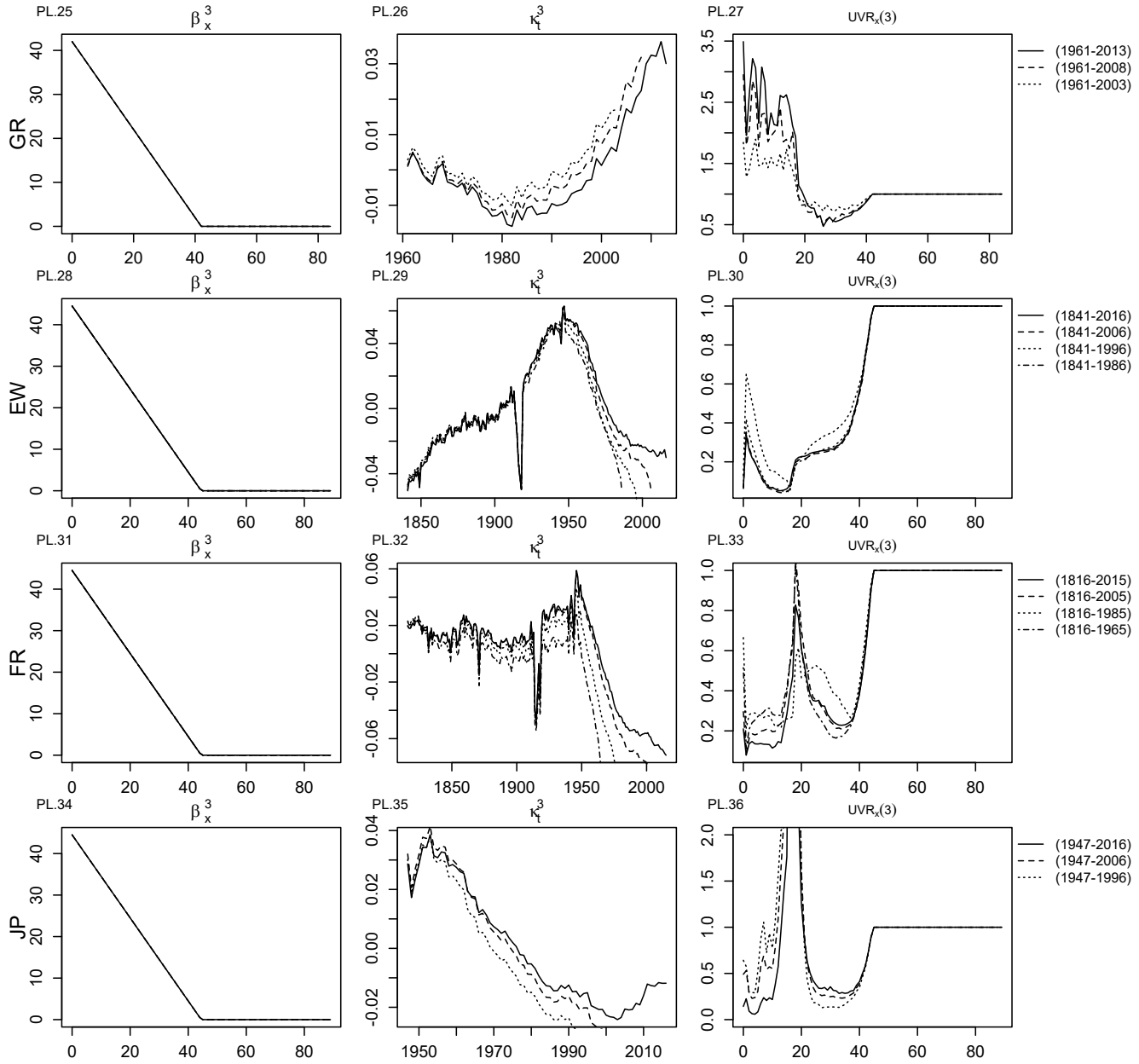

FIGURE A.8: PL model: third age–period component for all countries (GR, EW, FR, JP).

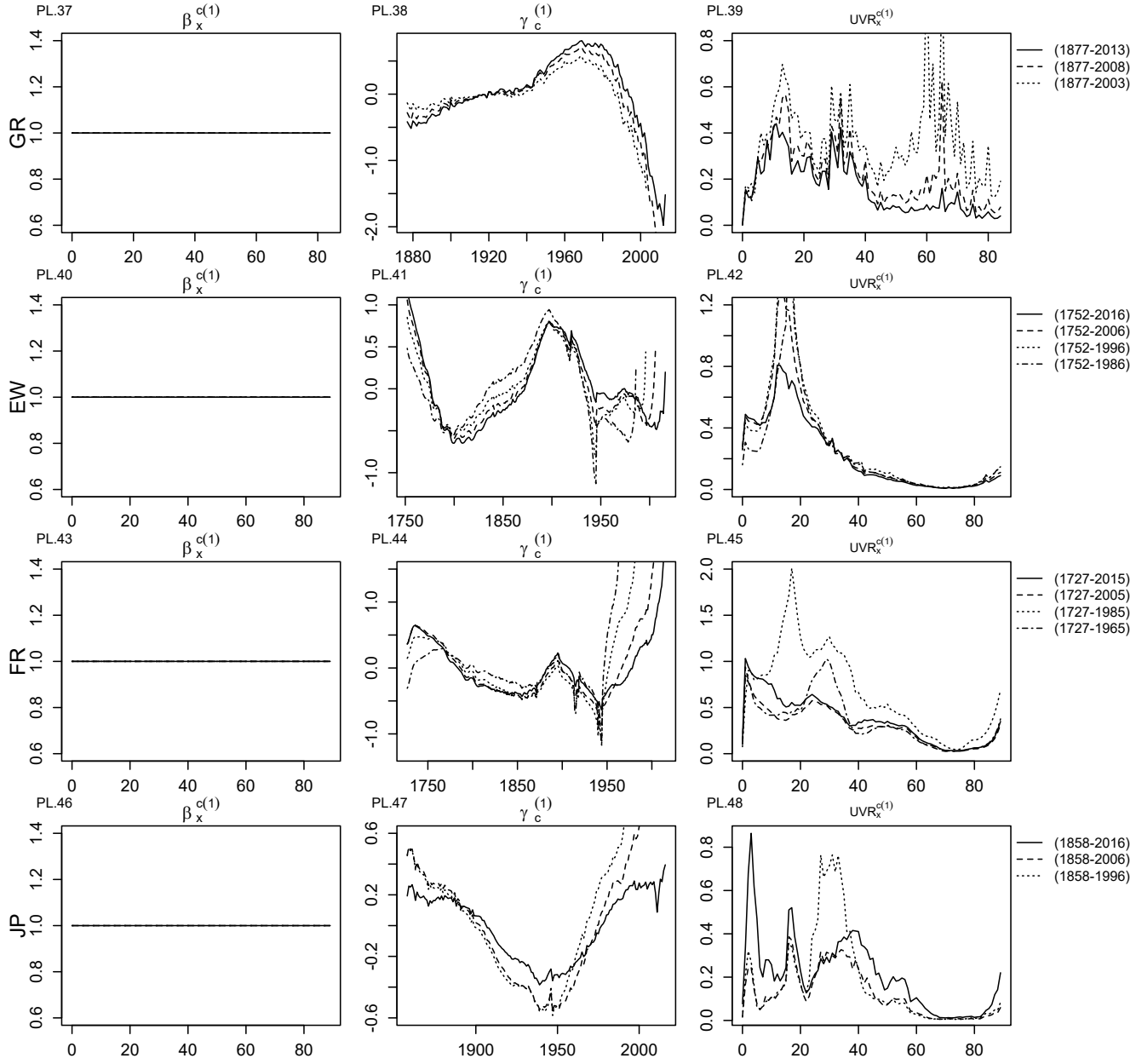

FIGURE A.9: PL model: age–cohort component for all countries (GR, EW, FR, JP).

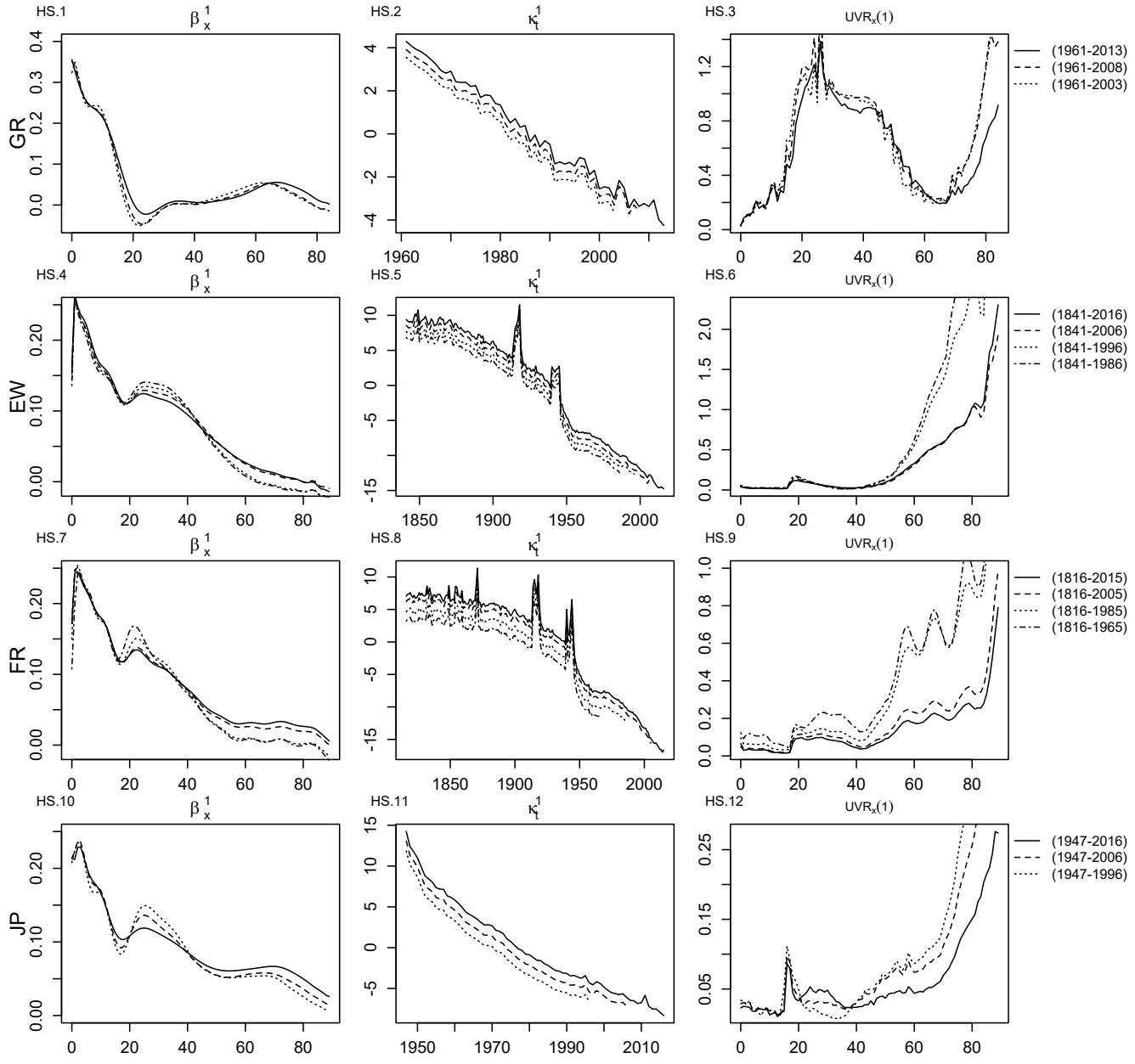

FIGURE A.10: HS model: first age–period component for all countries (GR, EW, FR, JP).

