University of Piraeus

School of Finance and Statistics
Department of Statistics and Insurance Science

Actuarial Models in Demography

by

Apostolos E. Bozikas

A Thesis Submitted in Fulfilment of the
Requirements for the Degree of
Doctor of Philosophy

Piraeus, Greece
May 2019

To my beloved Elizabeth
(1912-2016)

Acknowledgements

After four years, this academic journey comes to an end. First, I would like to express my deep gratitude to my MSc and PhD supervisor professor Georgios Pitselis, for giving me the opportunity to extend my studies and providing me knowledge, help, motivation and inspiration for research. I am extremely grateful for what he has offered me all these years.

I would like to thank the members of my PhD consulting committee professors Cleon Tsimbos and Georgia Verropoulou for their encouragement and positiveness during my studies. I am also grateful to the members of my PhD examining committee professors Konstantinos Politis, Petros Hatzopoulos, Alex Karagrigoriou and Anastasia Kostaki for the careful reading and the comments that improved my thesis.

Furthermore, I want to thank my friends for supporting me over this period. Special thanks go to my friends, professor Andreas Koutsogiannis, Dr. Fotis Ioakeimidis and Dr. Spyridon Tzaninis for our discussions and their encouragement during my research.

I am extremely grateful to my parents for their love, patience and unconditional support over all these years of studies. They deserve a big thank and so do all the people who supported me during this period. A special dedication goes to Nadia El-Basha for all her love and support from the beginning of my PhD studies.

I would like to express my gratitude to my bro Georgios Tsetsenekos and my godmother Sotiria Tsetsenekou for their support in my life. I am also thankful to my classmate Georgios Laskaridis for urging me to attend extra probability and statistics courses during our BSc studies and my friends Christos Bozantzidis, Michalis Ioakeimidis, Soraya El-Basha and Yiannis Spanos. I will never forget the time spent in the inspiring Dtale café and the London coffee house that offered a place for my head.

Finally, I acknowledge the financial support from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship grant (GA. no. 1286).

H.F.R.I. Hellenic Foundation for Research & Innovation

GSRT GENERAL SECRETARIAT FOR RESEARCH AND TECHNOLOGY

Abstract

During the last decades, a significant increase in life expectancy has been observed in most countries around the world. This change is mainly due to the improvement of living conditions and the development of medical science. Consequently, a serious demographic problem arises from the increasing number of elderly, combined with low fertility rates. Population ageing creates an additional cost for life insurers and annuity providers. In this spirit, the development of efficient methods to model and forecast the mortality rates of a population is a key challenge for actuaries and demographers. This thesis exploits actuarial credibility techniques to propose novel mortality modelling methods, aiming to contribute in more accurate demographic projections. Before introducing these methods, we firstly examine and review the existing modelling techniques. Greek population data are incorporated into the most used stochastic mortality models under a common age-period-cohort framework. The fitting performance of each model is thoroughly evaluated, while projection results for both genders are also illustrated in pricing insurance-related products. In addition, we propose a credibility regression approach with random coefficients to model and forecast the mortality dynamics for populations with limited data. The results on Greek mortality data indicate that credibility regression contributes to more accurate forecasts, compared with those produced from the Lee and Carter (1992) and Cairns et al. (2006) models. Then, the credibility regression model is extended to a multi-level hierarchical credibility regression model for mortality data of multiple populations in a hierarchical form. The forecasting performances between the hierarchical model, the Lee-Carter model and two Lee-Carter extensions for multiple populations are compared for both genders of three northern European countries (Ireland, Norway, Finland). Empirical illustrations show that the proposed method produces more accurate forecasts. Finally, we present a credibility formulation of the Lee-Carter method particularly designed for multi-population mortality modelling. Differently from the standard Lee-Carter methodology, where the time index is assumed to follow an appropriate time series process, herein, the period dynamics of mortality are estimated under a crossed classification credibility framework. The forecasting performances between the proposed model, the Lee-Carter model and two Lee-Carter extensions for multiple populations are compared for both genders of three developed countries (United Kingdom, USA, Japan). The numerical results indicate that the proposed model contributes to more accurate forecasts.

Περίληψη

Τις τελευταίες δεκαετίες παρατηρήθηκε σημαντική αύξηση του προσδόκιμου ζωής στις περισσότερες χώρες του κόσμου. Η αλλαγή αυτή οφείλεται κυρίως στη βελτίωση των συνθηκών διαβίωσης και στην ανάπτυξη της ιατρικής επιστήμης. Κατά συνέπεια, ένα σοβαρό δημογραφικό πρόβλημα προκύπτει από τον αυξανόμενο αριθμό των ηλικιωμένων σε συνδυασμό με το χαμηλά ποσοστά γονιμότητας. Η γήρανση του πληθυσμού δημιουργεί ένα επιπλέον κόστος για τους ασφαλιστές ζωής και τους παρόχους συντάξεων. Στο πλαίσιο αυτό, η ανάπτυξη αποτελεσματικών μεθόδων για τη μοντελοποίηση και πρόβλεψη των ποσοστών θνησιμότητας ενός πληθυσμού αποτελεί βασική πρόκληση για τους αναλογιστές και τους δημογράφους. Η παρούσα διατριβή προτείνει νέες τεχνικές πρόβλεψης θνησιμότητας, χρησιμοποιώντας αναλογιστικές μεθόδους μοντελοποίησης από τη θεωρία αξιοπιστίας χαρτοφυλακίου, με σκοπό τη συμβολή τους σε πιο ακριβείς δημογραφικές προβολές. Πριν παρουσιάσουμε τις μεθόδους αυτές, αρχικά εξετάζουμε τις υπάρχουσες τεχνικές μοντελοποίησης. Τα ελληνικά δεδομένα προσαρμόζονται στα κυριότερα μοντέλα θνησιμότητας, τα οποία υπόκεινται σε ένα ευρύτερο πλαίσιο μελέτης μοντέλων ηλικίας-περιόδου-γενεάς. Η καταλληλότητα προσαρμογής του κάθε μοντέλου αξιολογήθηκε διεξοδικά, ενώ τα αποτελέσματα των προβλέψεων για τα δυο φύλλα αποτυπώνονται και ως προς την τιμολόγηση ασφαλιστικών προϊόντων. Στη συνέχεια, προτείνουμε ένα μοντέλο παλινδρόμησης αξιοπιστίας με τυχαίους συντελεστές για τη μοντελοποίηση και την πρόβλεψη θνησιμότητας πληθυσμών με περιορισμένο αριθμό δεδομένων. Τα αποτελέσματα του μοντέλου πάνω στα ελληνικά δεδομένα έδειξαν ότι η χρήση μεθόδων αξιοπιστίας συμβάλλει σε ακριβέστερες προβλέψεις, σε σύγκριση με εκείνες που προκύπτουν από τα μοντέλα Lee and Carter (1992) και Cairns et al. (2006). Στη συνέχεια, το μοντέλο παλινδρόμησης αξιοπιστίας χαρτοφυλακίου επεκτείνεται σε ένα ιεραρχικό μοντέλο παλινδρόμησης χαρτοφυλακίου για τη μοντελοποίηση δεδομένων θνησιμότητας πολλών πληθυσμών σε ιεραρχική μορφή. Οι προβλέψεις μεταξύ του ιεραρχικού μοντέλου, του μοντέλου Lee-Carter και δύο επεκτάσεων του μοντέλου Lee-Carter για πολλαπλούς πληθυσμούς συγκρίνονται ανά φύλο, για τρεις χώρες της Βόρειας Ευρώπης (Ιρλανδία, Νορβηγία και Φινλανδία). Τα αποτελέσματα των προβλέψεων δείχνουν ότι η προτεινόμενη μέθοδος οδηγεί σε πιο ακριβείς προβλέψεις. Τέλος, παρουσιάζουμε μια μέθοδο τροποποίησης του μοντέλου Lee-Carter, μέσω της θεωρίας αξιοπιστίας χαρτοφυλακίου, ειδικά σχεδιασμένη για τη μοντελοποίηση της θνησιμότητας μεταξύ πολλαπλών πληθυσμών. Διαφορετικά από τη βασική μεθοδολογία Lee-Carter, όπου η παράμετρος περιόδου ακολουθεί ένα κατάλληλο μοντέλο χρονολογικών σειρών, στην προτεινόμενη μέθοδο, η θνησιμότητα εκτιμάται βάσει της μεθόδου αξιοπιστίας σταυρωτής ταξινόμησης. Οι προβλέψεις μεταξύ του προτεινόμενου μοντέλου, του μοντέλου Lee-Carter και δύο επεκτάσεων του μοντέλου Lee-Carter για πολλαπλούς πληθυσμούς συγκρίνονται ανά φύλο, για τα δεδομένα τριών ανεπτυγμένων χωρών (Ηνωμένο Βασίλειο, ΗΠΑ και Ιαπωνία). Τα αποτελέσματα των προβλέψεων δείχνουν ότι το προτεινόμενο μοντέλο συμβάλλει σε πιο ακριβείς προβλέψεις.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

During the last decades, mortality has significantly declined in most developed countries around the world, mainly due to the continuous improvement of living conditions and the evolution of medical science and technology. Eventually, the decline in mortality creates higher financial responsibilities for governments and annuity providers. Consequently, finding ways to manage the mortality dynamics of a population is a very important step in building a sustainable health and pension system. In this spirit, actuaries and demographers are focused on the development of novel methods to model and forecast the mortality rates of a population.

1.2 Single Population Mortality Models

In the literature, several methods have been proposed in order to capture the mortality trends of a population. Lee and Carter (1992) proposed a pioneer modelling method to forecast the mortality of the total population of the United States, by decomposing the mortality rates into age and period parameters. A remarkable variant of the Lee–Carter method, particularly designed for higher ages, was proposed by Cairns et al. (2006). In the literature, we can find many extensions to these methods. Renshaw and Haberman (2006) extended the Lee–Carter model by including a cohort effect, while Plat (2009) proposed a model which combines preferable characteristics of the Lee and Carter (1992) and Cairns et al. (2006) models. Despite its variants and extensions, the Lee-Carter model inspired many authors to introduce more sophisticated methods by including additional parameters. Hyndman and Ullah (2007) used functional data analysis and penalized regression splines in their modelling framework and Hatzopoulos and Haberman (2009) proposed a mortality modelling approach under the framework of generalized linear models (GLM), while Hatzopoulos and Haberman (2011) extended this approach by incorporating cohort effects.

Last years, many studies have been conducted to compare mortality models on datasets of various countries. Booth et al. (2006) compared the accuracy of the forecasts obtained by five extensions of the Lee-Carter method using data from ten developed countries, while Shang et al. (2011) extended this accuracy comparison by using ten methods and incorporating data from fourteen selected countries. Cairns et al. (2009) and Haberman and Renshaw (2011) compared the fitting and forecasting performance of different stochastic models for England & Wales and the United States mortality experience. Gaille (2012) applied the Lee-Carter and the Heligman-Pollard models to Swiss mortality rates and compared the financial impacts of their forecasts on future pension liabilities. Stoeldraijer et al. (2013) compared the forecasts obtained from the Lee-Carter method and its extensions with the official forecasts obtained from the statistical offices in Europe to evaluate the differences for the case of Netherlands. Hatzopoulos and Haberman (2015) proposed a dynamic parametric model, within the GLM framework, for analyzing the cohort mortality survival function for Sweden, Norway, England & Wales and Denmark. In addition, Van Berkum et al. (2016) analyzed the impact of allowing for multiple structural changes on a large collection of mortality models fitted on Dutch and Belgian male data and Maccheroni and Nocito (2017) backtested the forecasting performance of the Lee and Carter (1992) and the Cairns et al. (2006) models on Italian data.

A common issue in mortality modelling is that, for some countries, there are not enough historical data. This issue affects the existing modelling methods, which inevitably base their forecasts on datasets of a limited period of observations. Li et al. (2004) extended the Lee–Carter model to be applied for Chinese and South Korean mortality data, which are available at only a few points in time and at unevenly spaced intervals. Zhao (2012) modified the Lee–Carter model by incorporating linearized cubic splines and other additive functions to approximate the model parameters and forecast mortality for short-base-period Chinese data. Huang and Browne (2017) presented a stochastic modification of the CMI (Continuous Mortality Investigation) model to project mortality improvement rates for limited Chinese data using clustering analysis techniques and Wang et al. (2018) proposed a mortality modelling approach for small populations with a combination of data aggregation and graduation. Also, Kostaki and Zafeiris (2019) reviewed the typical problems and limitations of empirical mortality data in small populations and proposed some ways to deal with them.

Differently from the above Lee–Carter variants and extensions, alternative modelling approaches can also serve as a tool in mortality modelling for populations with limited data. Based on the actuarial credibility theory, these approaches aim to model the period patterns of limited mortality data for a specific age, using information from a wider age span. Bühlmann (1967) established the theoretical foundation of modern credibility theory (also known as greatest accuracy credibility theory) and Hachemeister (1975) introduced a credibility regression model to estimate auto-mobile bodily injury claims for various states in the USA. For an extensive review on credibility theory for non-life insurance, we also refer to Goovaerts et al. (1990), Bühlmann and Gisler (2005) and Klugman et al. (2012).

Regarding some life insurance applications of credibility theory, Hardy and Panjer (1998) used empirical Bayes credibility theory to provide a theoretical basis for the calculation of risk measures associated with mortality risk for insurance companies. Salhi et al. (2016) proposed a credibility approach, which consists on reviewing the fitting parameters of a Makeham mortality curve, as new observations arrive. Schinzinger et al. (2016) presented a multivariate evolutionary credibility model for mortality improvement rates to describe the joint dynamics of mortality through time in several populations. Tsai and Lin (2017a) applied Bühlmann credibility to mortality data of Japan, the United Kingdom and the United States, while Tsai and Lin (2017b) incorporated Bühlmann credibility into the Lee and Carter (1992) model, the Cairns et al. (2006) model and the linear relational model of Tsai and Yang (2015) to improve forecasting performance for the United Kingdom dataset. Moreover, Li and Lu (2018) proposed a Bayesian non-parametric model for the mortality of a small population, when a benchmark mortality table of a larger population is also available and serves as part of the prior information. By using an adaptive smoothing procedure based on the local likelihood, Salhi and Thérond (2018) proposed a methodology to adjust the graduated mortality table based on credibility techniques and Gong et al. (2018) highlighted the importance of using credibility procedures in individual life and annuity business.