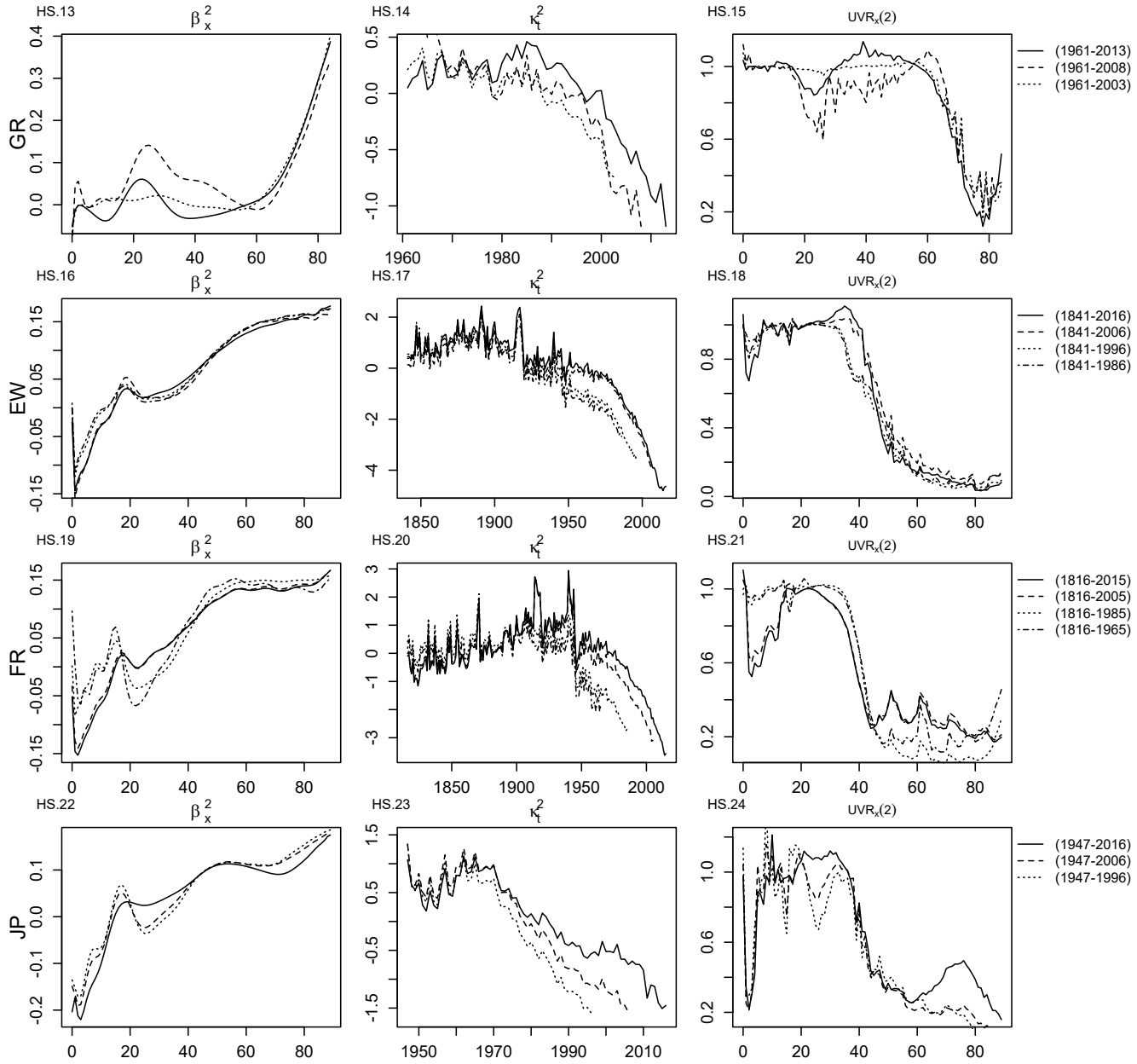

FIGURE A.11: HS model: second age–period component for all countries (GR, EW, FR, JP).

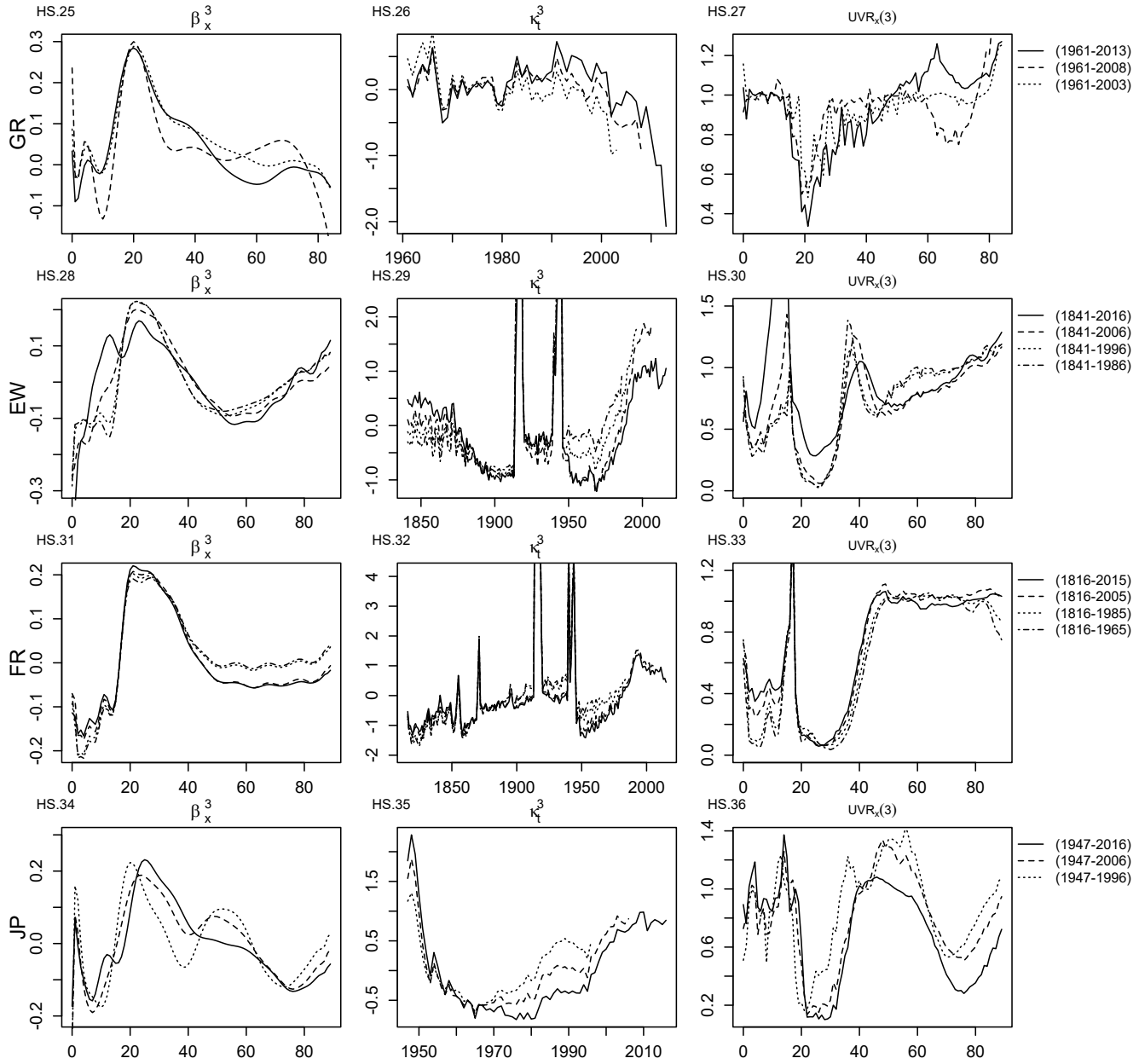

FIGURE A.12: HS model: third age–period component for all countries (GR, EW, FR, JP).

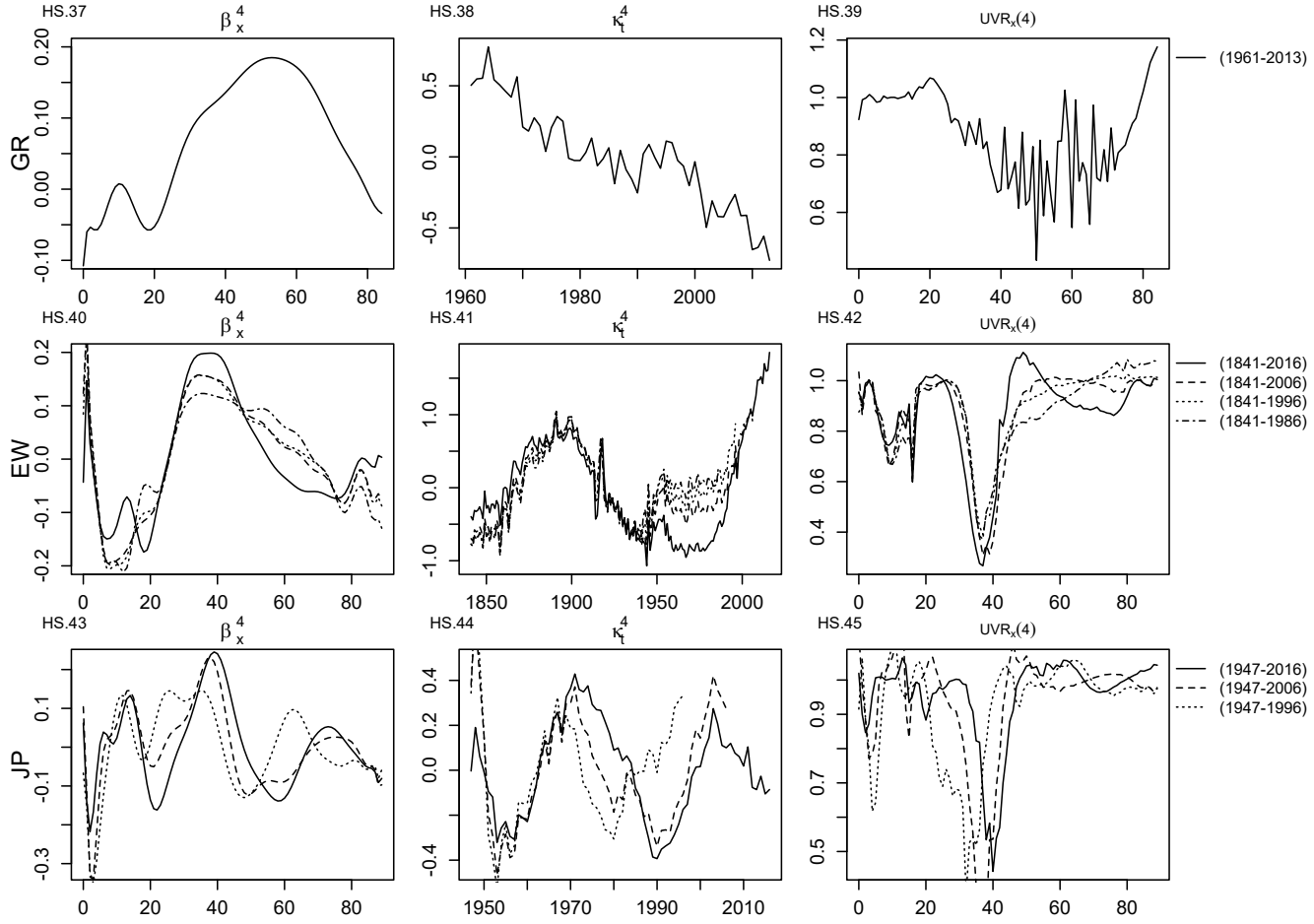

FIGURE A.13: HS model: fourth age–period component for countries GR, EW and JP.

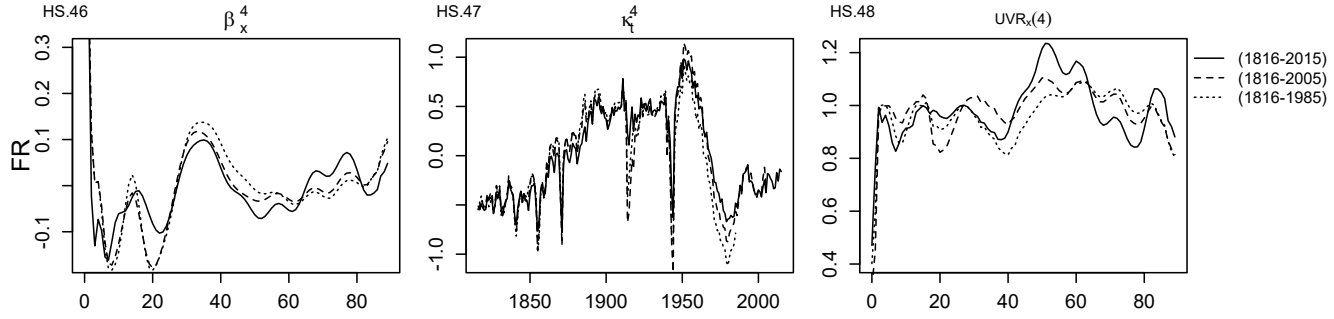

FIGURE A.14: HS model: fourth age–period component for FR country.