1.3 Multi-Population Mortality Models

Wilson (2001) observed a global convergence in mortality, setting a basis for the development of multi-population mortality models. Models like the aforementioned ones ignore the dependence structure across populations and may lead to isolated and divergent forecasts among populations. On the other hand, considering the mortality dependency across populations could eliminate this divergent behaviour and potentially improve the forecasting performance by incorporating more data. The most widely used extensions of the Lee-Carter model for multiple populations are the Carter and Lee (1992) approach, which applies a common time-varying index to all populations and the Li and Lee (2005) approach, which proposes a two-step procedure to model mortality dynamics for multiple populations.

In the literature, we can find many contributions regarding multi-population mortality modelling. Li and Hardy (2011) assumed that there is a linear relationship between the time varying index of a base population and the other populations and Cairns et al. (2011) introduced a Bayesian framework to jointly model two populations. Hatzopoulos and Haberman (2013) presented a coherent mortality modelling structure under the

GLM framework for analyzing mortality dynamics using worldwide data from the Human Mortality Database. D'Amato et al. (2014) extended the Lee-Carter model in order to take into account the existence of dependence in mortality data across multiple populations and Kleinow (2015) developed a common age effect model for multiple populations. Li et al. (2015) generalized a single-population mortality model in different possible ways to fit two or more populations and measured the basis risk in longevity hedges. Wan and Bertschi (2015) proposed a coherent model for Swiss data. Tsai and Wu (2018) incorporated the hierarchical credibility of Jewell (1975) to model the mortality rates for multiple populations and Tsai and Zhang (2019) proposed a multi-dimensional Bühlmann credibility approach to model mortality rates for multiple populations.

1.4 Objectives and Thesis Outline

This thesis first investigates the fitting and forecasting performance of the most used single and multi-population mortality models in the literature and then exploits actuarial credibility modelling techniques in order to build novel mortality models, which could potentially contribute to more accurate demographic projections. The rest of this thesis is organized as follows:

Chapter 2 comparatively applies the most widely used mortality models in the literature, for the first time, to Greek data. An analysis of their fitting behaviour was conducted and the corresponding forecasting results were evaluated. More specifically, we incorporated the Greek mortality data into seven mortality models under a common age-period-cohort framework. The fitting performance of each model was thoroughly evaluated based on information criteria values, as well as the likelihood ratio test, and their robustness to period changes was investigated. In addition, parameter risk in forecasts was assessed by employing bootstrapping techniques. For completeness, projection results for both genders were also illustrated in pricing insurance-related products.

Chapter 3 proposes a credibility regression approach with random coefficients to model and forecast the mortality dynamics of a given population with limited data. Age-specific mortality rates are modelled and extrapolation methods are utilized to estimate future mortality rates. The results on Greek mortality data indicate that credibility regression contributed to more accurate forecasts than those produced from the Lee and Carter (1992) and Cairns et al. (2006). An application on pricing insurance-related products is also provided.

Chapter 4 proposes a multi-level hierarchical credibility regression method to model multi-population mortality data in a hierarchical form. Future mortality rates are derived using different extrapolation techniques, while the forecasting performances between the proposed model, the classical Lee-Carter model and two Lee-Carter extensions for multiple populations are compared for both genders of three northern European countries (Ireland, Norway, Finland). Empirical illustrations show that the proposed method produces more accurate forecasts, based on the mean absolute percentage forecast error (MAPFE) values.

Chapter 5 presents a credibility formulation of the Lee-Carter method particularly designed for multi-population mortality modelling. Differently from the standard Lee-Carter methodology, where the Lee-Carter time index is assumed to follow an appropriate time series process, herein, the period dynamics of mortality are estimated under a crossed classification credibility framework. The forecasting performances between the proposed model, the classical Lee-Carter model and two Lee-Carter extensions for multiple populations are compared for both genders of three developed countries (United Kingdom, USA, Japan). The numerical results indicate that the proposed model contributes to more accurate forecasts, based on the mean absolute percentage forecast error (MAPFE) values. Finally, Chapter 6 presents the general conclusions of this thesis.

Chapter 2

Stochastic Mortality Modelling under the Age-Period-Cohort Framework

2.1 Introduction

During the last decades, a significant increase in life expectancy has been observed worldwide. This change is mainly due to the human race dynamics, the improvement of living conditions and the development of medical science. Due to these factors, life expectancy in Greece has been increased from 70.2 to 78 years for males and 73.8 to 83.3 years for females during the period from 1961 to 2010, almost 9 years on average for both genders in 50 years (http://ec.europa.eu/eurostat/data/database).

From a human point of view, this increase in life expectancy constitutes positive news. However, for governments and annuity providers this is not necessarily the case, because higher life expectancy increases future pension costs, as benefits have to be provided over a longer period.

Especially, for the case of Greece, Tsimbos et al. (2011) presented estimates of life expectancy at birth for males and females based on regional life tables, constructed for the 51 administrative national departments for years 1991, 2001 and 2007. Kalogirou et al. (2012) estimated appropriate mortality measures for the three main categories of causes of death, for the 51 national prefectures and Verropoulou and Tsimbos (2016) examined, for the first time in Greece, mortality by cause of death among immigrants. Hatzopoulos and Haberman (2009) applied a parametrized approach, under the GLM framework, to forecast mortality using Greek data for years 1957-2006 and Zafeiris and Kostaki (2017) examined the mortality characteristics of the Greek population for years 1961 to 2014.

In general, actuaries and demographers are focused on the development of methods that could estimate future mortality trends of a population. In this direction Hunt and Blake (2015) introduced an age-period-cohort (APC) classification scheme for the existing mortality models that was then deployed by Villegas et al. (2017). Our study builds upon these works to investigate how the APC framework can be implemented with Greek data. A comparative analysis of the fitting methods is performed and the corresponding forecasting results for the Greek population are illustrated. In addition, forecasts are applied to price net premiums of insurance-related products.

The rest of this chapter is organized as follows. Section 2.2 illustrates an overview of the stochastic mortality models that Greek data fit. Section 2.3 describes fitting procedures, while Section 2.4 illustrates the mortality projection results for each model, along with an application in pricing insurance-related products. Our findings in comparison with those from the original papers are presented in Section 2.5. Concluding remarks are given in Section 2.6.

2.2 Mortality Modelling

In this section, we review the most widely used mortality models in the literature that belong to a common APC framework. According to Booth and Tickle (2008), mortality forecasting methods have been mainly developed under three notions, the “expectation”, the “explanation” and the “extrapolation”, each one of them having its positive and negative points.

In expectation methods, mortality forecasting is based on an expert’s opinion, which incorporates specific demographic or other relevant knowledge, but sometimes can lead to subjectivity or bias errors. Explanatory methods are based on structural or epidemiological models of certain causes of death involving known risk factors and they are generally limited to short-term forecasting. Extrapolative is the most promising and modern research method as it assumes that past mortality trends will continue in the future. Hence, all the models that will be discussed in the following sections incorporate the extrapolative method and they take the advantage of using time series models that give a probabilistic confidence interval for the forecasts.

Recent research activity aims to investigate the similarities among stochastic models in order to highlight their common properties. Aro and Pennanen (2011) fitted a general modelling framework into Finnish data that allows for multiple risk factors and guarantees that the parameter estimates are well-defined. Later, Hunt and Blake (2015) proposed a general APC modelling structure that encloses most of the existing mortality models. In the following, this APC framework of stochastic mortality modelling is described and then, it is illustrated using Greek data.

2.2.1 The Age-Period-Cohort Framework

Let us denote the observed number of deaths at at age x and year t as $d_{x,t}$ and central (at the middle of year t) population exposures as $E_{x,t}$. Initial exposures are then approximated by $E^0_{x,t} \approx E_{x,t} + (1/2)d_{x,t}$. Therefore, the one-year probability of death at age x

and year t is defined by $q_{x,t} = d_{x,t}/E^0_{x,t}$ and the death rate by $m_{x,t} = d_{x,t}/E_{x,t}$. According to Cairns et al. (2009), under the assumption that force of mortality remains constant over each year of integer age and over each calendar year, death rate $m_{x,t}$ and force of mortality $\mu_{x,t}$[1] coincide. Above conventions are adopted in this study.

A stochastic APC model links a response variable (usually the one-year probability of death $q_{x,t}$ or the force of mortality $\mu_{x,t}$) to an appropriate predictor, dependent on age $x = x_1, \ldots, x_k$, period $t = t_1, \ldots, x_n$, and cohort (year of birth) $c = t_1 - x_k, \ldots, t_n - x_1$ for a population. This structure is given by the following formula

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x}, \tag{2.1}$$

where $\eta_{x,t}$ denotes the link function, which transforms a mortality rate measure into a suitable modelling form, α_x is the static age function that expresses the general shape of mortality by age, $\beta_x^{(i)} \kappa_t^{(i)}$ is a set of N age-period terms, determining the mortality trends, where $\kappa_t^{(i)}$ indicates the general pattern of mortality through the time, while $\beta_x^{(i)}$ shows this pattern of mortality change across ages and $\beta_x^{(0)} \gamma_{t-x}$ is the age-cohort term, where $\gamma_{t-x} \equiv \gamma_c$ captures the effects of each year of birth c and $\beta_x^{(0)}$ modifies this effect across ages.

The choice of the response variable that is transformed by the link function $\eta_{x,t}$ depends on the format of mortality data. For instance, if the random variable of the number of deaths at age x and year t, $D_{x,t} \sim Binomial(E^0_{x,t}, q_{x,t})$ with $E(D_{x,t}/E^0_{x,t}) = q_{x,t}$, then initial exposures $E^0_{x,t}$ should be used. If random variable $D_{x,t} \sim Poisson(E_{x,t}\mu_{x,t})$ with $E(D_{x,t}/E_{x,t}) = \mu_{x,t}$, the central exposures $E_{x,t}$ are used. Hence, under the *Binomial* distribution assumption, the logit expression for the probability of death is used and link function takes the form $\eta_{x,t} = \text{logit} q_{x,t} = \log \frac{q_{x,t}}{1-q_{x,t}}$, while if a *Poisson* distribution of deaths is assumed, then $\eta_{x,t} = \log \mu_{x,t}$. For details, see Hunt and Blake (2015) and Villegas et al. (2017). We note that presence of the bilinear terms $\beta_x \kappa_t$ classifies the APC modelling structure into the generalised non linear family of models, discussed by Currie (2016). It has to be mentioned that models with smoothing functions will be not considered in this study[2].

Finally, we have to point out that in a mortality study, specific structural characteristics of the dataset should affect model choice. For instance, if there is evidence for cohort effects in our data, then a model with a cohort parameter should be selected. Moreover, if we believe that there is randomness in mortality rates from one year to the next, then our choice lies between models that incorporate more than one period factors.

[1] According to Cairns et al. (2009), the force of mortality can be viewed as the instantaneous death rate at exact time t for a person aged exactly x at time t.

[2] For instance, Hyndman and Ullah (2007) used functional data analysis and penalized regression splines in their modelling framework.

2.2.2 Data and Assumptions

The observed number of deaths $d_{x,t}$ and the central exposures $E_{x,t}$ for the Greek population were directly obtained from the Human Mortality Database (HMD, 2017). In HMD, Greek data are available by gender and age for the observation period of 1981 to 2013. Moreover, as suggested by Haberman and Renshaw (2011), for consistency in model comparison, all models should be fitted using the same distributional assumptions and results should be shown using the same mortality measure. Therefore, we assume a *Binomial* distribution of deaths using link function $\eta_{x,t} = \text{logit} q_{x,t}$.

For our study, only the ages from $x_1 = 60$ to $x_k = 89$ will be considered, as most of the models that will be discussed in next sections have been particularly designed for higher ages. Also, in order to obtain more reliable fitting and forecasting conclusions, only data of the historical period from $t_1 = 1981$ to $t_n = 2010$ were exploited, leaving last three years[3] out for backtesting reasons.

Furthermore, Cairns et al. (2009) point out that the reliability of the estimated cohort parameters γ_{t-x} depends on the number of the observations for each birth year. Our analysis was repeated by excluding cohorts with less than three to ten observations. Especially for datasets with short periods of time, excluding more than five cohorts seems to be excessive. Nevertheless, excluding male and female cohorts with less than eight observations (1892–1899 and 1943–1950) provides a better balance between the fitting and forecasting behaviour of Greek data. This choice gave us almost the same fitting results in comparison with the fact of excluding less cohorts, but led us to more reasonable forecasts, possibly due to avoiding overfitting of the cohort effect.

2.2.3 Age-Period-Cohort Mortality Models: A Review of Methods

In this subsection, we review the seven most widely used stochastic mortality models, labelled $M_i, i = 1, \ldots, 7$. The models of this section can be classified in the APC framework (2.1), assuming a *Binomial* distribution of deaths with $\eta_{x,t} = \text{logit}\, q_{x,t}$. The seven models M_1–M_7 are listed in Table 4.1.

M_1: The Lee-Carter Model

One of the most popular and widely applied models was proposed by Lee and Carter (1992) to forecast the mortality rates of the United States. In its original version, the model uses principal component analysis in order to decompose the bilinear age-period matrix of log death rates into a single age parameter and a time index used in forecasting. Many variants and extensions of this model followed. For some works related to the Lee-Carter method and its modifications, we refer to Lee and Miller (2001), Booth et al. (2002) and De Jong and Tickle (2006). In Booth et al. (2002), the Lee-Carter method

[3] Due to the limited availability of Greek data in HMD, years 2011–2013 correspond to a percentage of 10% of the whole fitting year span.

was embedded in a Poisson regression setting to model the Belgian death rates. The Lee-Carter model predictor is given by $\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}$, imposing the following constraints $\sum_x \beta_x^{(1)} = 1$ and $\sum_t \kappa_t^{(1)} = 0$ to ensure identifiability of the model predictor.

M$_2$: The Renshaw-Haberman Model

Renshaw and Haberman (2006) extended the Lee-Carter model by including a cohort parameter to $\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(0)} \gamma_{t-x}$. Haberman and Renshaw (2011) investigated certain complications of this model associated with predictions efficiency and the capture of the cohort effect for both United States and England & Wales mortality experience. These issues were resolved by using a simpler model predictor given by $\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \gamma_{t-x}$, assuming independence between the period and cohort parameters. This model will be considered in our application, incorporating the following identifiability constraints $\sum_x \beta_x^{(1)} = 1,\ \sum_t \kappa_t^{(1)} = 0$ and $\sum_c \gamma_c = 0$, where c is the set of cohort years of birth that have been fitted in the model.

M$_3$: The "Age-Period-Cohort" Currie Model

Currie (2006) presented in the actuarial literature a demographic model structure, firstly discussed by Hobcraft et al. (1982). Its predictor is given by $\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$. We can easily observe that this simple APC structure is a simplification of the previous model, considering that $\beta_x^{(1)} = 1$. Hence, period and cohort parameter estimates can be obtained as in M_2 under the identifiability constraints $\sum_t \kappa_t^{(1)} = 0,\ \sum_c \gamma_c = 0$ and $\sum_c c\gamma_c = 0$.

M$_4$: The Plat Model

After combining characteristics of other models, Plat (2009) proposed a three period factor model $\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x-\bar{x})^{+}\kappa_t^{(3)} + \gamma_{t-x}$, where $\bar{x}$ is the average age in the data. Then, he compared the fitting quality with datasets from the United States, England & Wales and Netherlands and noticed that when the age range is limited to higher ages (60 years or older), the reduced expression of his model predictor $\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$ should be ideally used. Thus, latter model structure is adopted for our application, using $\sum_t \kappa_t^{(1)} = 0,\ \sum_t \kappa_t^{(2)} = 0,\ \sum_c \gamma_c = 0,\ \sum_c c\gamma_c = 0$ and $\sum_c c^2\gamma_c = 0$ constraints to eliminate identifiability issues.

M$_5$: The Cairns-Blake-Dowd Model

In order to reduce the number of free parameters, Cairns et al. (2006) proposed a parsimonious model for the data from England & Wales, incorporating only two period factors in the absence of a static age function and cohort terms. This model predictor is given by $\eta_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}$. This structure has no identifiability issues, hence no constraints were taken into consideration.

M_6: The "Cohort" Cairns et al. Model

Cairns et al. (2009) extended M_5 to include a cohort effect as $\eta_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$. Note that this model structure is a reduced version of Plat's structure without a static age term, under the following constraints $\sum_c \gamma_c = 0$ and $\sum_c c\gamma_c = 0$.

M_7: The "Quadratic" Cairns et al. Model

A more complicated structure for M_6 was introduced by Cairns et al. (2009), which includes an additional quadratic age effect with a period term. The model predictor is given by $\eta_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \left((x-\bar{x})^2 - \hat{\sigma}_x^2\right)\kappa_t^{(3)} + \gamma_{t-x}$, where the constant $\hat{\sigma}_x^2$ is the mean of $(x-\bar{x})^2$. This model is identifiable under the transformations $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$ and $\sum_c c^2\gamma_c = 0$.

Finally, another extension of M_6 with a decreasing cohort effect was also discussed in Cairns et al. (2009), given by $\eta_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + (x_d - x)\gamma_{t-x}$, where x_d is a constant parameter and its predictor is identifiable under $\sum_c \gamma_c = 0$. Unfortunately, this model revealed some dangers associated with its use, according to Cairns et al. (2011), where it led to very implausible results in forecasting the United States male mortality. Hence, due to the above issues, this model structure will not be considered in this study[4].

Table 2.1 Structure overview of $M_1 - M_7$ mortality models.

Model	Structure	Original Papers
M_1	$\text{logit}\, q_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)}$	Lee and Carter (1992)
M_2	$\text{logit}\, q_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \gamma_{t-x}$	Renshaw and Haberman (2006)
M_3	$\text{logit}\, q_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_{t-x}$	Currie (2006)
M_4	$\text{logit}\, q_{x,t} = \alpha_x + \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$	Plat (2009)
M_5	$\text{logit}\, q_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)}$	Cairns et al. (2006)
M_6	$\text{logit}\, q_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + \gamma_{t-x}$	Cairns et al. (2009)
M_7	$\text{logit}\, q_{x,t} = \kappa_t^{(1)} + (x-\bar{x})\kappa_t^{(2)} + ((x-\bar{x})^2 - \hat{\sigma}_x^2)\kappa_t^{(3)} + \gamma_{t-x}$	Cairns et al. (2009)

[4]As Hunt and Blake (2015) point out, in practice, M_7 has been proved the most popular extension of the original Cairns et al. (2006) model, since it gives a better fit than M_6, while the age function for the cohort parameters in M_8 may be more complicated to fit data due to the estimation of the additional constant parameter x_d.

2.3 Model Fit

In this section, we describe the fitting methods whereby model parameters can be estimated. Lee and Carter (1992) estimated model parameters using singular value decomposition (SVD) in the context of least squares fitting method, while Renshaw and Haberman (2003) minimised the deviance of their predictor structure. Following Brouhns et al. (2002), we estimate age, period and cohort parameters by maximising model's likelihood. Under the assumption of a *Binomial* distribution, log-likelihood for models M_1 - M_7 is given in Villegas et al. (2017) as

$$L(d_{x,t}) = \sum_{x,t} \omega_{x,t} \left\{ d_{x,t} \log\left(\frac{\hat{d}_{x,t}}{E^0_{x,t}}\right) + (E^0_{x,t} - d_{x,t}) \log\left(\frac{E^0_{x,t} - \hat{d}_{x,t}}{E^0_{x,t}}\right) + \log\binom{E^0_{x,t}}{d_{x,t}} \right\}, \quad (2.2)$$

where $E^0_{x,t}$ is the initial exposure, while f^{-1} now denotes the inverse link function of $f(u) = \operatorname{logit} u$. Then, the expected number of deaths for each model is given by

$$\hat{d}_{x,t} = E^0_{x,t} f^{-1}\left(\alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x}\right), \quad (2.3)$$

with $N = 1$ for M_1 - M_3, $N = 2$ for M_4 - M_6 and $N = 3$ for M_7 and the prior weights $\omega_{x,t}$ are defined as

$$\omega_{x,t} = \begin{cases} 0, & \text{if a } (x,t) \text{ data cell is omitted,} \\ 1, & \text{if data cell is included.} \end{cases} \quad (2.4)$$

For the implementation of mortality models, there are various *R*-packages in the literature. In particular, the standard Lee-Carter model and some of its extensions are included in the “demography” (Hyndman et al., 2014) package, while Butt et al. (2014) developed the “ilc” package that contains the Renshaw-Haberman family related models. Turner and Firth (2015) provided the “gnm” package, which facilitates the fitting procedure and the parameters estimation of generalized nonlinear models and Villegas et al. (2017) introduced the powerful “StMoMo” package that incorporates all the fitting algorithms that we used in this chapter.

It is worth mentioning that when we firstly fitted $M2$ to Greek female data, $\kappa_t^{(1)}$ showed an upward trend compared to decreasing γ_{t-x} values. This is the result of the well-known identifiability issues of the Renshaw-Haberman model. To overcome this issue, we considered an additional constraint for the cohort parameter, according to Hunt and Villegas (2015).

We also note that fitting models M_1–M_4 under the *Poisson* assumption ($\eta_{x,t} = \log \mu_{x,t}$), as they were firstly adjusted in original papers, gave us similar parameter estimates. In addition, robustness of parameter estimates was examined by using two different fitting periods of data: 1981–2000 and 1981–2010. Figures 2.1–2.7 illustrate the maximum likelihood estimates under the *Binomial* assumption for models M_1–M_7 respectively, for Greek males and females, aged 60–89. Solid lines correspond to parameter estimates for the fitting period 1981–2010, while dotted lines for the period 1981–2000. In the following, we give some explanatory comments on parameter estimates.

2.3.1 Parameter Estimates

The α_x estimates (Figure 2.1) show an almost linear upward trend for both genders, which is similar for models M_2–M_4, therefore it is omitted from Figures 2.2–2.4. The estimates for $\kappa_t^{(1)}$ decrease in every mortality model, indicating a general mortality improvement for both genders over the time.

For each one of the (five) models that incorporate a cohort parameter, estimates cannot be safely interpreted as they depend on the whole model setting, including possible interactions with $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ parameters and the corresponding age effects. More precisely, cohort estimates of M_2 (Figure 2.2), M_3 (Figure 2.3), M_6 (Figure 2.6) and M_7 (Figure 2.7) show an increase until year 1915 and decreasing fluctuations for the rest of the cohort years, while M_4 cohort estimates (Figure 2.4) fluctuate over the entire period.

2.3.2 Robustness

As Cairns et al. (2009) pointed out, an important property of a model is the robustness of its parameter estimates relative to changes in the range of fitted data. That is, parameter estimates should not change significantly when fitting to a shorter data range. Consequently, a possible lack of robustness for a model means that is sensitive to changes in the period of fitted data and brings into question the appropriateness of its use for projections or other relevant applications that wholly rely on them.

Dotted lines in Figures 2.1–2.7 indicate that none of the seven models suffers from serious robustness issues. However, use of a fitting range with less data results to an abrupt increase of $\beta_x^{(1)}$ female estimates (bottom-left panel of Figure 2.2) and that remains unchanged even if we repeated model fit, considering less cohorts to be excluded. On the contrary, models M_5 (Figure 2.5) and M_6 (Figure 2.6) seem to be the most robust ones for both genders.

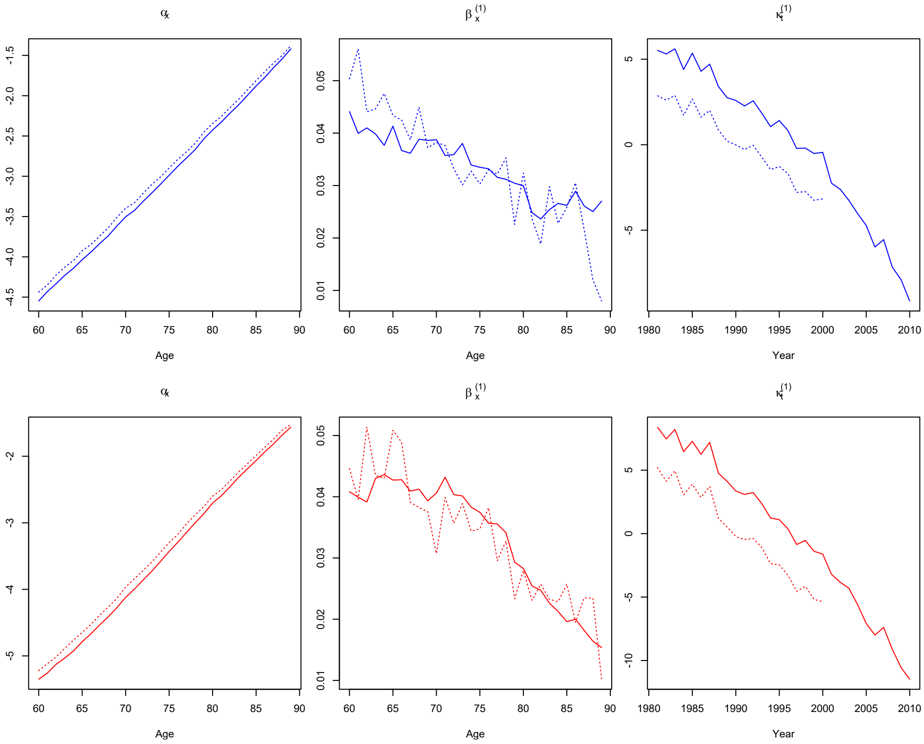


Figure 2.1 M_1: α_x, $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

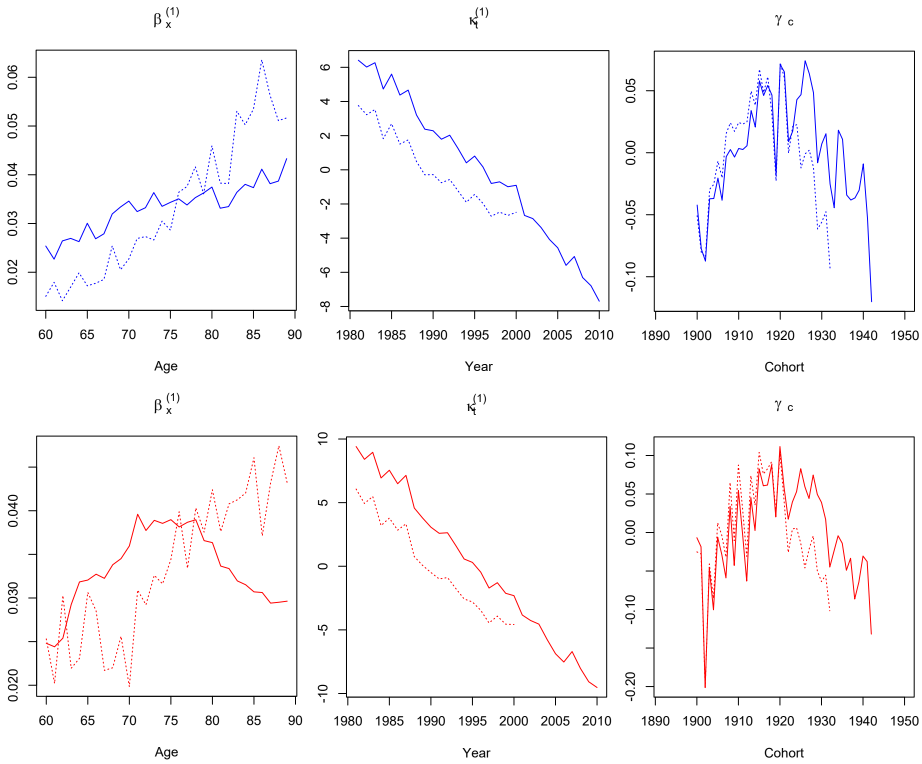


Figure 2.2 M_2: $\beta_x^{(1)}$, $\kappa_t^{(1)}$ and γ_c estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