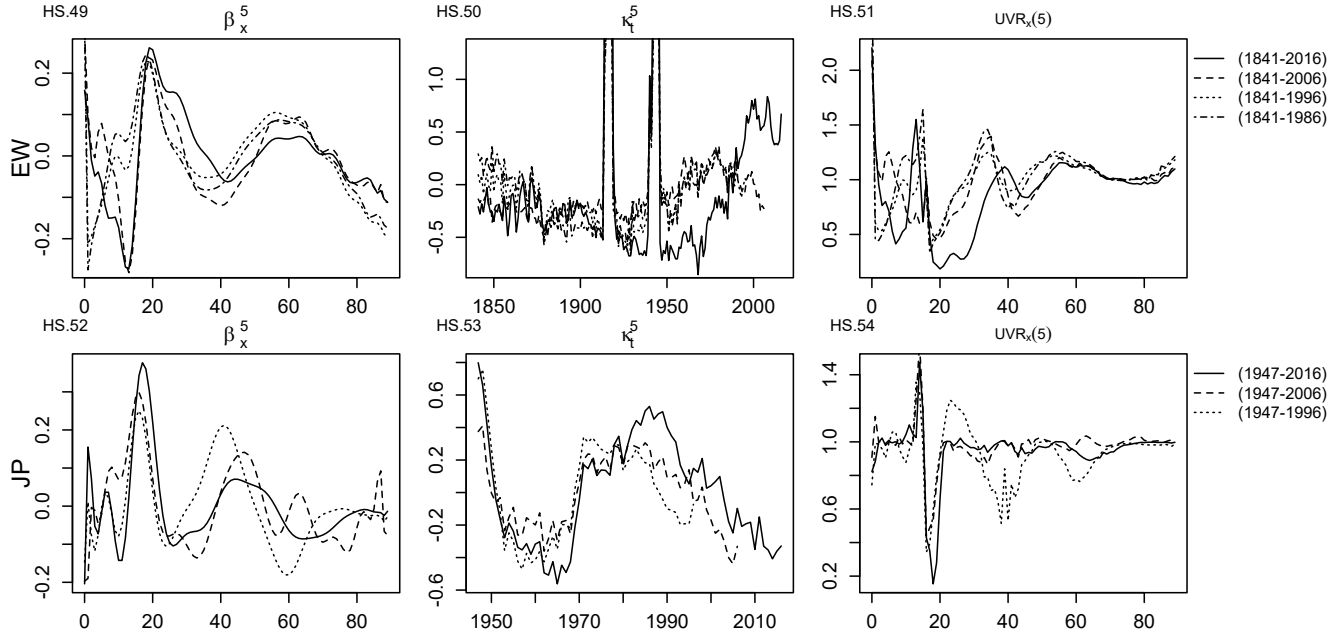

FIGURE A.15: HS model: fifth age–period component for EW and JP.

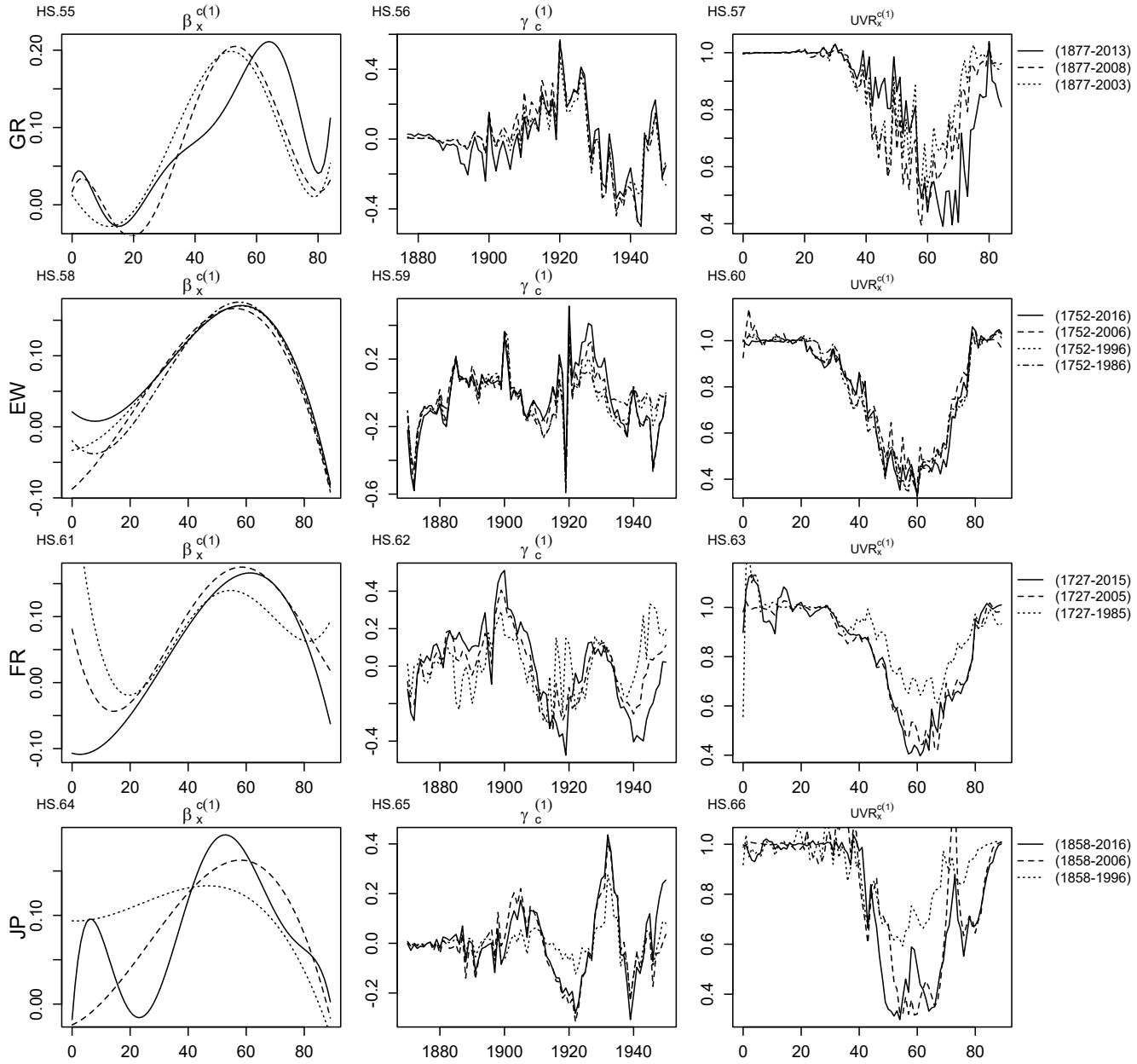

FIGURE A.16: HS model: first age–cohort component for all countries (GR, EW, FR, JP).

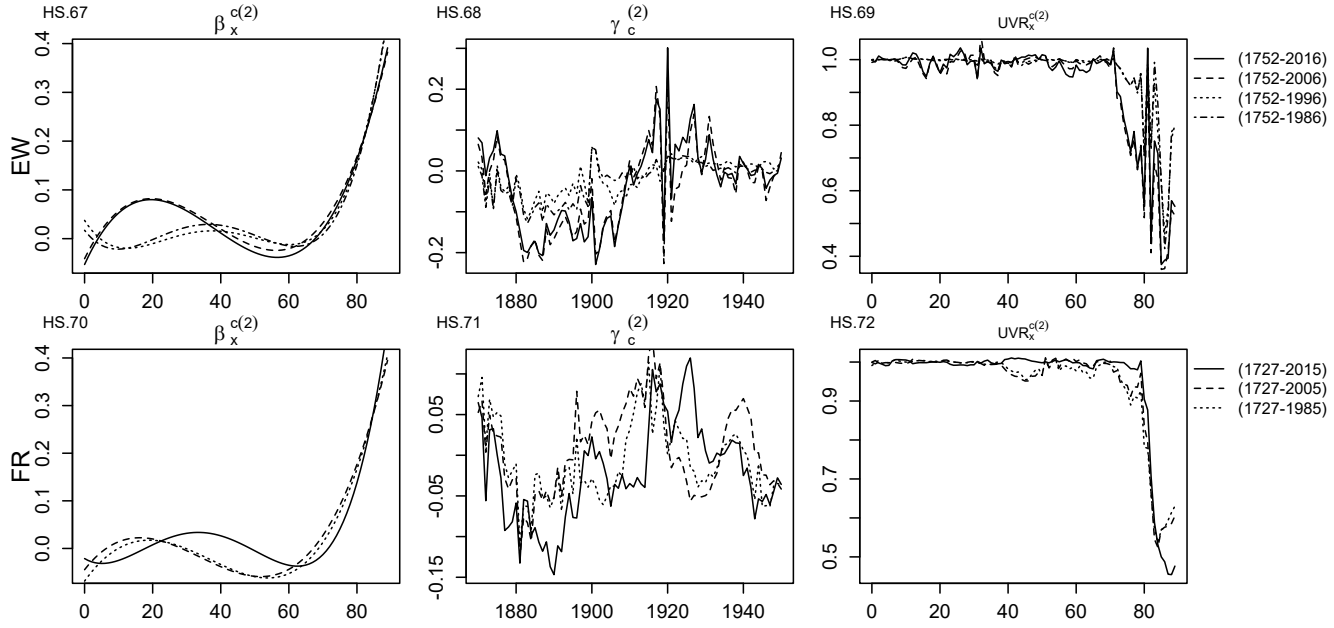

FIGURE A.17: HS model: second age–cohort component for EW and FR.

A.2 Deviance Residuals according to the proposed age–period model structure

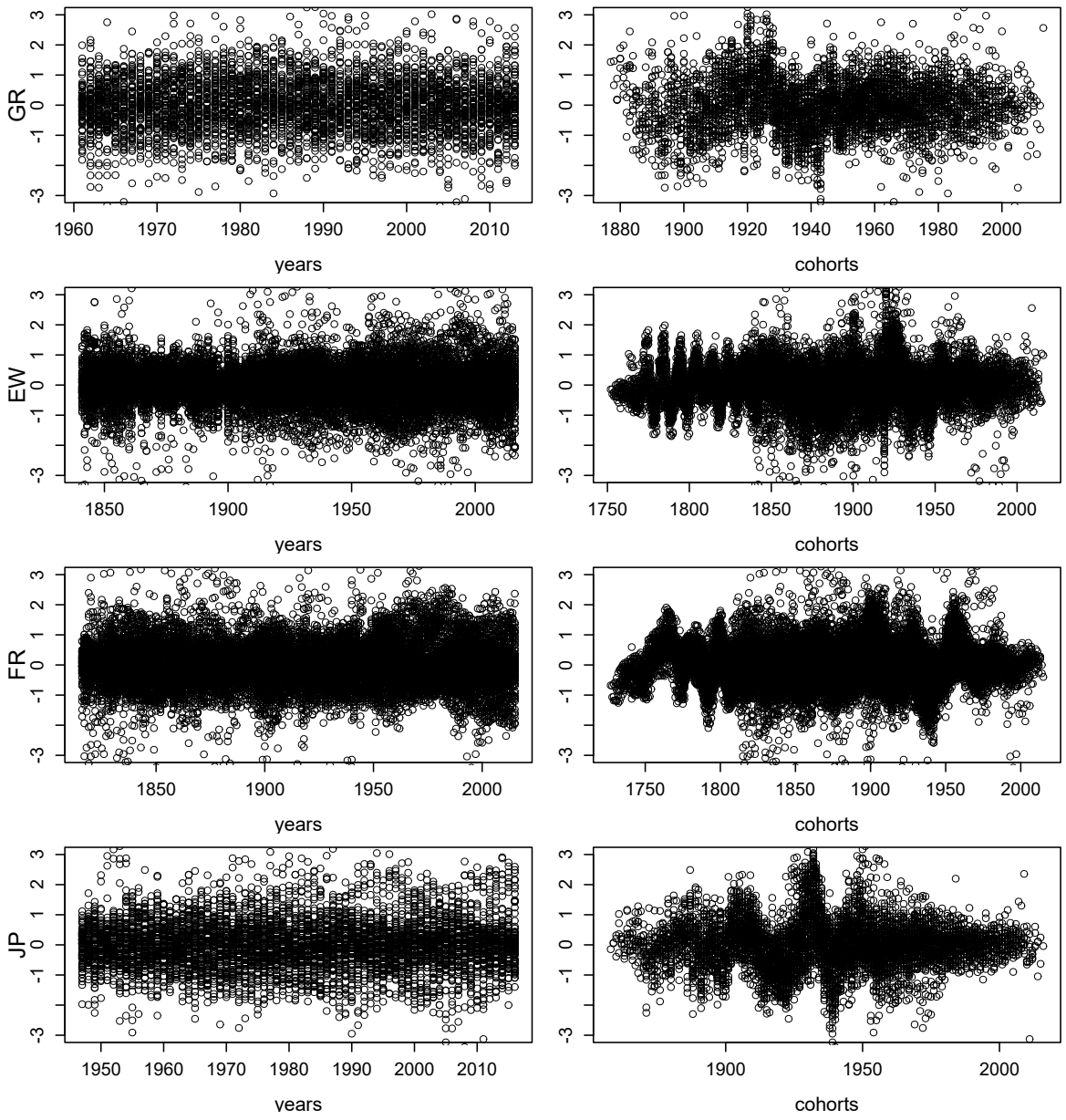

FIGURE A.18: Deviance Residuals presented in calendar years (left) and cohort (right) according to the HS model structure (4.3), for all countries (GR, EW, FR, JP).