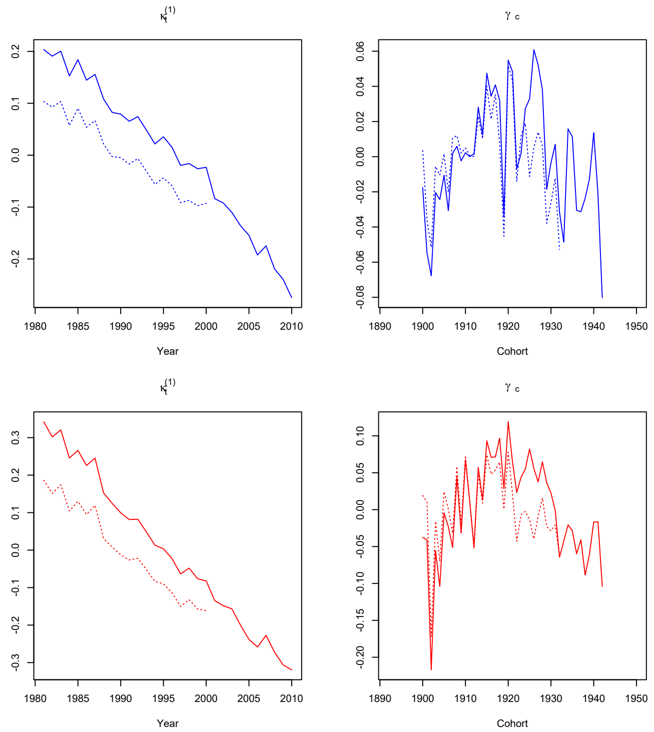


Figure 2.3 M_3: $\kappa_t^{(1)}$ and γ_c estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

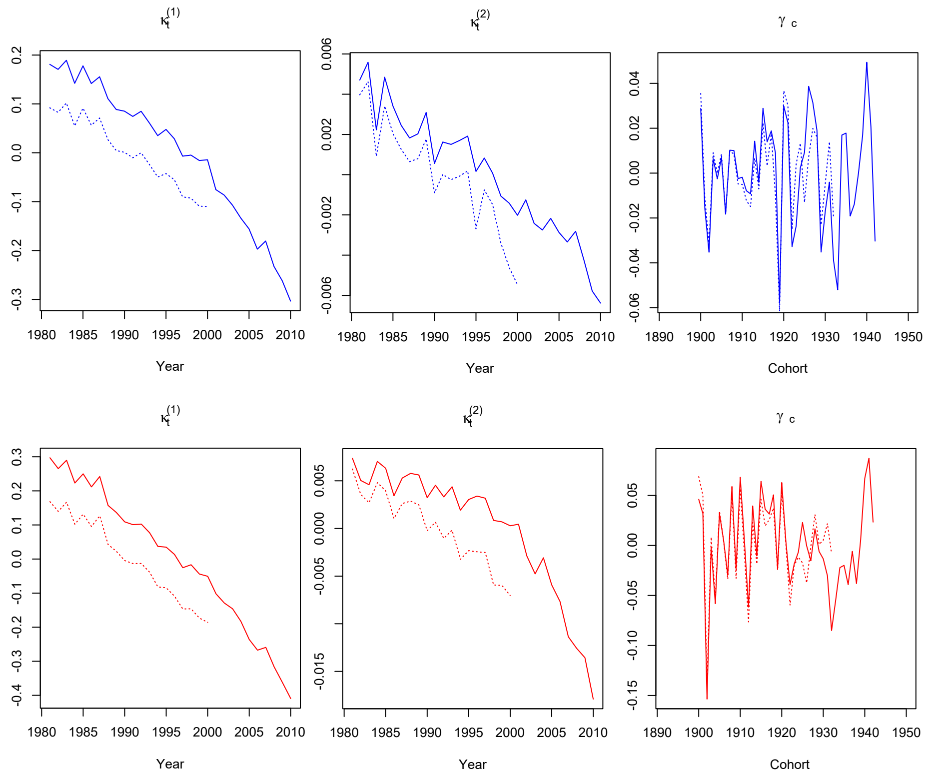


Figure 2.4 M_4: $\beta_x^{(1)}$, $\kappa_t^{(1)}$ and γ_c estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

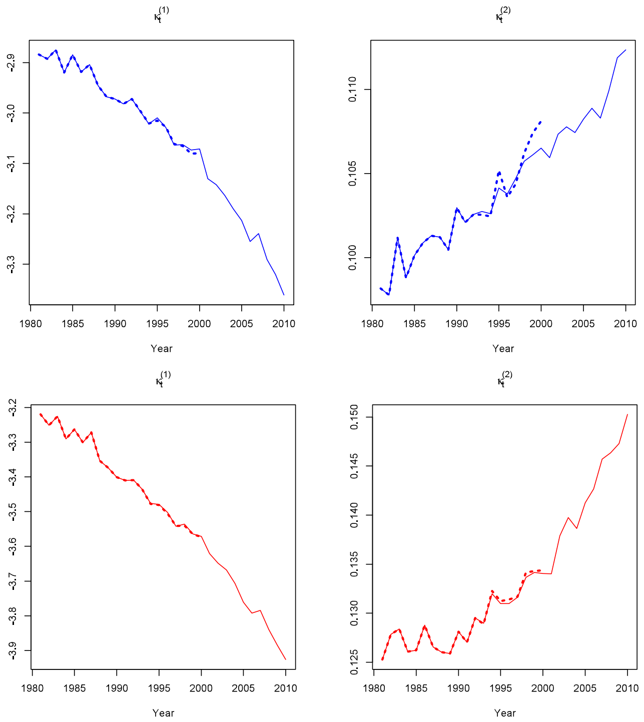


Figure 2.5 M_5: $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

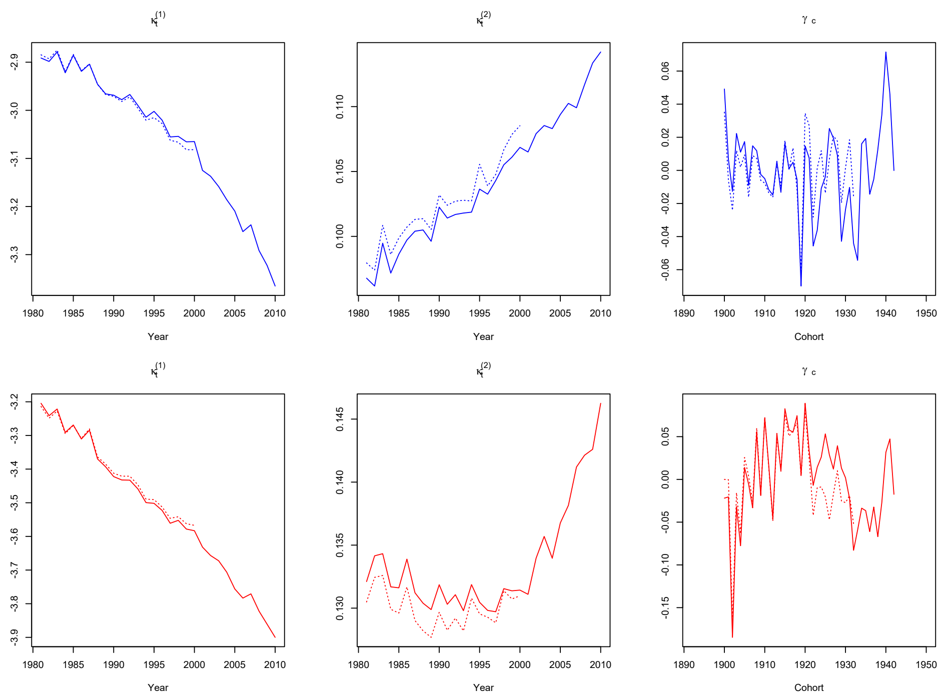


Figure 2.6 M_6: $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and γ_c estimated parameters for males (top panels) and females (bottom panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

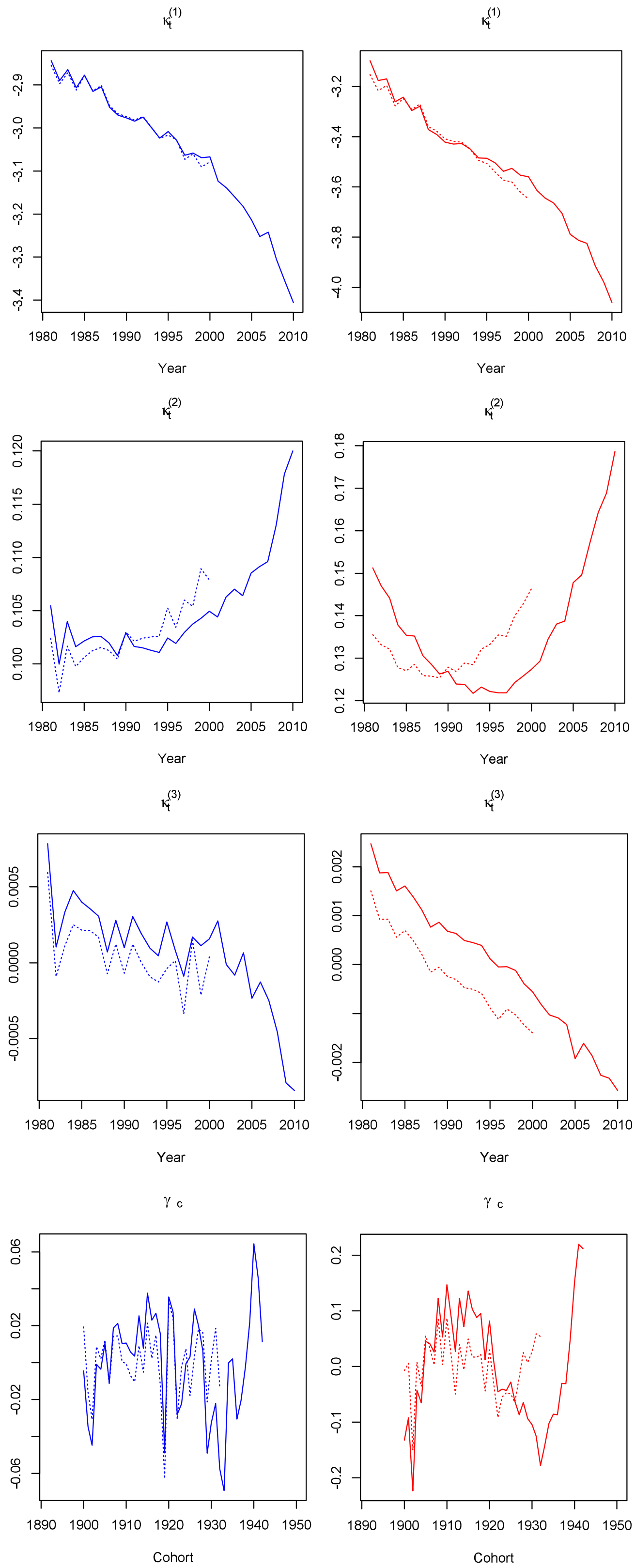


Figure 2.7 M_7: $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and γ_c estimated parameters for males (left panels) and females (right panels), aged 60–89, fitted in 1981–2010 (solid lines) and 1981–2000 (dotted lines).

2.3.3 Goodness of Fit Diagnostics

A model's goodness of fit is measured by the scaled residual deviance between the observed and the fitted data, which depends on the chosen distributional assumption. As discussed in Pitacco et al. (2009), lack of randomness in the residuals patterns indicates the inability of a model to capture specific age, period or cohort effects.

Under the *Binomial* distribution assumption of deaths, residual deviance for each model is defined by Debón et al. (2010) as

$$D(d_{x,t},\hat{d}_{x,t}) = \sum_{x,t} \mathrm{dev}(x,t) = \sum_{x,t} 2\omega_{x,t}\left\{ d_{x,t}\log\left(\frac{d_{x,t}}{\hat{d}_{x,t}}\right) + (E^0_{x,t} - d_{x,t})\log\left(\frac{E^0_{x,t}-d_{x,t}}{E^0_{x,t}-\hat{d}_{x,t}}\right)\right\}.$$

Then, standardised deviance is given by Pitacco et al. (2009) as

$$r_{x,t} = \mathrm{sign}(d_{x,t} - \hat{d}_{x,t})\left(\frac{\mathrm{dev}(x,t)}{\hat{\phi}}\right)^{1/2}. \tag{2.5}$$

The weights $\omega_{x,t}$ in (2.3.3) are defined as in (2.4) and

$$\hat{\phi} = \frac{D(d_{x,t},\hat{d}_{x,t})}{\nu},$$

where ν expresses the degrees of freedom of the model (the number of the observations minus the number of the model parameters).

Figures 2.8–2.14 plot the residuals deviance against age, period (calendar year) and cohort (year of birth) for models M_1–M_7, fitted for ages 60–89 of period 1981–2010 for males and females. According to the structural features of each model, we can make some comments.

- The evident dispersion of residuals in the right panels of Figures 2.8 and 2.12 reveal the inability of models M_1 and M_5, respectively to capture the cohort effect.
- The strong patterns, appeared in left panels of Figures 2.12 and 2.13 illustrate the weakness of models M_5 and M_6 respectively to capture the age effects, especially for females.
- All the models capture effectively the period effects.

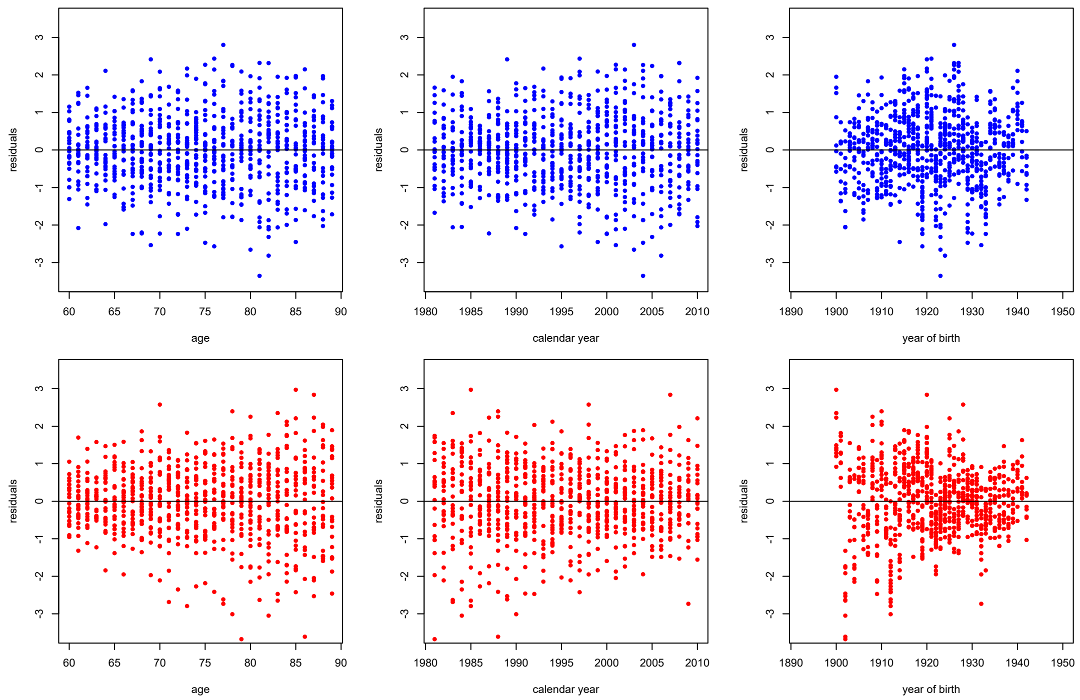


Figure 2.8 Residuals deviance of M_1 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

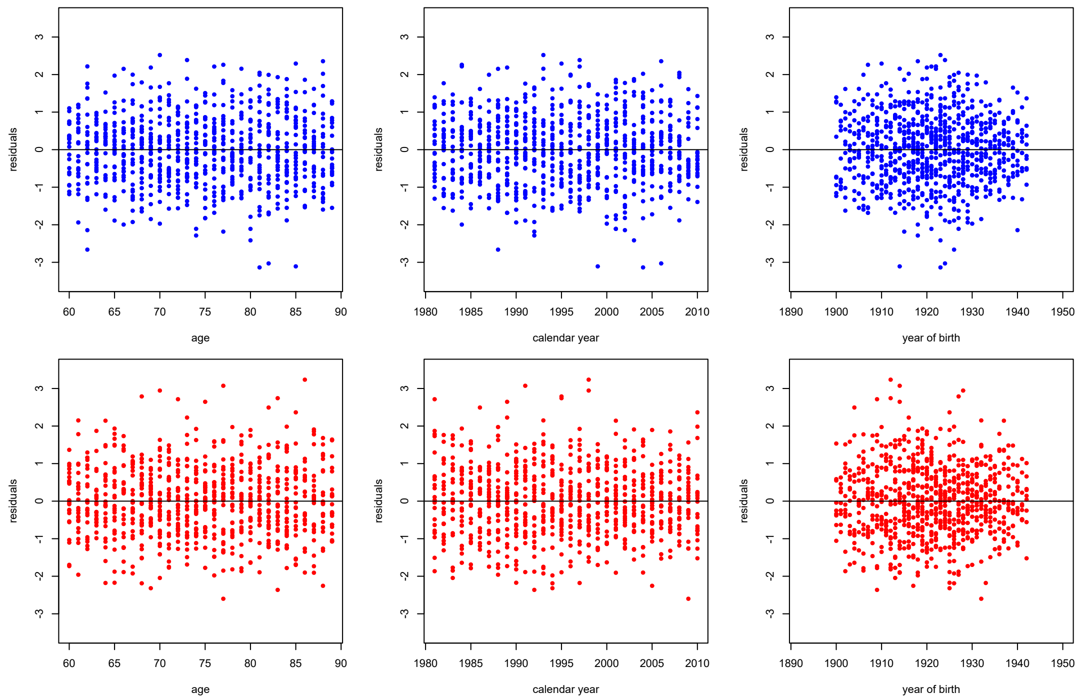


Figure 2.9 Residuals deviance of M_2 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

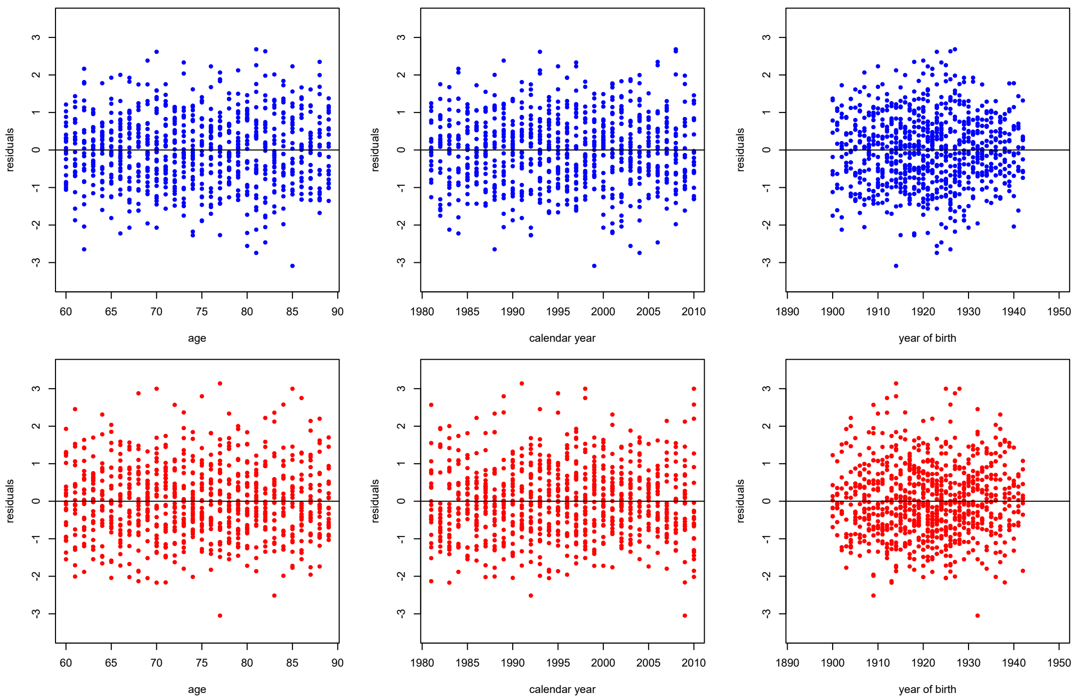


Figure 2.10 Residuals deviance of M_3 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

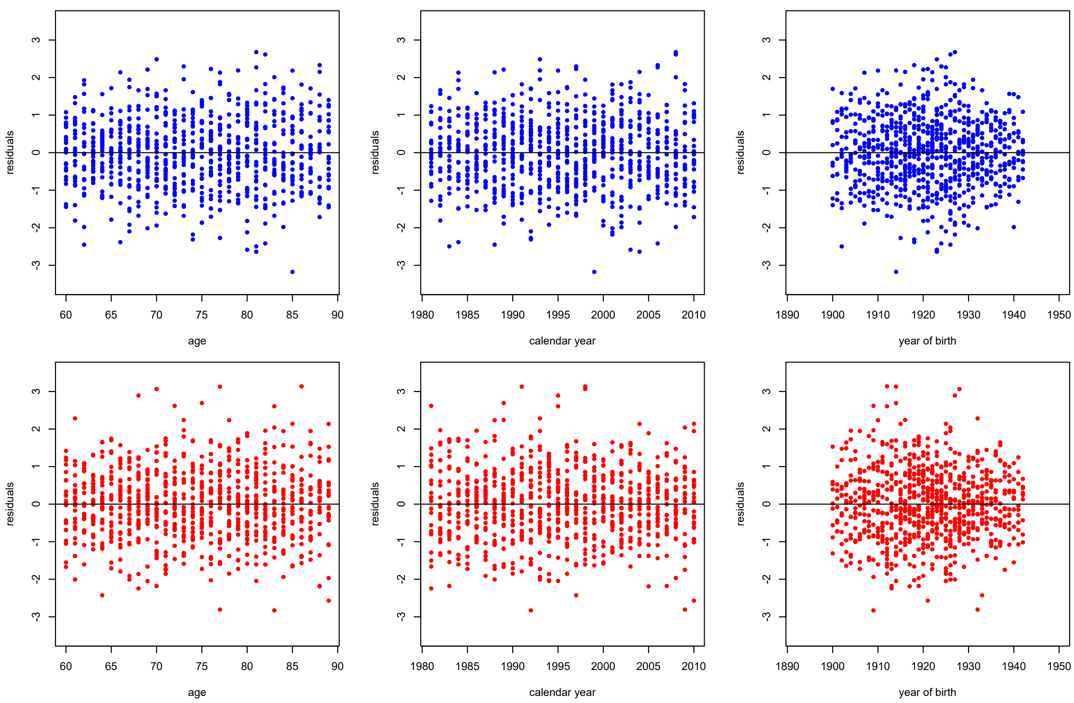


Figure 2.11 Residuals deviance of M_4 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

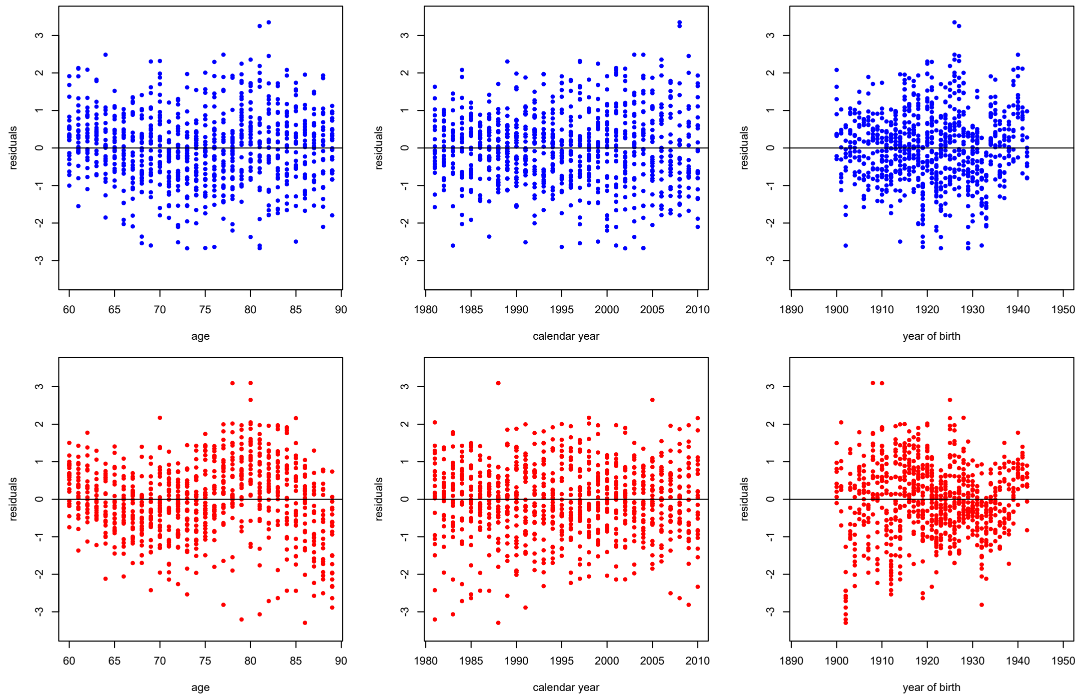


Figure 2.12 Residuals deviance of M_5 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