A.3 Unexplained Variance for all models

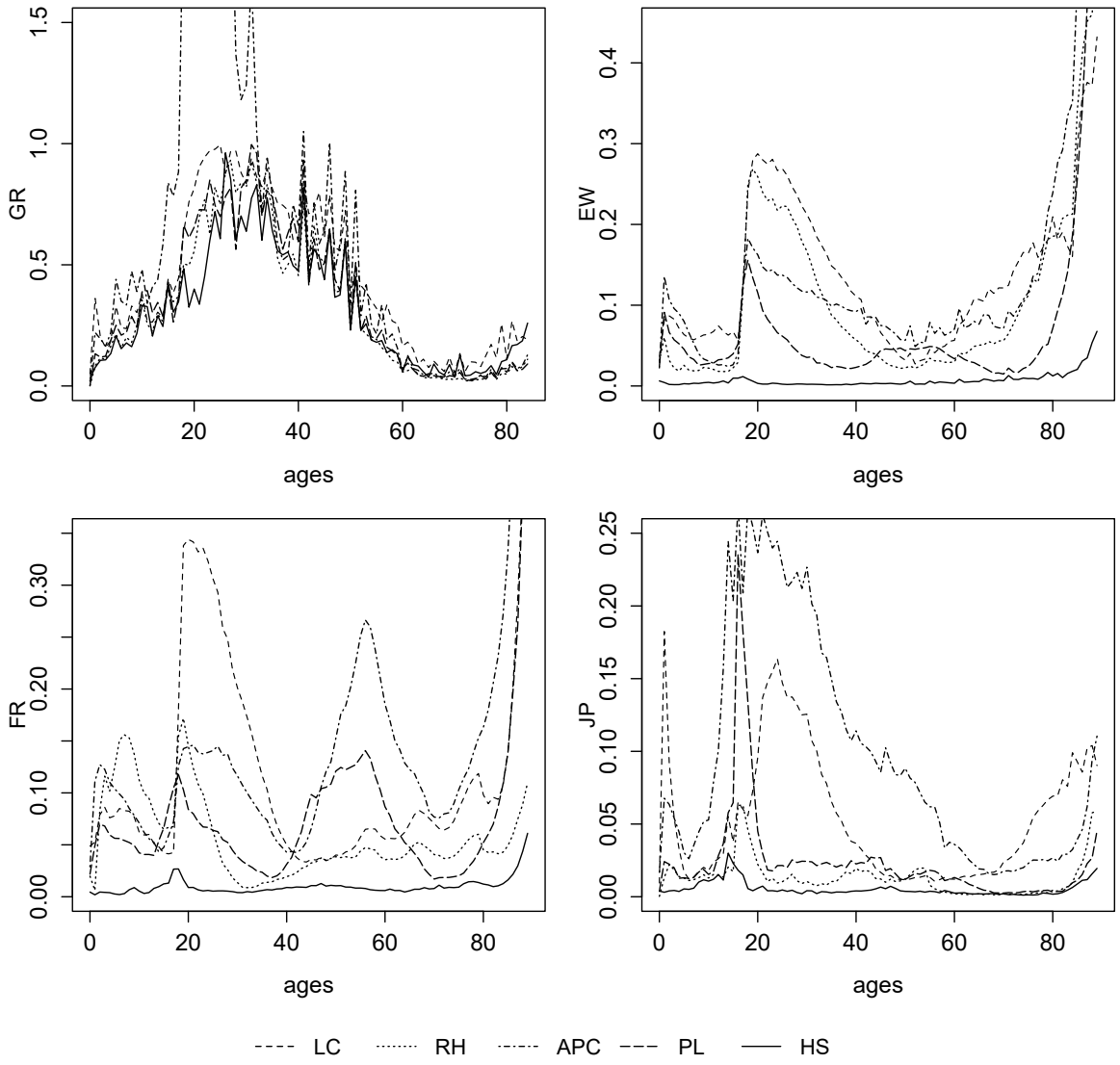

FIGURE A.19: Unexplained Variance for all ages and each model, for all countries (GR, EW, FR, JP).

A.4 Quantitative tests for the fitting process

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	221	-21,213.44	42,868.89	44,286.15	4.287	11.379
RH	441	**-18,633.10**	**38,148.19**	**40,976.30**	**3.790**	**9.460**
APC	272	-21,317.67	43,179.33	44,923.65	5.405	12.715
PL	378	-19,113.68	38,983.36	41,407.45	5.395	10.371
HS	447	-20,176.13	41,246.26	44,112.84	4.109	10.115

TABLE A.1: Greece, males: quantitative tests for the fitting process

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	354	-1,075,745.35	2,152,198.70	2,154,913.98	7.131	18.473
RH	707	-553,002.06	1,107,418.13	1,112,841.02	5.369	14.896
APC	528	-1,157,738.25	2,316,532.50	2,320,582.42	20.541	23.522
PL	880	-762,205.25	1,526,170.51	1,532,920.37	9.136	15.737
HS	1,598	**-135,970.84**	**275,137,68**	**287,394.81**	**0.329**	**3.696**

TABLE A.2: England & Wales, males: quantitative tests for the fitting process

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	378	-1,683,773.46	3,368,302.92	3,371,250.62	10.810	17.430
RH	755	-494,751.55	991,013.10	996,900.68	3.581	11.262
APC	576	-1,935,221.43	3,871,594.86	3,876,086.58	22.775	24.065
PL	976	-1,243,376.24	2,488,704.49	2,496,315.46	9.199	16.179
HS	1560	**-234,746.31**	**472,612.62**	**484,777.70**	**0.529**	**4.748**

TABLE A.3: France, males: quantitative tests for the fitting process

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	248	-136,173.79	272,843.59	274,517.17	1.918	9.668
RH	495	-53,081.31	107,152.61	110,493.02	0.541	4.772
APC	316	-228,562.40	457,756.80	459,889.26	4.425	14.414
PL	456	-93,807.81	188,527.62	191,604.85	1.321	7.524
HS	663	**-43,898.17**	**89,122.35**	**93,596.47**	**0.277**	**3.387**

TABLE A.4: Japan, males: quantitative tests for the fitting process

A.5 Quantitative tests for the forecasting process

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG	ARIMA		DLR	AVG
	RWD	BstAr		MSPE	RWD	BstAr		MAPE
LC	10.672	10.576	10.633	10.627	21.006	21.285	21.491	21.261
RH	9.697	9.792	9.564	9.684	16.952	17.016	16.865	16.944
APC	**8.378**	**8.325**	**8.483**	**8.395**	19.814	19.938	19.787	19.846
PL	9.181	18.251	16.081	14.504	**14.216**	16.386	14.959	**15.187**
HS	10.420	10.390	11.052	10.621	15.331	**16.098**	**14.651**	15.360

TABLE A.5: Greece: percentage error tests for 5 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG	ARIMA		DLR	AVG
	RWD	BstAr		MSPE	RWD	BstAr		MAPE
LC	10.398	10.333	10.985	10.572	20.320	20.964	21.932	21.072
RH	**8.361**	**8.581**	8.838	**8.593**	15.440	15.275	15.383	15.366
APC	10.054	10.077	10.206	10.112	21.364	21.599	21.037	21.333
PL	8.860	11.163	9.400	9.808	18.402	**15.148**	19.028	17.526
HS	8.691	10.390	**6.966**	8.682	**15.332**	15.857	**13.989**	**15.059**

TABLE A.6: Greece: percentage error tests for 10 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG	ARIMA		DLR	AVG
	RWD	BstAr		MSPE	RWD	BstAr		MAPE
LC	29.508	29.509	35.256	31.424	49.838	49.837	52.518	50.731
RH	271.861	271.861	148.885	230.869	58.085	58.085	48.008	54.726
APC	86.494	84.316	90.684	87.165	46.192	46.131	46.626	46.316
PL	586.784	371.845	49.022	335.884	62.748	55.635	33.099	50.494
HS	**3.221**	**7.135**	**2.101**	**4.152**	**13.271**	**20.800**	**9.887**	**14.653**

TABLE A.7: E&W: percentage error tests for 10 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG	ARIMA		DLR	AVG
	RWD	BstAr		MSPE	RWD	BstAr		MAPE
LC	27.069	26.967	30.428	28.155	45.517	45.316	49.972	46.935
RH	42.439	42.439	63.437	49.438	53.258	53.258	62.328	56.281
APC	81.839	72.442	104.406	86.229	45.766	45.547	48.532	46.615
PL	585.665	135.145	48.224	256.345	77.350	49.652	37.688	54.897
HS	**8.072**	**15.473**	**2.970**	**8.838**	**21.374**	**31.008**	**12.670**	**21.684**

TABLE A.8: E&W: percentage error tests for 20 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG	ARIMA		DLR	AVG
	RWD	BstAr		MSPE	RWD	BstAr		MAPE
LC	29.369	29.495	28.063	28.976	44.253	44.342	43.967	44.187
RH	53.006	53.006	63.493	56.502	59.295	59.295	63.043	60.544
APC	67.311	52.618	112.472	77.467	41.885	43.421	48.096	44.467
PL	285.011	27.417	34.811	115.746	73.084	**30.423**	37.915	47.141
HS	**16.632**	**24.894**	**4.638**	**15.388**	**30.189**	38.640	**17.094**	**28.641**

TABLE A.9: E&W: percentage error tests for 30 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG MSPE	ARIMA		DLR	AVG MAPE
	RWD	BstAr			RWD	BstAr		
LC	17.665	16.921	20.539	18.375	34.905	34.657	36.288	35.283
RH	16.023	15.415	16.093	15.844	33.337	31.565	30.131	31.678
APC	124.585	126.383	132.376	127.781	55.970	56.185	56.919	56.358
PL	425.771	1,659.819	138.278	741.289	73.922	132.045	51.059	85.675
HS	**2.868**	**5.497**	**1.354**	**3.240**	**12.929**	**19.007**	**8.961**	**13.632**

TABLE A.10: France: percentage error tests for 10 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG MSPE	ARIMA		DLR	AVG MAPE
	RWD	BstAr			RWD	BstAr		
LC	53.237	59.410	43.466	52.038	57.113	59.630	52.800	56.514
RH	53.144	53.861	43.539	50.181	61.797	62.209	56.704	60.237
APC	67.024	61.199	205.816	111.346	45.698	45.219	63.328	51.415
PL	135,771.806	141,779.438	2,821.128	93,457.457	1,012.532	1,026.364	277.732	772.209
HS	**19.314**	**27.480**	**8.387**	**18.394**	**34.344**	**40.784**	**19.145**	**31.424**

TABLE A.11: France: percentage error tests for 30 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG MSPE	ARIMA		DLR	AVG MAPE
	RWD	BstAr			RWD	BstAr		
LC	106.298	401.509	30.881	179.563	66.921	103.098	42.813	70.944
RH	54.834	55.401	57.353	55.863	63.138	63.485	64.525	63.716
APC	85.590	181.443	12,725.362	4,330.798	50.313	60.429	291.085	133.942
PL	937.737	919.067	8,253.562	3,370.122	112.916	118.272	355.095	195.428
HS	**34.427**	**45.672**	**8.998**	**29.699**	**37.558**	**43.695**	**18.577**	**33.277**

TABLE A.12: France: percentage error tests for 50 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG MSPE	ARIMA		DLR	AVG MAPE
	RWD	BstAr			RWD	BstAr		
LC	7.760	5.863	6.911	6.845	19.519	17.117	18.275	18.304
RH	3.354	4.639	2.788	3.594	15.298	14.943	12.435	14.225
APC	14.879	12.346	13.689	13.638	31.272	28.614	29.739	29.875
PL	4.219	3.583	4.131	3.978	14.937	14.094	14.558	14.530
HS	**1.689**	**3.414**	**0.930**	**2.011**	**9.122**	**12.222**	**6.770**	**9.371**

TABLE A.13: Japan: percentage error tests for 10 years out–of–sample forecasted mortality rates