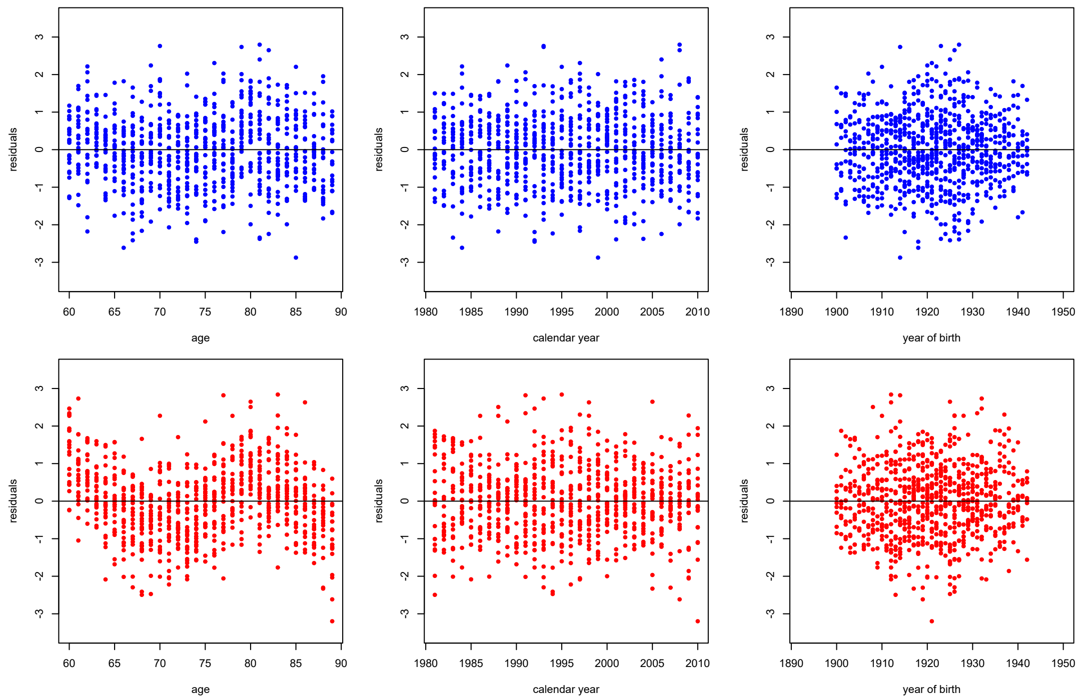


Figure 2.13 Residuals deviance of M_6 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

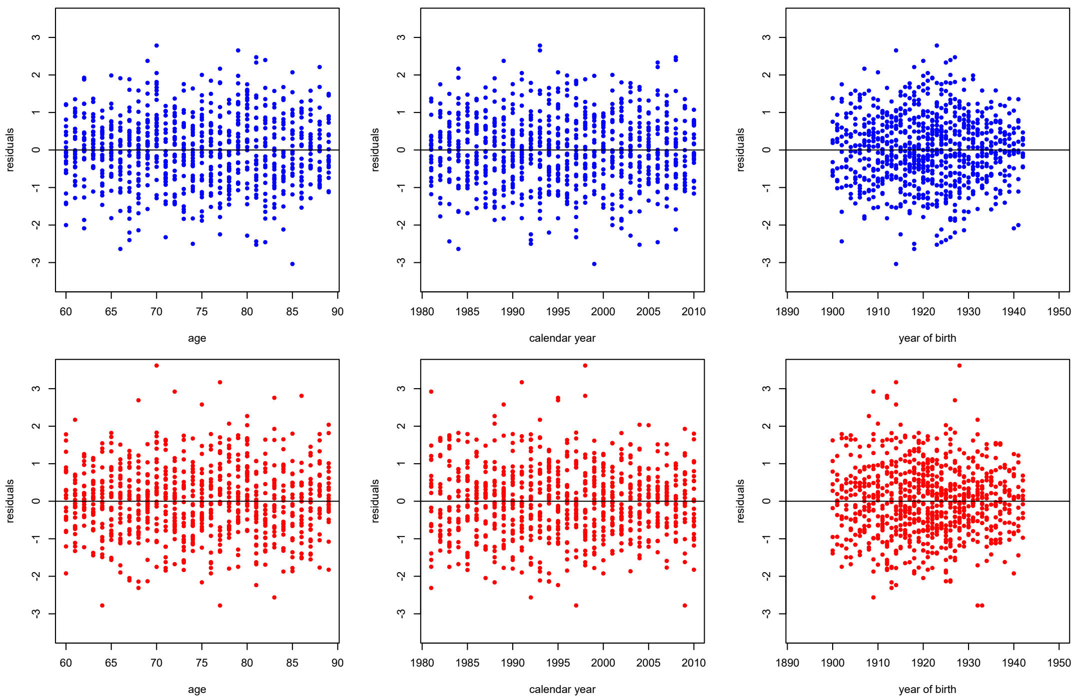


Figure 2.14 Residuals deviance of M_7 for males (top panels) and females (bottom panels) for period 1981–2010 and ages 60–89 in Greece.

Information Criteria

Generally, a better fit is expected from models with more parameters. According to Haberman and Renshaw (2011), an alternative way to address this conjecture is to penalize the model parameters using AIC (Akaike, 1974) and BIC (Schwarz, 1978) information criteria for each model. In addition, Hurvich and Tsai (1989) derived a correction of the Akaike criterion, the AIC(c), which is more suitable for small samples. Therefore, we use AIC, AIC(c) and BIC, which are defined for $M_i,\ i = 1, \ldots, 7$ as

$$AIC_i(c) = AIC_i + \frac{2k_i(k_i+1)}{n-k_i-1}, \text{ with } AIC_i = 2k_i - 2\log\hat{L}_i$$

and

$$BIC_i = (\log n)k_i - 2\log\hat{L}_i,$$

where $\hat{L}_i$ is the maximum likelihood estimate, k_i is the number of the effective parameters[5] estimated by M_i and n is the number of the observations. Smaller AIC, AIC(c) and BIC values indicate a better model fitting. Table 5.2 presents the log likelihood estimates along with the number of the effective parameters and the corresponding AIC, AIC(c) and BIC values of M_1–M_7 (ranking order in brackets), for males and females.

[5] The sum of the estimated parameters minus those that reflect each model's constraints.

Table 2.2 The log likelihood and the number of the effective parameters along with AIC(c), AIC and BIC values (ranking order in brackets) of the mortality models for males and females.

Model	Log Likelihood	Effective Parameters	AIC	AIC(c)	BIC
Males					
M_1	−4487.643	88	9151.287(7)	9172.483(7)	9566.560(7)
M_2	−4191.779	129	8641.558(4)	8689.610(4)	9250.311(4)
M_3	−4218.961	100	8637.922(3)	8665.708(3)	9109.823(2)
M_4	−4202.953	128	8661.907(5)	8709.151(5)	9265.940(5)
M_5	−4501.146	60	9122.291(6)	9131.835(6)	9405.432(6)
M_6	−4209.024	101	8620.048(2)	8648.429(2)	9096.669(1)
M_7	−4160.547	130	8581.094(1)	8629.960(1)	9194.565(3)
Females					
M_1	−4980.632	88	10,137.265(6)	10,158.461(6)	10,552.538(6)
M_2	−4254.321	129	8766.643(3)	8814.694(3)	9375.395(3)
M_3	−4367.542	100	8935.085(4)	8962.870(4)	9406.986(4)
M_4	−4235.015	128	8726.030(2)	8773.275(2)	9330.064(2)
M_5	−5279.019	60	10,678.038(7)	10,687.581(7)	10,961.178(7)
M_6	−4474.985	101	9151.969(5)	9180.349(5)	9628.590(5)
M_7	−4209.487	130	8678.975(1)	8727.841(1)	9292.447(1)

In line with BIC male results, M_6 is on top, M_3 follows and M_7 is third, while AIC and AIC(c) male rankings coincide, with M_7, M_6, M_3 being on top. Note that BIC penalizes model parameters stronger than AIC and AIC(c), we therefore expect to rank better models containing less parameters. For females, all information criteria coincide to M_7, M_4 and M_2 rank order. Unsurprisingly, M_1 and M_5 models hold the worst criteria ranking for both genders, indicating that cohort effect must be taken into account in Greek male and female mortality modelling.

Likelihood-Ratio Test

In Table 4.1, we can easily observe that some models are special cases of others. More specifically models M_1 and M_3 are nested within M_2, M_3 in M_4, M_5 nests in M_6 and M_7 and finally, M_6 is nested within M_7. In order to test the null hypothesis that the nested model is the correct versus the alternative hypothesis that the more general is correct, we follow Cairns et al. (2009) in using the Likelihood Ratio (LR) test. Six pairs of tested models and their statistics are presented for both genders in Table 5.3. LR statistic is given by $\psi^{LR} = 2\log\frac{\hat{L}_2}{\hat{L}_1}$, where $\hat{L}_2$ is the maximum likelihood estimate of the general model and $\hat{L}_1$ of the nested model, while ψ^{LR} approximates a χ^2 distribution, with $n_2 - n_1$ degrees of freedom, where n_2 are the degrees of the general model and n_1 of the nested model. For each pair of models in Table 5.3, null hypothesis is rejected in a significance level α, since $\psi^{LR} > \chi^2_{(n_2-n_1),\alpha}$ or the p-value $= 1 - F_{\chi^2_{(n_2-n_1)}}(\psi^{LR})$. Our testing results confirm information criteria rankings, suggesting that models with more parameters fit on Greek data better than more parsimonious models.

Table 2.3 Likelihood ratio test statistics for pairs of nested models (H_0) within general models (H_1).

H_0: Nested Model	**H_1: General Model**	**Likelihood Ratio Test Statistic**	**Degrees of Freedom**	**p-Value**
		Males		
M_1	M_2	591.730	41	<0.0001
M_3	M_2	54.364	29	<0.0001
M_3	M_4	32.015	28	<0.0001
M_5	M_6	584.240	41	<0.0001
M_5	M_7	681.200	70	<0.0001
M_6	M_7	96.955	29	<0.0001
		Females		
M_1	M_2	1452.600	41	<0.0001
M_3	M_2	226.440	29	<0.0001
M_3	M_4	265.050	28	<0.0001
M_5	M_6	1608.100	41	<0.0001
M_5	M_7	2139.100	70	<0.0001
M_6	M_7	530.990	29	<0.0001

2.4 Mortality Projection

In this section, we estimate future mortality rates using models M_1–M_7 for both genders. Projection methods are based on the extrapolation of period and cohort parameters for each model fitted on Greek data. Currie (2016) stated that the key point in order to obtain as accurate as possible mortality forecasts is to select the most appropriate time series models that reflect to period and cohort dynamics for a given population. In similar comparative studies, Cairns et al. (2011) and Haberman and Renshaw (2011) modelled period indices using a multivariate random walk with a drift and cohort indices with univariate ARIMA models.

In our case, we thoroughly selected an appropriate univariate ARIMA model for each period and cohort index over a range of candidate models, according to KPSS (Kwiatkowski et al., 1992), ADF (Dickey and Fuller, 1979) and PP (Phillips and Perron, 1988) unit root tests and the information criteria values. More precisely, our choice was based on time series overall performance against AIC, AICc and BIC penalized scores. Discordance issues between criteria values were addressed by preferring simpler time series models on grounds of parsimony. Therefore, κ_t's in models M_2, M_3, M_4, M_6 and M_7 are assumed to be independent of the corresponding γ_c's for each mortality model, following respectively univariate ARIMA(p,d,q) processes of the forms

$$(1-\phi_1 B-\cdots-\phi_p B^p)\,(1-B)^d\,\kappa_t=\delta+(1+\theta_1 B+\cdots+\theta_q B^q)\,e_t\,, \tag{2.6}$$

$$(1-\phi_1^{'} B-\cdots-\phi_p^{'} B^p)\,(1-B)^d\,\gamma_c=\delta^{'}+(1+\theta_1^{'} B+\cdots+\theta_q^{'} B^q)\,e_c^{'}\,, \tag{2.7}$$

where B^d is a time lag operator (also known as back-shift operator) that shifts data d periods back, δ and $\delta^{'}$ are constant drift parameters, $\phi_1,\cdots,\phi_p$ and $\phi_1^{'},\cdots,\phi_p^{'}$ are the autoregressive coefficients with $\phi_p\neq 0$, $\phi_p^{'}\neq 0$, while $\theta_1,\cdots,\theta_q$ and $\theta_1^{'},\cdots,\theta_q^{'}$ are the moving average parameters with $\theta_q\neq 0$, $\theta_q^{'}\neq 0$ and e_t, $e_c^{'}$ are white noise processes. Tables 2.4 and 2.5 present the selected ARIMA models for period and cohort indices, respectively for males and females. For all models period indices are assumed to be modelled independently. Also remind that M_1 and M_5 do not incorporate a cohort index.

Time series equations (2.6) and (2.7) were simulated to produce 1000 trajectories for future values of the period $\hat{\kappa}_{t_n+s}$ and the cohort $\hat{\gamma}_{t_n+s-x}$ indices, where $s=1,2,\cdots,20$ denotes the years of the forecasting horizon. Then, future simulated mortality values are extracted using

$$\text{logit}\hat{q}_{x,t_n+s}=\alpha_x+\sum_{i=1}^{N}\beta_x^{(i)}\,\hat{\kappa}_{t_n+s}^{(i)}+\hat{\gamma}_{t_n+s-x}\,,$$

or

$$\hat{q}_{x,t_n+s}=\frac{\exp(\alpha_x+\sum_{i=1}^{N}\beta_x^{(i)}\,\hat{\kappa}_{t_n+s}^{(i)}+\hat{\gamma}_{t_n+s-x})}{1+\exp(\alpha_x+\sum_{i=1}^{N}\beta_x^{(i)}\,\hat{\kappa}_{t_n+s}^{(i)}+\hat{\gamma}_{t_n+s-x})}\,, \tag{2.8}$$

where $t_n=2010$ is the last year of the fitting period and $\text{logit}\hat{q}_{x,t_n+s}$ denotes the logit-transform of future probabilities of death for each age x for models M_1–M_7.

Short-term male and female forecast errors of period 2011–2013 were extracted for models M_1–M_7, while for the sake of comparison, extrapolation was firstly performed by using fitted jump-off rates[6] and secondly by using actual rates for the year 2010 (Table 2.6), taken directly from HMD. Measures show that models M_2, M_3, M_4 and M_6 produce better forecasts for both genders (ranking order in brackets), either by using fitted or actual jump-off rates. Especially, when fitted rates are used, models M_2 and M_3 distinguish for both genders, while for actual rates M4 and M3 are dominant for males and M_6 outperforms for females. In any case, all of the three error measures give the higher error values for M_1, M_5 and M_7, indicating the presence of cohort effects in male and female mortality indices that cannot be captured by models M_1, M_5 and a possible overfitting behaviour of M_7.