	MSPE(%)				MAPE(%)			
	ARIMA		DLR	AVG MSPE	ARIMA		DLR	AVG MAPE
	RWD	BstAr			RWD	BstAr		
LC	10.948	15.332	8.516	11.599	24.708	30.667	21.348	25.574
RH	93.601	143.306	121.922	119.610	51.572	52.387	50.549	51.503
APC	20.432	15.873	17.310	17.872	37.895	32.581	33.981	34.819
PL	8.741	15.017	11.418	11.725	19.701	26.155	23.112	22.989
HS	**5.390**	**2.451**	**3.124**	**3.655**	**17.395**	**11.210**	**13.426**	**14.010**

TABLE A.14: Japan: percentage error tests for 20 years out–of–sample forecasted mortality rates

A.6 ARIMA and DLR models

		LC	RH		APC		PL				HS				
		$\kappa_t^{(1)}$	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\kappa_t^{(4)}$	$\kappa_t^{(5)}$
GR	1961-2008	(0,2,2)	(0,1,1)+c	(0,1,5)+c	(0,1,1)+c	(0,2,3)	(0,1,1)+c	(0,1,2)+c	(0,2,2)	(0,2,3)	(0,1,1)+c	(1,1,0)+c	(0,1,1)	-	-
	1961-2003	(0,2,2)	(4,2,0)	(0,2,3)	(0,1,1)+c	(0,2,3)	(0,1,1)+c	(0,1,1)+c	(0,2,2)	(1,2,2)	(0,1,1)+c	(5,1,0)	(0,1,1)	-	-
EW	1841-2006	(2,1,2)+c	(0,1,0)+c	(4,2,1)	(0,1,5)+c	(1,2,1)	(0,1,0)	(0,1,0)	(0,2,1)	(0,2,1)	(0,1,0)+c	(0,1,3)+c	(3,0,0)	(1,1,4)	(0,1,5)
	1841-1996	(2,1,2)+c	(0,1,0)+c	(2,2,2)	(0,1,5)+c	(1,2,1)	(0,1,0)	(0,1,0)	(0,2,1)	(1,2,1)	(0,1,0)+c	(0,1,3)+c	(3,0,0)	(2,1,2)	(2,0,1)
	1841-1986	(2,1,2)+c	(0,1,0)+c	(5,1,0)	(0,1,5)+c	(0,2,2)	(0,1,0)	(0,1,0)	(0,2,1)	(1,1,1)	(0,1,0)+c	(0,1,3)+c	(3,0,0)	(3,0,0)	(2,1,2)+c
FR	1816-2005	(0,1,5)+c	(1,2,2)	(1,1,2)+c	(0,1,3)	(0,2,1)	(0,1,5)	(0,1,5)	(5,1,0)	(3,2,0)	(1,1,1)+c	(1,1,1)	(0,1,5)	(1,1,3)	-
	1816-1985	(0,1,5)+c	(0,1,5)+c	(0,2,2)	(0,1,3)+c	(2,2,3)	(0,1,5)	(0,1,5)	(3,1,2)+c	(0,2,5)	(1,1,1)+c	(0,1,5)	(1,1,1)	(0,1,3)	-
	1816-1965	(0,1,5)	(1,1,1)+c	(4,2,1)	(1,1,1)	(2,2,3)	(0,1,5)	(0,1,5)	(3,1,2)	(0,0,5)	(1,1,1)+c	(0,1,5)	(0,1,1)	-	-
JP	1947-2006	(0,2,3)	(0,2,2)	(0,2,3)	(0,2,2)	(1,2,3)	(1,1,1)+c	(0,1,0)+c	(0,1,0)+c	(1,2,1)	(0,2,3)	(0,1,3)+c	(0,2,1)	(3,0,0)	(4,0,0)
	1947-1996	(0,2,2)	(3,2,0)	(1,1,0)+c	(3,2,0)	(1,2,1)	(0,2,1)	(0,1,0)+c	(0,1,0)+c	(0,2,2)	(0,1,3)+c	(0,1,3)+c	(0,2,1)	(2,0,0)	(2,0,1)

TABLE A.15: ARIMA(p,d,q) models for each stochastic model and country. The parameter c represents the drift value.

ϕ		LC	RH		APC		PL				HS				
		$\kappa_t^{(1)}$	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\gamma_c^{(1)}$	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\kappa_t^{(4)}$	$\kappa_t^{(5)}$
GR	1961-2008	1	0.814	0.867	1	0.818	1	0.768	0.980	0.953	1	1	1	-	-
	1961-2003	1	1	0.988	1	0.834	1	1	0.888	0.945	1	1	1	-	-
EW	1841-2006	0.969	0.944	0.958	0.948	1	0.949	1	0.971	0.731	0.897	1	0.852	1	0.939
	1841-1996	1	1	0.657	0.925	0.662	0.915	1	0.971	0.676	0.890	1	0.808	0.937	1
	1841-1986	0.921	1	0.839	0.862	0.899	0.892	1	0.968	0.632	0.888	1	1	1	0.939
FR	1816-2005	0.949	0.989	0.947	1	0.915	0.799	0.793	0.822	0.922	1	1	1	1	
	1816-1985	0.887	1	1	1	0.886	0.795	0.770	0.822	0.800	1	1	1	1	-
	1816-1965	1	1	1	1	1	0.792	0.737	0.805	0.648	1	1	1	-	-
JP	1947-2006	0.860	0.912	0.977	0.864	0.822	0.878	0.902	0.897	0.947	0.876	1	0.922	0.906	0.884
	1947-1996	0.856	0.886	0.814	0.859	0.851	0.872	0.894	0.880	0.952	0.879	1	0.920	0.889	0.933

TABLE A.16: The ϕ–parameters of the best DLR models for each stochastic mortality model and for each dataset.

A.7 Explanation Ratio

		LC	RH			APC			PL					HS							
		Total	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\kappa_t^{(4)}$	$\kappa_t^{(5)}$	$\gamma_c^{(1)}$	$\gamma_c^{(2)}$	Total
GR	1961-2013	70.85	18.65	60.53	79.18	36.07	25.04	61.11	16.36	4.45	-37.98	92.78	75.61	73.50	1.73	2.58	1.33		0.71		79.85
	1961-2008	71.13	3.56	74.59	78.15	30.76	29.98	60.74	18.90	6.12	-31.58	79.94	73.38	72.51	2.66	0.98			0.08		76.23
	1961-2003	71.48	42.53	34.19	76.72	27.26	36.00	63.26	24.48	7.42	-16.64	56.62	71.88	72.52	1.30	1.89			0.96		76.67
EW	1841-2016	87.98	-12.31	104.35	92.04	60.10	30.78	90.88	35.58	-8.27	47.02	20.57	94.90	92.75	3.49	1.43	1.25	0.63	0.06	0.02	99.63
	1841-2006	90.95	-239.76	336.60	96.84	57.06	32.91	89.97	30.55	-9.08	51.91	20.55	93.93	93.85	2.09	2.66	0.48	0.43	0.06	0.03	99.59
	1841-1996	92.84	-278.01	374.60	96.59	54.30	34.94	89.24	25.89	-7.84	52.81	21.72	92.58	92.27	3.35	2.89	0.41	0.50	0.07	0.01	99.50
	1841-1986	92.94	-249.26	345.65	96.39	52.28	35.70	87.98	13.25	-5.79	49.67	33.56	90.69	92.20	2.78	3.32	0.61	0.46	0.08	0.01	99.46
FR	1816-2015	89.93	-71.38	163.37	91.99	52.84	36.77	89.61	49.55	-43.19	33.90	53.90	94.16	93.91	1.98	3.06	0.31		0.05	0.02	99.33
	1816-2005	88.37	-84.32	178.41	94.09	48.35	39.83	88.18	38.29	6.35	35.30	12.94	92.88	92.79	2.08	3.86	0.47		0.06	0.03	99.29
	1816-1985	85.54	-82.55	176.90	94.35	41.24	44.23	85.47	22.97	7.47	41.15	17.59	89.18	89.33	6.33	2.52	0.71		0.04	0.04	98.97
	1816-1965	76.70	-182.46	274.75	92.29	41.74	36.00	77.74	26.62	13.55	35.37	6.03	81.57	84.82	10.95	2.13					97.90
JP	1947-2016	94.12	91.63	6.98	98.61	82.94	6.64	89.58	69.29	7.05	12.57	8.55	97.46	95.95	1.41	1.56	0.28	0.13	0.08		99.41
	1947-2006	94.58	78.34	20.50	98.84	81.62	7.81	89.43	57.67	9.83	9.66	19.75	96.91	95.08	2.54	1.22	0.21	0.18	0.03		99.26
	1947-1996	94.92	88.13	10.29	98.42	80.02	9.90	89.92	56.21	11.33	8.33	20.34	96.21	95.18	2.68	0.89	0.26	0.35	0.04		99.40

TABLE A.17: Explanation Ratio values for all stochastic mortality models and for all datasets (all values are percentages (%)).

A.8 Results for the fitting process using shorter datasets

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	234	-38,293.21	77,054.41	78,581.30	0.889	6.621
RH	467	**-25,891.57**	**52,717.14**	**55,764.39**	0.460	4.172
APC	288	-36,943.05	74,462.09	76,341.34	1.301	7.630
PL	400	-27,041.30	54,882.60	57,492.66	0.895	5.236
HS	569	-27,718.17	56,574.35	60,287.16	**0.342**	**4.171**

TABLE A.18: England & Wales, males: quantitative tests for the fitting process using shorter datasets

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	233	-40,757.42	81,980.84	83,497.00	0.968	6.516
RH	465	**-27,624.86**	**56,179.72**	**59,205.54**	0.419	4.140
APC	286	-39,793.13	80,158.26	82,019.30	0.793	6.237
PL	396	-29,615.83	60,023.65	62,600.48	0.573	4.794
HS	491	-27,944.53	56,871.06	60,066.06	**0.333**	**3.787**

TABLE A.19: France, males: quantitative tests for the fitting process using shorter datasets

Model	npar	Log-Likelihood	AIC	BIC	MSPE(%)	MAPE(%)
LC	234	-45,914.59	92,297.18	93,824.07	0.621	5.682
RH	467	**-32,388.50**	**65,711.01**	**68,758.26**	0.361	3.976
APC	288	-58,615.67	117,807.34	119,686.59	1.219	7.612
PL	400	-41,060.84	82,921.69	85,531.75	0.613	5.233
HS	579	-32,510.93	66,179.85	69,957.92	**0.307**	**3.481**

TABLE A.20: Japan, males: quantitative tests for the fitting process using shorter datasets

		LC	RH			APC			PL					HS						
		Total	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\gamma_c^{(1)}$	Total	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$	$\kappa_t^{(4)}$	$\kappa_t^{(5)}$	$\gamma_c^{(1)}$	Total
EW	1961-2016	92.89	62.19	34.26	96.45	77.49	12.46	89.95	62.28	-11.74	19.80	18.14	94.48	94.07	1.59	0.60	0.68		0.24	97.18
FR	1961-2015	93.27	43.10	53.86	96.96	81.82	12.19	94.01	71.85	-0.07	-20.66	44.91	96.03	92.82	1.44	2.15	1.28		0.14	97.82
JP	1961-2015	96.55	76.48	21.58	97.98	87.50	5.08	92.58	84.56	-1.85	8.62	5.60	96.93	96.69	0.53	0.93	0.28	0.15	0.11	98.68

TABLE A.21: Explanation Ratio values for all stochastic mortality models, for shorter datasets (all values are percentages (%)).

Appendix B

Evaluation results of the experimental testbeds of Chapter 4

B.1 Definition of User-Defined Link Functions in Matlab

We consider the class of models for the probability of deaths with Binomial errors and a user-defined link function. We suppose $D \sim B(E, q)$ have the Binomial distribution and let $Q = D/E^0$ be the random variable corresponding to q.

Here, $q = E(Q)$ is the mean of Q, E^0 is the initial exposure and η is a linear function of the explanatory variables. A user-defined link in MATLAB requires three functions: the link function itself, i.e., η as a function of q, the inverse link function, i.e., q as a function of η, and the derivative of η with respect to q. As defined in Section 4.2.2, we reported three cases depending on the cumulative distribution chosen.

1. The cumulative distribution as link function g, maps q, $0 < q < 1$, to $-\infty < F^{-1}(q;\xi,\theta) < \infty$, so that:

$$\eta = F^{-1}(q;\xi,\theta)$$

$$q = F\left(\eta;\xi,\theta\right)$$

$$\frac{d\eta}{dq} = \left(F^{-1}(q;\xi,\theta)\right)' = \frac{1}{f(F^{-1}(q;\xi,\theta);\xi,\theta)}$$

2. The cumulative distribution as link function, g, maps q, $0 < q < 1$, to $0 < F^{-1}(q;\xi,\theta) < \infty$, so the logarithmic form of the cumulative distribution is needed to map q, to $-\infty < \log(F^{-1}(q;\xi,\theta)) < \infty$, the natural scale for regression, so that:

$$\eta = \log\left(F^{-1}(q;\xi,\theta)\right)$$

$$q = F\left(\exp(\eta);\xi,\theta\right)$$

$$\frac{d\eta}{dq} = \frac{1}{F^{-1}(q;\xi,\theta)} \cdot \left(F^{-1}(q;\xi,\theta)\right)' = \frac{1}{F^{-1}(q;\xi,\theta) \cdot f(F^{-1}(q;\xi,\theta);\xi,\theta)}$$

3. The cumulative distribution as link function, g, maps q, $0 < q < 1$, to $0 < F^{-1}(q;\xi,\theta) < 1$, so we need the logit of the cumulative distribution so that maps q, to $-\infty < \text{logit}(F^{-1}(q;\xi,\theta)) < \infty$, the natural scale for regression, so that:

$$\eta = \text{logit}\left(F^{-1}(q;\xi,\theta)\right) = \log\left(\frac{F^{-1}(q;\xi,\theta)}{1 - F^{-1}(q;\xi,\theta)}\right)$$

$$q = F\left(\frac{\exp(\eta)}{1+\exp(\eta)};\xi,\theta\right)$$

$$\frac{d\eta}{dq} = \frac{\left(F^{-1}(q;\xi,\theta)\right)'}{F^{-1}(q;\xi,\theta)\cdot(1-F^{-1}(q;\xi,\theta))} = \frac{1}{f\left(F^{-1}(q;\xi,\theta)\right)\cdot F^{-1}(q;\xi,\theta)\cdot(1-F^{-1}(q;\xi,\theta))}$$

Note that there is no need to define the derivative of any cumulative distribution, as this can be achieved by using the probability distribution function as described below.

By using the chain rule:

$$f(f^{-1}(x)) = x$$

and therefore

$$\frac{d}{dx}f(f^{-1}(x)) = \frac{d}{dx}x = 1 \tag{B.1}$$

and also by chain rule

$$\frac{d}{dx}f(f^{-1}(x)) = f'(f^{-1}(x))\cdot(f^{-1})'(x) \tag{B.2}$$

Thus, by (B.1) and (B.2), we have

$$f'(f^{-1}(x))\cdot(f^{-1})'(x) = 1$$

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

Thus, in our case, by replacing f with F, we have

$$(F^{-1})'(x) = \frac{1}{F'(F^{-1})(x)} = \frac{1}{f(F^{-1}(x))}$$

In addition, the Matlab code for the user-defined link function is:

```
link = @(mu) log(gpinv(mu, xi, theta));
derlink = @(mu) 1./(gpinv(mu,xi,theta).*gppdf(gpinv(mu,xi,theta),xi,theta));
invlink = @(eta) gpcdf(exp(eta),xi,theta);
new_F = {link, derlink, invlink};

B = glmfit(Lx,qtx,'binomial','link',new_F,'weights',etx0,'constant','off')
```

where B contains the GLM-estimated parameters, etx0 contains the initial exposures, Lx is the matrix of the orthonormal polynomials and qtx is the probability of deaths.

B.2 Age–Period, Age–Cohort Components and Unexplained Variance Ratio Graphical Representations for E&W Dataset

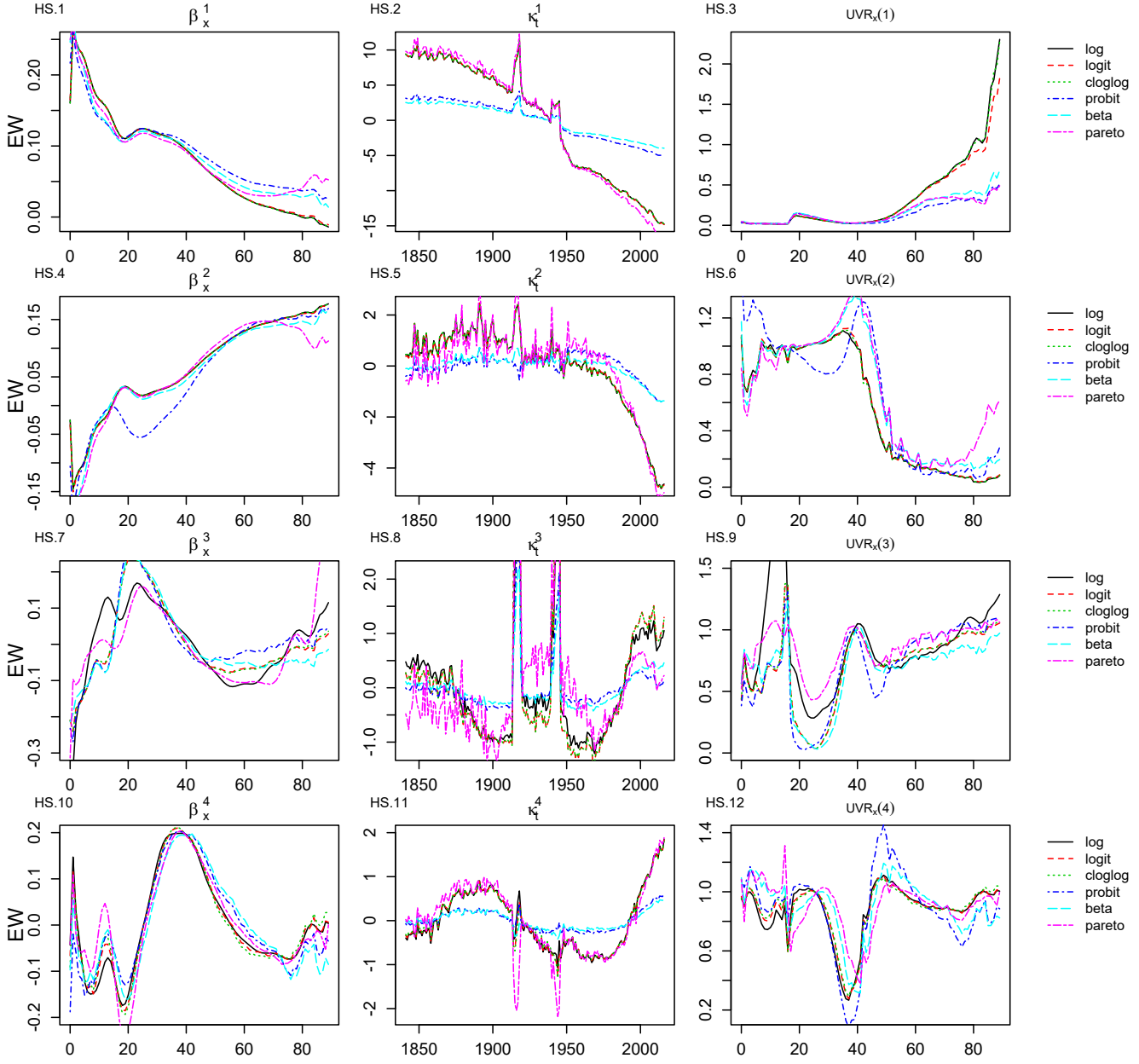

FIGURE B.1: First to fourth age-period component for E&W dataset.

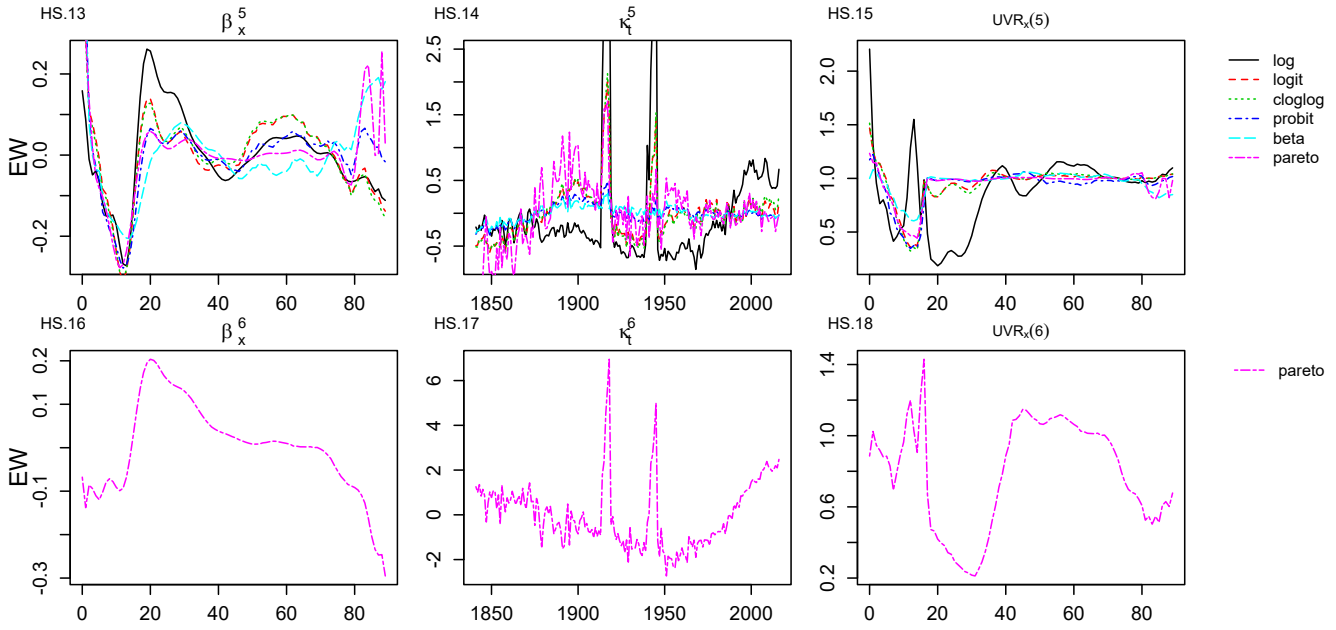

FIGURE B.2: Fifth and sixth age-period component for E&W dataset.

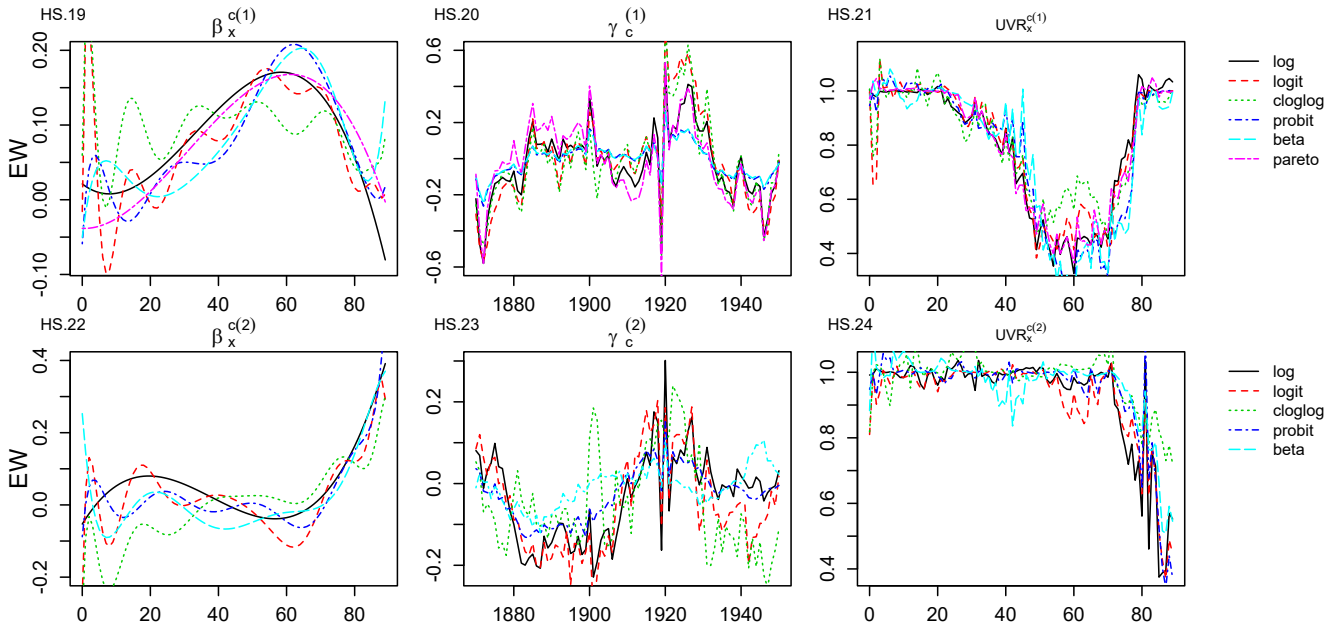

FIGURE B.3: First and second age-cohort component for E&W dataset.