[6]The probabilities of death in the last year of the fitting period.

Table 2.4 Selected ARIMA(p,d,q) models for the period index $\kappa_t^{(i)}$, $i = 1,2,3$ of male and female mortality models.

Model	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$
Males			
M_1	ARIMA(0,2,2)	—	—
M_2	ARIMA(0,1,1) with drift	—	—
M_3	ARIMA(1,1,0) with drift	—	—
M_4	ARIMA(0,2,2)	ARIMA(2,1,0) with drift	—
M_5	ARIMA(1,2,1)	ARIMA(2,1,0) with drift	—
M_6	ARIMA(0,2,2) with drift	ARIMA(0,1,1) with drift	—
M_7	ARIMA(1,2,1)	ARIMA(2,2,0)	ARIMA(0,1,1) with drift
Females			
Model	$\kappa_t^{(1)}$	$\kappa_t^{(2)}$	$\kappa_t^{(3)}$
M_1	ARIMA(1,1,0) with drift	—	—
M_2	ARIMA(3,1,0) with drift	—	—
M_3	ARIMA(3,1,0) with drift	—	—
M_4	ARIMA(1,1,0) with drift	ARIMA(1,1,0) with drift	—
M_5	ARIMA(0,2,2)	ARIMA(0,1,0) with drift	—
M_6	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift	—
M_7	ARIMA(2,1,0) with drift	ARIMA(2,2,0)	ARIMA(0,1,1) with drift

Table 2.5 Selected ARIMA(p,d,q) models for the cohort index γ_c of male and female mortality models.

Model	γ_c for Males	γ_c for Females
M_2	ARIMA(2,1,0)	ARIMA(2,1,1) with drift
M_3	ARIMA(0,0,1)	ARIMA(4,1,1)
M_4	ARIMA(0,0,2)	ARIMA(4,1,1)
M_6	ARIMA(0,1,3)	ARIMA(3,0,2)
M_7	ARIMA(0,0,1)	ARIMA(4,0,1)

The predictive power of mortality models was evaluated by measuring the differences (errors) between the observed and the forecasted values for the same period. For the first three out-of-sample years of projection ($t_n = 2010,\ s = 1,2,3$), in which Greek mortality data are available in HMD, forecast accuracy of models M_1–M_7 can be evaluated by averaging the mean absolute error (MAE) and the mean absolute percentage error (MAPE) values over the 3-years period for ages 60 to 89, defined by

$$MAE_{avg} = \frac{1}{3 \times (89-60+1)} \sum_{s=1}^{3} \sum_{x=60}^{89} \left| \hat{q}_{x,2010+s} - q_{x,2010+s} \right| \times 100\,, \tag{2.9}$$

$$MAPE_{avg} = \frac{1}{3 \times (89-60+1)} \sum_{s=1}^{3} \sum_{x=60}^{89} \left| \frac{\hat{q}_{x,2010+s} - q_{x,2010+s}}{q_{x,2010+s}} \right| . \tag{2.10}$$

Table 2.6 Averaged values (ranking order in brackets) of MAE and MAPE measures of the forecasting period 2011–2013 using fitted or actual jump-off rates for males and females.

Fitted Jump-off Rates							
Males							
Error	M_1	M_2	M_3	M_4	M_5	M_6	M_7
MAE_{avg}	0.332(6)	0.251(1)	0.253(2)	0.287(3)	0.327(5)	0.295(4)	0.346(7)
$MAPE_{avg}$	10.194(4)	6.496(1)	6.583(2)	9.385(3)	10.935(6)	10.559(5)	15.697(7)
Females							
MAE_{avg}	0.207(4)	0.147(1)	0.165(2)	0.219(5)	0.234(6)	0.198(3)	0.281(7)
$MAPE_{avg}$	10.363(3)	6.052(1)	7.981(2)	12.239(5)	13.396(6)	11.216(4)	22.340(7)
Actual Jump-off Rates							
Males							
Error	M_1	M_2	M_3	M_4	M_5	M_6	M_7
MAE_{avg}	0.273(6)	0.213(3)	0.208(2)	0.192(1)	0.289(7)	0.237(4)	0.247(5)
$MAPE_{avg}$	6.780(5)	5.222(2)	5.086(1)	5.371(3)	6.916(6)	6.020(4)	8.545(7)
Females							
MAE_{avg}	0.213(6)	0.180(3)	0.168(2)	0.196(4)	0.200(5)	0.165(1)	0.250(7)
$MAPE_{avg}$	7.073(5)	5.570(2)	5.336(1)	6.225(4)	7.283(6)	5.866(3)	11.818(7)

Long-term mortality projections for a 20 year horizon ahead were obtained using (2.8) with actual jump-off rates for the seven mortality models, incorporating 1000 simulation trajectories of the selected period and cohort indices. The simulated one year probabilities of death (in logarithmic scale) for models M_1–M_7 are illustrated for both genders in Figure 2.15.

Plotting results reveal the appropriateness of a mortality model for long-term forecasting. In addition, according to Cairns et al. (2011); Villegas et al. (2017) differences in uncertainty levels of each model indicates the significance of model risk in mortality forecasting.

Figure 2.15 shows that M_1, M_2 and M_3 forecasts seem to be implausible for both genders, since fans at age 85 are notably narrower than at age 65. Furthermore, M_4 and M_6 fans at age 75 show a weak, but not significant increase for both genders, while fans at age 85 show some decreasing fluctuations. On the other hand, parsimonious model M_5 performs well in general for both genders. Finally, female fans of model M_7 are narrower age 75 and 85 than at 65 and show an unreasonable increase in older ages. This is mainly because forecasts are linked to the estimated cohort effect of Figure 2.7 that exhibits a steep, upward and linear trend between cohort years 1930 and 1940.

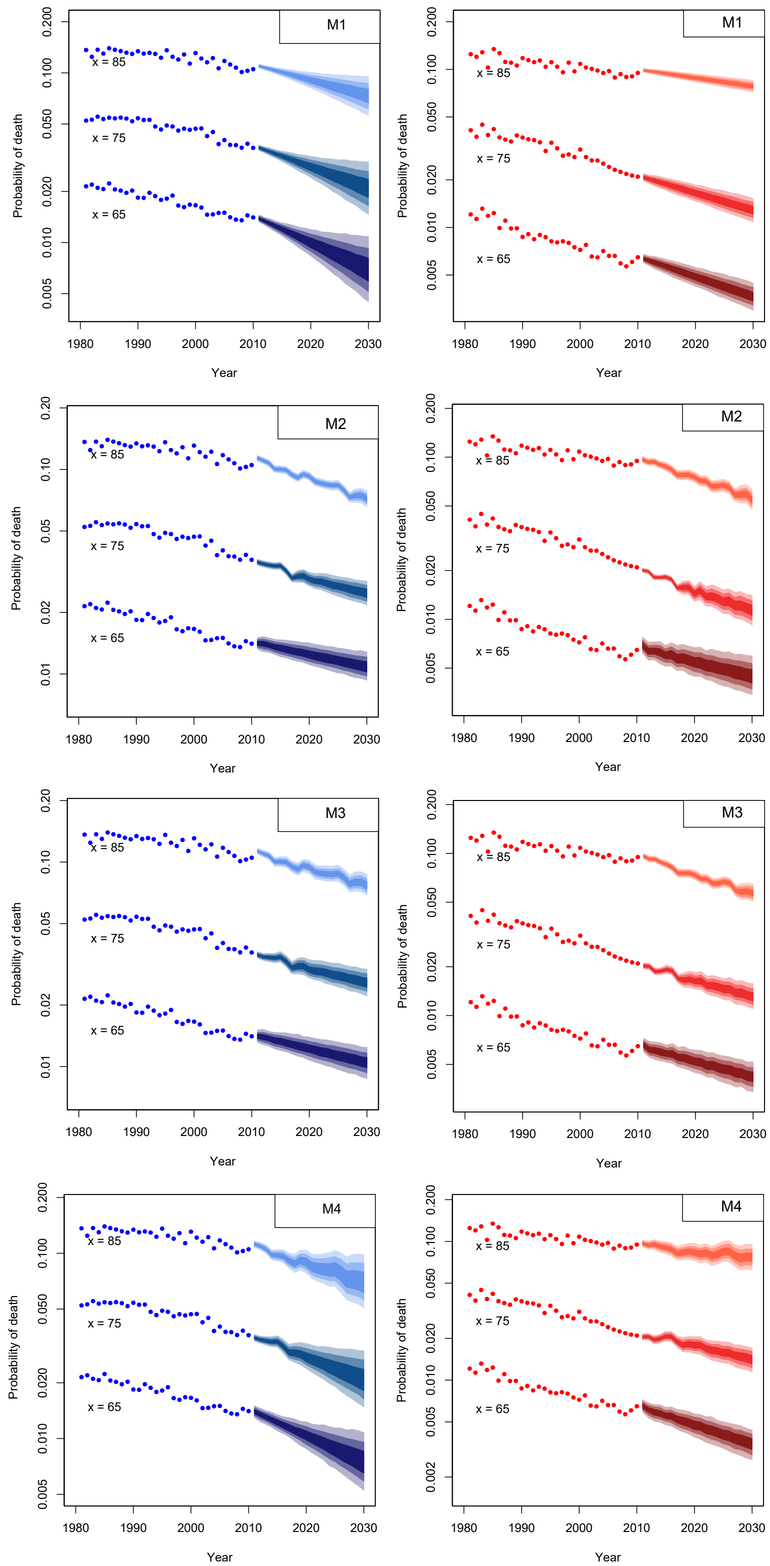

M1
M2
M3
M4
x = 85
x = 75
x = 65
Probability of death
Year
1980
1990
2000
2010
2020
2030

Fig. 2.15 *Cont.*

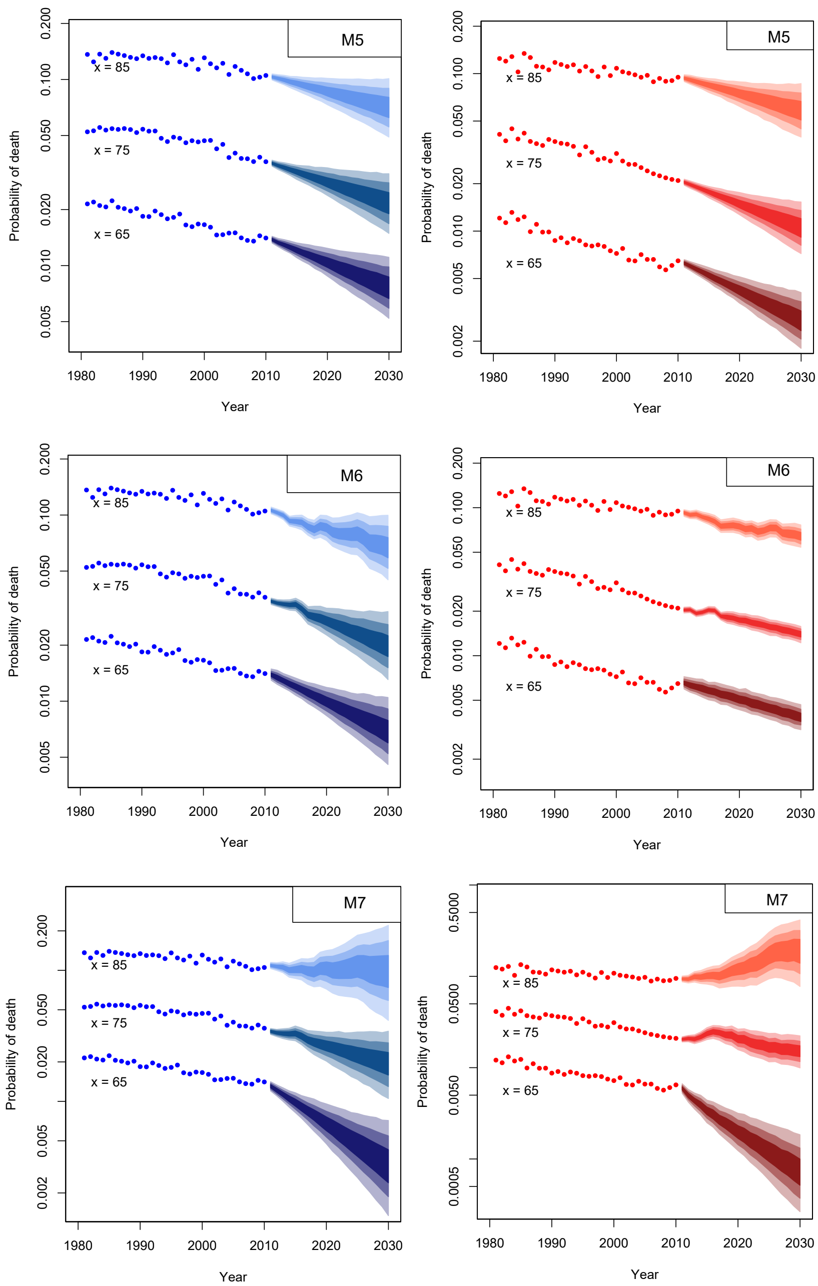


Figure 2.15 Long-term mortality projection results at ages $x = 65$ (bottom lines), $x = 75$ (middle lines) and $x = 85$ (top lines) derived from models M_1–M_7 fitted to males (left panels) and females (right panels) for ages 60–89 of the period 1981–2010. The shades regions in the projection period 2011–2030 denote the 50%, 80% and 95% prediction intervals.

2.4.1 Assessing Parameter Risk

We observe that mortality projections obtained with stochastic models incorporate only the forecast error that arises from the estimation of the period and the cohort indices, ignoring the effects of the so called parameter risk.