B.3 Age–Period, Age–Cohort Components and Unexplained Variance Ratio Graphical Representations for GR Dataset

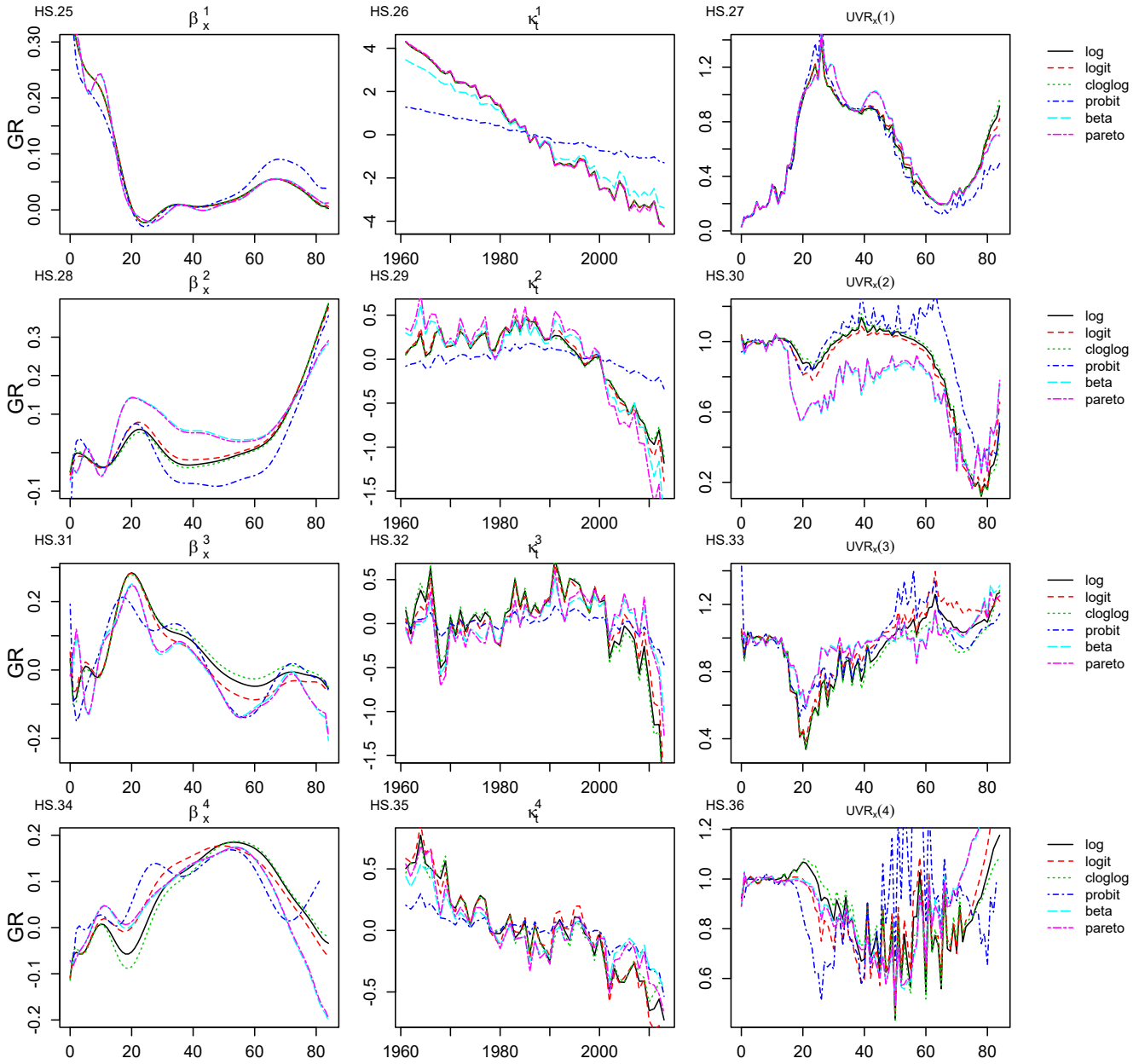

FIGURE B.4: First to fourth age-period component for GR dataset.

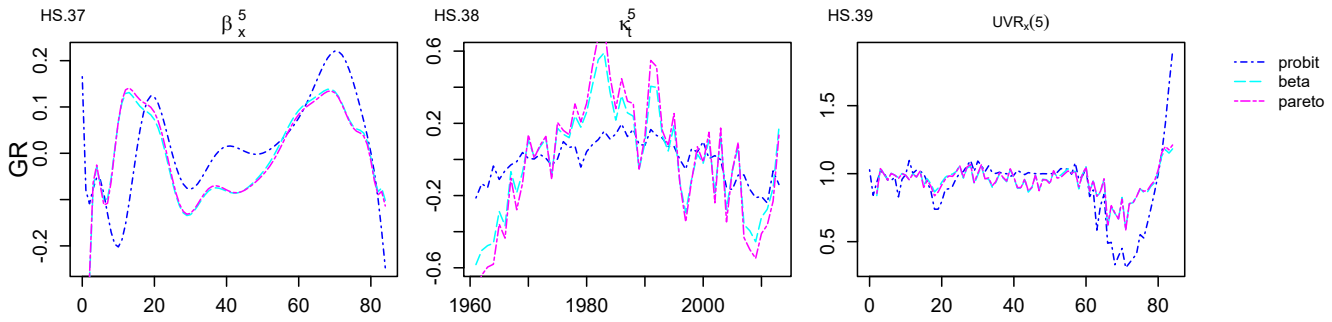

FIGURE B.5: Fifth age-period component for GR dataset.

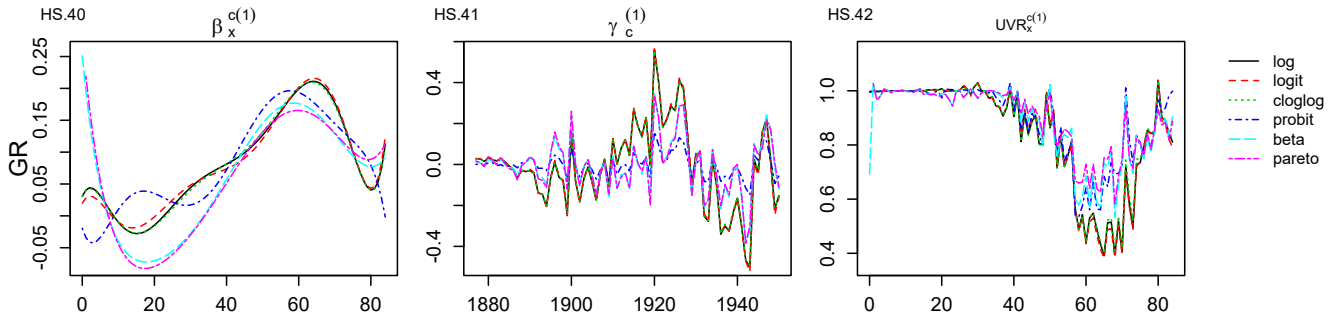

FIGURE B.6: Age-cohort component for GR dataset.

Appendix C

Updated HS Pseudocodes of Chapter 5

Algorithm 4: `hsfit()` age-period components estimation workflow

Input: D : $(n \times \text{a})$-matrix with values $D_{t,x}, \forall t, x$
Input: E : $(n \times \text{a})$-matrix with values $E_{t,x}, \forall t, x$
Output: M : $(n \times \text{a})$-matrix representing $\log(\hat{m}_{t,x}) = \tilde{\alpha}_x + \sum_{i=1}^{p} \beta_x^{(i)} \cdot \kappa_t^{(i)}$

Define parameters
$x \leftarrow x_1, \ldots, x_{\text{a}}$; $\text{a} \leftarrow \text{length}(x)$; `// ages`
$t \leftarrow t_1, \ldots, t_n$; $n \leftarrow \text{length}(t)$; `// calendar years`
$L \leftarrow$ create orthonormal polynomials using Gram–Schmidt process;
$[L_{best}, k_1] \leftarrow \text{MCDM}(L)$; `// Best orthon.pol. with degree of` $k_1 - 1$ `using MCDM`
$B \leftarrow \text{GLM}(L_{best}, D)$ with offset $\log(E)$, $\forall t$;
$A \leftarrow \text{Cov}(B)$;
initial_$s \leftarrow$ is a set of values around the area of the $\text{Var}(b_{t_1,0})$;
for $s \in$ initial_s **do**
  $U \leftarrow \text{SPCA}(A, s)$;
  $P \leftarrow \text{eig}(U)$; `// P : eigenvectors of U`
  $\bar{b} \leftarrow \text{mean}(B)$;
  $B^r \leftarrow B - \bar{b}$;
  $\tilde{\alpha}_x \leftarrow \bar{b} \cdot L$; `// main age-profile`
  $Y \leftarrow B^r \cdot P$; `//` $Y = \{\kappa^{(i)}\}$ `i-period effect,` $\forall i \in [1, k_1]$
  $G \leftarrow P^{-1} \cdot L^T$; `//` $G = \{\beta^{(i)}\}$ `i-age effect,` $\forall i \in [1, k_1]$
  for $i \leftarrow 1$ **to** k_1 **do**
    $[points, UVR_{values}] \leftarrow$ get local minima of UVR(i)
    $threshold \leftarrow$ get cluster thresholds ($[points, UVR_{values}]$) `// Define UVR(i) thresholds`
    if ($threshold > \delta$) **and** ($UVR(i)$ reveals unique age cluster) **then**
      $Components \leftarrow$ keep this i-component to the model structure
    end
  end
  cand_solutions $\leftarrow$ calculate evaluation metrics for *Components*;
  `// cand_solutions is a 2D array holding the candidate p components`
  `// repeat the same procedure for the next` s `value`
end
optimal_solution $\leftarrow \text{MCDM}(method, cand\_solutions, weights)$; `// Define best p using MCDM`
`// optimal_solution holds the results of the optimum` s `value and the optimum age-period components (`p`)`
for $i \leftarrow 1$ **to** n **do**
  `// Generation of log-graduated mortality rates for age-period effects`
  $M(i, 1 : \text{a}) \leftarrow Y(i, 1 : p) \cdot G(1 : p, 1 : \text{a}) + \tilde{\alpha}$;
end
return M;

Algorithm 5: `hsfit()` age-cohort components estimation workflow

Input: $D : (n_c \times a)$-matrix with values $D_{c,x}, \forall c, x$
Input: $E : (n_c \times a)$-matrix with values $E_{c,x}, \forall c, x$
Input: $M : (n_c \times a)$-matrix with values $M_{c,x}, \forall c, x$
Output: $M^c : (n_c \times a)$-matrix representing $\log(\hat{\hat{m}}_{c,x}) = \tilde{\alpha}^c_x + \sum_{j=1}^{q} \beta^{c(j)}_x \cdot \gamma^{(j)}_c$

Define parameters
$x \leftarrow x_1, \ldots, x_a$; $a \leftarrow \text{length}(x)$; `// ages`
$c_1 \leftarrow t_1 - x_a$; $c_{n_c} \leftarrow t_n - x_1$;
$c \leftarrow c_1, \ldots, c_{n_c}$; $n_c \leftarrow \text{length}(c)$; `// cohorts`
$L^c \leftarrow$ create orthonormal polynomials using Gram–Schmidt process;
$[L^c_{best}, k_2] \leftarrow \text{MCDM}(L^c)$; `// Best orthon.pol. with degree of` $k_2 - 1$ `using MCDM`
$B^c \leftarrow \text{GLM}(L^c_{best}, D)$, offset: $\log(E)$ and the age–period effects, (M), $\forall c$;
$P^c \leftarrow \text{PCA}(B^c)$; `//` P^c `: eigenvectors of` B^c
$\bar{b}^c \leftarrow \text{mean}(B^c)$;
$B^r \leftarrow B^c - \bar{b}^c$;
$\tilde{\alpha}^c \leftarrow \bar{b}^c \cdot L^c$; `// main age-profile`
$Y^c \leftarrow B^r \cdot P^c$; `//` $Y^c = \{\gamma^{(j)}\}$ `j-cohort effect,` $\forall i \in [1, k_2]$
$G^c \leftarrow (P^c)^{-1} \cdot (L^c)^T$; `//` $G^c = \{\beta^{c(j)}_x\}$ `j-age effect,` $\forall i \in [1, k_2]$
for $i \leftarrow 1$ **to** k_2 **do**
  $[points, UVR^c_{values}] \leftarrow$ get local minima of $UVR^c(i)$
  $threshold \leftarrow$ get cluster thresholds $([points, UVR^c_{values}])$
  `// Define` $UVR^c(i)$ `thresholds`
  if ($threshold > \delta$) **and** ($UVR^c(i)$ reveals unique age cluster) **then**
    include this i-component to the model structure
  end
end
`// after UVR criterion we have concluded in the optimum age-cohort components (`q`)`
for $i \leftarrow 1$ **to** n_c **do**
  `// Generation of log-graduated mortality rates for age-cohort effects`
  $M^c(i, 1:a) \leftarrow Y^c(i, 1:q) \cdot G^c(1:q, 1:a) + \tilde{\alpha}^c$;
end
return M^c;

Bibliography

Aggarwal, A., Beck, M. B., Cann, M., Ford, T., Georgescu, D., Morjaria, N., Smith, A., Taylor, Y., Tsanakas, A., Witts, L. & Ye, I. (2016), 'Model risk – daring to open up the black box', *British Actuarial Journal* **21**(2), 229–296.

Booth, H., Hyndman, R., Tickle, L. & De Jong, P. (2006), 'Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions', *Demographic Research* **15**(9), 289–310.

Booth, H., Maindonald, J. & Smith, L. (2002), 'Applying lee-carter under conditions of variable mortality decline', *Population Studies* **56**(3), 325–336.

Booth, H. & Tickle, L. (2008), 'Mortality modelling and forecasting: a review of methods', *Annals of Actuarial Science* **3**(1-2), 3–43.

Bradley, J. V. & Bradley, J. V. (1968), 'Distribution-free statistical tests'.

Brillinger, D. (1986), 'The natural variability of vital rates and associated statistics', *Biometrics* **42**(4), 693–734.

Brock, A. & Griffiths, C. (2003), 'Trends in the mortality of young adults aged 15-44 in England and Wales 1961–2001', *Health Statistics Quarterly* **19**, 22–31.

Brouhns, N., Denuit, M. & Van Keilegom, I. (2005), 'Bootstrapping the poisson log-bilinear model for mortality forecasting', *Scandinavian Actuarial Journal* **2005**(3), 212–224.

Brouhns, N., Denuit, M. & Vermunt, J. K. (2002), 'A poisson log-bilinear regression approach to the construction of projected lifetables', *Insurance: Mathematics and Economics* **31**(3), 373–393.

Butt, Z., Haberman, S. & Shang, H. L. (2014), '**ilc**: Lee-carter mortality models using iterative fitting algorithms', *URL http://cran.r-project.org/package=ilc* .

Cairns, A., Blake, D. & Dowd, K. (2006*a*), 'Pricing death: Frameworks for the valuation and securitization of mortality risk', *ASTIN Bulletin: The Journal of the IAA* **36**(1), 79–120.

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

Cairns, A., Blake, D. & Dowd, K. (2008), 'Modelling and management of mortality risk: a review', *Scandinavian Actuarial Journal* **2008**(2-3), 79–113.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D. & Khalaf-Allah, M. (2011), 'Mortality density forecasts: An analysis of six stochastic mortality models', *Insurance: Mathematics and Economics* **48**(3), 355–367.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A. & 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.

Clayton, D. & Schifflers, E. (1987), 'Models for temporal variation in cancer rates. II: Age-period-cohort models', *Statistics in Medicine* **6**(4), 469–481.

Coughlan, G., Epstein, D., Ong, A., Sinha, A., Hevia-Portocarrero, J., Gingrich, E., Khalaf-Allah, M. & Joseph, P. (2007), 'Lifemetrics: A toolkit for measuring and managing longevity and mortality risks', *Technical document. JPMorgan Pension Advisory Group* .

Currie, I. D. (2006), 'Smoothing and forecasting mortality rates with P-splines', *Talk given at the Institute of Actuaries, June 2006* .

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

Demšar, J. (2006), 'Statistical comparisons of classifiers over multiple data sets', *Journal of Machine Learning Research* **7**, 1–30.

Deprez, P., Shevchenko, P. V. & Wüthrich, M. V. (2017), 'Machine learning techniques for mortality modeling', *European Actuarial Journal* **7**(2), 337–352.

Dixon, W. J. & Mood, A. M. (1946), 'The statistical sign test', *Journal of the American Statistical Association* **41**(236), 557–566.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Epstein, D. & Khalaf-Allah, M. (2010), 'Evaluating the goodness of fit of stochastic mortality models', *Insurance: Mathematics and Economics* **47**(3), 255–265.

Dunn, O. J. (1961), 'Multiple comparisons among means', *Journal of the American Statistical Association* **56**(293), 52–64.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation', *Econometrica* **50**(4), 987–1007.

Eurostat (), 'Database-Your key to European Statistics.', http://ec.europa.eu/eurostat/data/database.

Federal Reserve (2011), Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency, *Supervisory guidance on model risk management.*

Fishburn, P. C. (1967), 'Additive utilities with incomplete product sets: Application to priorities and assignments', *Operations Research* **15**(3), 537–542.

Friedman, M. (1940), 'A comparison of alternative tests of significance for the problem of m rankings', *The Annals of Mathematical Statistics* **11**(1), 86–92.

Glei, D. & Horiuchi, S. (2007), 'The narrowing sex differential in life expectancy in high-income populations: Effects of differences in the age pattern of mortality', *Population Studies* **61**(2), 141–159.

Goodman, L. (1979), 'Simple models for the analysis of association in cross-classifications having ordered categories', *Journal of the American Statistical Association* **74**(367), 537–552.

Haberman, S. & Renshaw, A. E. (1996), 'Generalized linear models and actuarial science', *Journal of the Royal Statistical Society: Series D (The Statistician)* **45**(4), 407–436.

Hatzopoulos, P. & Haberman, S. (2009), 'A parameterized approach to modeling and forecasting mortality', *Insurance: Mathematics and Economics* **44**(1), 103–123.

Hatzopoulos, P. & Haberman, S. (2011), 'A dynamic parameterization modeling for the age-period-cohort mortality', *Insurance: Mathematics and Economics* **49**(2), 155–174.

Hatzopoulos, P. & Haberman, S. (2015), 'Modeling trends in cohort survival probabilities', *Insurance: Mathematics and Economics* **64**, 162 – 179.

Hatzopoulos, P. & Sagianou, A. (2020), 'Introducing and evaluating a new multiple-component stochastic mortality model', *North American Actuarial Journal* **24**(3), 393–445.

Hatzopoulos, P. & Sagianou, A. (n.d.), 'HSTool: A matlab toolbox for the hatzopoulos–sagianou multiple-component stochastic mortality model', *Working paper* .

HMD-Greek-data (), 'Human mortality database-Greek data', https://www.mortality.org/cgi-bin/hmd/country.php?cntr=GRC&level=1.

Hobcraft, J., Menken, J. & Preston, S. (1985), *Age, Period, and Cohort Effects in Demography: A Review*, New York, NY.

Horiuchi, S. & Wilmoth, J. (1998), 'Deceleration in the age pattern of mortality at olderages', *Demography* **35**(4), 391–412.

Human-Mortality-Database (), 'University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany)', http://www.mortality.org.

Hunt, A. & Blake, D. (2021), 'On the structure and classification of mortality models', *North American Actuarial Journal* **25**(sup1), S215–S234.

Hunt, A. & Villegas, A. M. (2015), 'Robustness and convergence in the Lee-Carter model with cohort effects', *Insurance: Mathematics and Economics* **64**, 186 – 202.

Hwang, C.-L. & Yoon, K. (1981), *Methods for Multiple Attribute Decision Making*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 58–191.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F. (2019), 'forecast: Forecasting functions for time series and linear models'. R package version 8.17.0. Available in https://cran.r-project.org/web/packages/forecast/index.html.

Hyndman, R. J., Booth, H., Tickle, L. & Maindonald, J. (2015), 'demography: Forecasting mortality, fertility, migration and population data'. R package version 1.22. Available in https://cran.r-project.org/package=demography.

Hyndman, R. J. & Shahid Ullah, M. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Iman, R. L. & Davenport, J. M. (1980), 'Approximations of the critical region of the fbietkan statistic', *Communications in Statistics - Theory and Methods* **9**(6), 571–595.

Janssen, F., for The Netherlands Epidemiology, A. K. & of Morbidity research group†, D. C. (2005), 'Cohort patterns in mortality trends among the elderly in seven european countries, 1950-99', *International Journal of Epidemiology* **34**(5), 1149–1159.

Jarque, C. M. & Bera, A. K. (1987), 'A test for normality of observations and regression residuals', *International Statistical Review / Revue Internationale de Statistique* **55**(2), 163–172.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Lee, R. & Carter, L. (1992), 'Modeling and forecasting U.S. mortality', *Journal of the American Statistical Association* **87**(419), 659–671.

Lee, R. & Miller, T. (2001), 'Evaluating the performance of the lee-carter method for forecasting mortality', *Demography* **38**(4), 537–549.

Levantesi, S. & Pizzorusso, V. (2019), 'Application of machine learning to mortality modeling and forecasting', *Risks* **7**(1).

Li, N., Lee, R. & Tuljapurkar, S. (2004), 'Using the lee-carter method to forecast mortality for populations with limited data', *International Statistical Review / Revue Internationale de Statistique* **72**(1), 19–36.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**(2), 297–303.

Luss, R. & d'Aspremont, A. (2006), 'DSPCA: a toolbox for sparse principal component analysis'.

Marugame, T., Yoshimi, I., Kamo, K.-i., Imamura, Y., Kaneko, S., Mizuno, S. & Sobue, T. (2005), 'Trends in Lung Cancer Mortality Among Young Adults in Japan', *Japanese Journal of Clinical Oncology* **35**(4), 177–180.

Massey, F. J. (1951), 'The kolmogorov-smirnov test for goodness of fit', *Journal of the American Statistical Association* **46**(253), 68–78.

MATLAB (n.d.*a*), 'Beta distribution', https://www.mathworks.com/help/stats/beta-distribution.html.

MATLAB (n.d.*b*), 'Generalized pareto distribution', https://www.mathworks.com/help/stats/generalized-pareto-distribution.html.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models, second edition*, Chapman and Hall, London.

Meslé, F. (2006), 'Progrès récents de l'espérance de vie en France. les hommes comblent une partie de leur retard [Recent improvements in life expectancy in France: Men are starting to catch up]', *Population* **61**(4), 365–387.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R. & Muthuraman, K. (2013), 'Modeling and forecasting mortality rates', *Insurance: Mathematics and Economics* **52**(2), 275–285.

Nemenyi, P. B. (1963), *Distribution-free Multiple Comparisons*, Princeton University.

Osmond, C. (1985), 'Using age, period and cohort models to estimate future mortality rates', *International journal of epidemiology* **14**(1), 124–129.

Plat, R. (2009), 'On stochastic mortality modeling', *Insurance: Mathematics and Economics* **45**(3), 393–404.

Rae, R. A., Barrett, A., Brooks, D., Chotai, M. A., Pelkiewicz, A. J. & Wang, C. (2018), 'A review of solvency ii: Has it met its objectives?', *British Actuarial Journal* **23**, e4.

Renshaw, A. E. & Haberman, S. (2003), 'Lee–carter mortality forecasting with age-specific enhancement', *Insurance: Mathematics and Economics* **33**(2), 255–272.

Renshaw, A. E. & 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. & 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.

Sagianou, A. & Hatzopoulos, P. (2022), 'Extensions on the hatzopoulos & sagianou multiple-components stochastic mortality model', *Risks* **10**(7), 131.

Triantaphyllou, E. (2000), *Multi-Criteria Decision Making Methods*, Springer US, Boston, MA, pp. 5–21.

United Nations (2004), *World Population Prospects: The 2002 Revision. Analytical report*, Economic & social affairs, UN, Dept. of Economic and Social Affairs. Population Division.

Villegas, A. M., Kaishev, V. K. & Millossovich, P. (2018), 'StMoMo: An R Package for Stochastic Mortality Modelling', *Journal of Statistical Software* **84**(3), 1–38.

Willets, R. (2004), 'The cohort effect: Insights and explanations', *British Actuarial Journal* **10**(4), 833–877.

Young, P. C., Tych, W. & Taylor, C. J. (2009), 'The Captain Toolbox for Matlab', *IFAC Proceedings Volumes* **42**(10), 758 – 763. 15th IFAC Symposium on System Identification. Available in http://www.es.lancs.ac.uk/cres/captain/.

Yue, J., Yang, S. & Huang, H. (2008), A study of the Lee-Carter model with age-shifts, *in* 'Living to 100 and Beyond Symposium', Orlando, Fla.