Especially, for countries with limited data experience such as Greece, use of bootstrapping techniques is required to address this issue. Therefore, we exploit the advantages of a residual bootstrapping method to assess the parameter uncertainty in mortality projections for the seven models, described in the previous sections.

In their study, Renshaw and Haberman (2008) proposed a residual bootstrapped method to accommodate uncertainty in estimating the parameters of the Lee-Carter model under the *Poisson* assumption of deaths. In our case, we produce bootstrap samples under the *Binomial* distribution assumption, following a slightly modified approach described by Debón et al. (2010).

Simulations were carried out using "StMoMo" *R*-package and 1000 trajectories were generated to compare the prediction intervals of the forecast error and the parameter estimation uncertainty of projections. Figure 2.16 illustrates for both genders the 95% prediction intervals for the probabilities of death at ages $x = 65$, $x = 75$ and $x = 85$ for models M_1–M_7, fitted to Greek data for ages 60–89 of the period 1981–2010. The historical rates are denoted by thick dots, solid lines denote the corresponding fitted rates and dot-dashed lines depict the 95% confidence intervals including parameter uncertainty. For the projection period 2011–2030, dashed lines represent the central forecast values and dot lines show the 95% prediction intervals excluding parameter uncertainty. The dot-dashed lines depict the 95% prediction intervals accounting for parameter uncertainty.

Figure 2.16 shows an evident parameter uncertainty in the projection period for males (left panels) of models M_2 (age 85) and M_7 (ages 65 and 85). Parameter variability is also observed for females of the same models (right panels), with an implausible upward trend for M_7 at age 75 and 85, which indicates the inappropriateness of this model to forecast female mortality at higher ages.

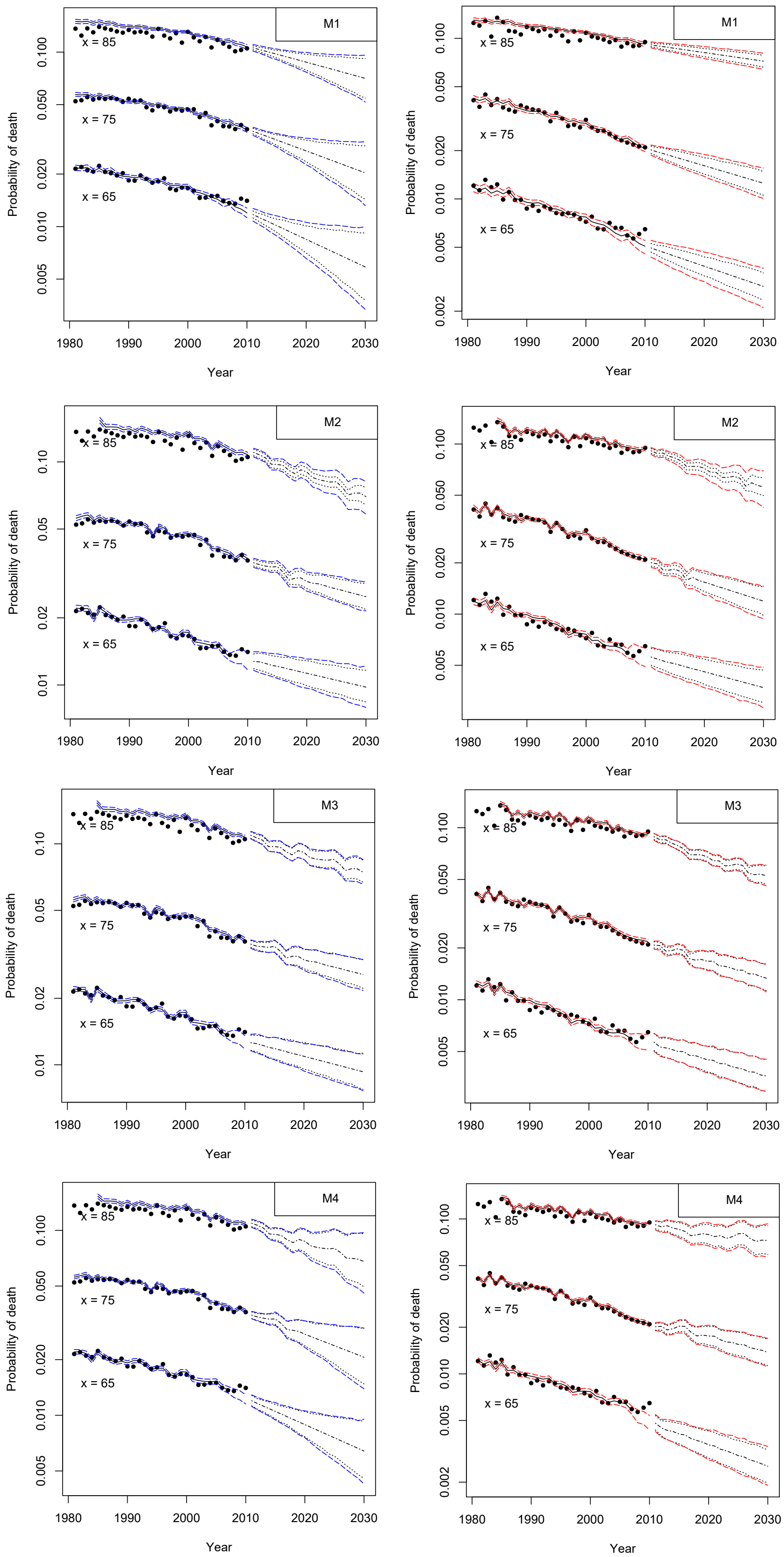


Fig. 2.16 *Cont.*

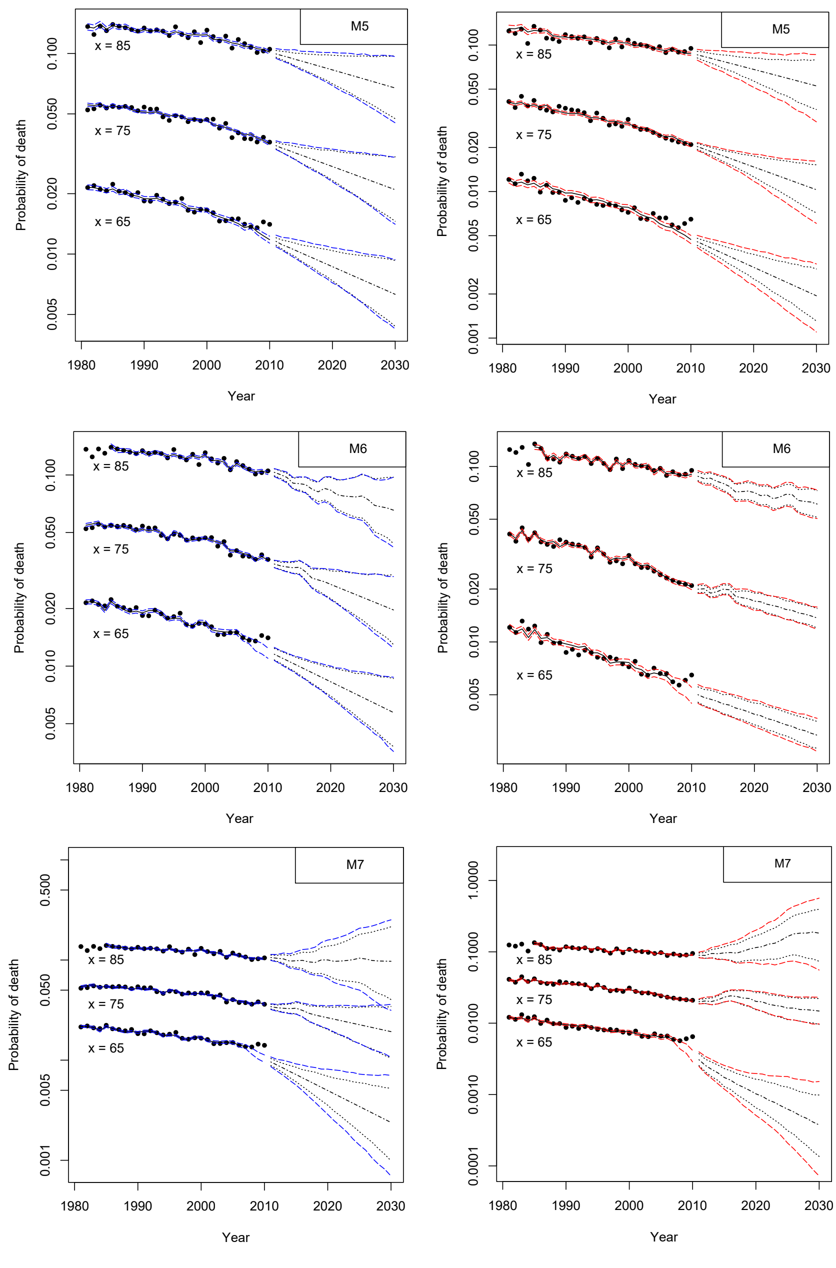


Figure 2.16 95% prediction intervals for the probabilities of death at ages $x = 65$, $x = 75$ and $x = 85$ for models M_1–M_7, fitted to males (left panels) and females (right panels) for ages 60–89 and the period 1981–2010 (thick dots). Solid lines denote the corresponding fitted rates and dot-dashed lines depict the 95% confidence intervals including parameter uncertainty. For the projection period 2011–2030, the central forecast values are given by dashed lines. Dashed lines and dot lines show the 95% prediction intervals with and without parameter uncertainty, respectively.

2.4.2 Application in Insurance-Related Products

An appropriate mortality modelling method constitutes an essential tool in pricing insurance products. In addition, as Lovász (2011) point out, insurance-related application results reflect the appropriateness of a model choice. In the following, we apply the cohort mortality forecasts obtained from M_1–M_7 to calculate life insurance premiums, similarly as in Tsai and Lin (2017b). Let us denote as $A^{\;1}_{x,t_n+1:\overline{K}|}$ the fully discrete life insurance premium issued to an insured aged x in year t_n+1, payable at the end of the year of death, if it occurs within a term of K years and as $A_{x,t_n+1:\overset{\;1}{\overline{K}|}}$ the pure endowment issued to an insured aged x in year t_n+1, payable at the end of K years in case of being alive. Net premiums (NP) are given respectively by

$$A^{\;1}_{x,t_n+1:\overline{K}|} = \sum_{k=0}^{K-1} {}_kp_{x,t_n+1} \cdot q_{x+k,t_n+1+k} \cdot (1+i)^{-(k+1)} , \tag{2.11}$$

$$A_{x,t_n+1:\overset{\;1}{\overline{K}|}} = {}_Kp_{x,t_n+1} \cdot (1+i)^{-K} , \tag{2.12}$$

where ${}_kp_{x,t_n+1}$ denotes the k-year survival probability for age x in year t_n+1, while its estimate is given by ${}_k\hat{p}_{x,t_n+1} = \hat{p}_{x,t_n+1} \cdot \ldots \cdot \hat{p}_{x+k-1,t_n+1+k-1}$, $k = 1, \ldots, K-1$ (similarly for ${}_K\hat{p}_{x,t_n+1}$), i is the interest rate and ${}_0\hat{p}_{x,t_n+1} = 1$.

Since mortality projection models are typically used for pension applications, it would also be beneficial to see the performance of a life annuity product. Let us denote as $\ddot{a}_{x,t_n+1:\overline{k}|}$ a discrete life annuity-due of an insured aged x in year t_n+1, payable on an annual basis for up to K years, so long as insured survives. Actuarial present value (APV) is given by

$$\ddot{a}_{x,t_n+1:\overline{K}|} = \sum_{k=0}^{K-1} {}_kp_{x,t_n+1} \cdot (1+i)^{-k} . \tag{2.13}$$

Hence, we apply the estimated mortality rates obtained from M_1–M_7, fitted to 1981–2000 with actual jump-off rates to calculate life insurance NPs and annuity APV for ages 60–79 with $K = 10$, assuming $i = 4\%$. As before, we use averaged MAE and MAPE to evaluate the errors between forecasted NPs and those produced from the observed mortality rates for the years 2001–2010. For each model, error measures for life insurance premiums are given by

$$MAE_x^{(K=10)} = \frac{1}{20} \sum_{x=60}^{79} \left| \hat{A}^{\;1}_{x,2001:\overline{10}|} - A^{\;1}_{x,2001:\overline{10}|} \right| \times 100 , \tag{2.14}$$

$$MAPE_x^{(K=10)} = \frac{1}{20} \sum_{x=60}^{79} \left| \frac{\hat{A}^{\;1}_{x,2001:\overline{10}|} - A^{\;1}_{x,2001:\overline{10}|}}{A^{\;1}_{x,2001:\overline{10}|}} \right| . \tag{2.15}$$

Similarly, MAE and MAPE formulas are adjusted for pure endowment or annuity products by replacing $A^{1}_{x,t_n+1:\overline{K}|}$ with $A_{x,t_n+1:\overline{K}|}^{\;\;\;1}$ or $\ddot{a}_{x,t_n+1:\overline{K}|}$ in (2.14) and (2.15). Table 2.7 presents the averaged values of MAE and MAPE values in ranking order for a 10 year forecasted life insurance, pure endowment and life annuity using actual jump-off rates for males and females, aged 60–79 in 2001–2010.

Table 2.7 Averaged values (ranking order in brackets) of MAE and MAPE measures for 10 year forecasted life insurance, pure endowment and life annuity values using actual jump-off rates for males and females, aged 60–79 in 2001–2010.

Error	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Life Insurance							
Males							
MAE_x	2.222(6)	1.242(1)	2.284(7)	2.199(5)	2.020(4)	1.456(2)	1.799(3)
$MAPE_x$	7.651(6)	5.536(1)	8.895(7)	7.626(5)	7.412(4)	5.557(2)	6.490(3)
Females							
MAE_x	1.605(6)	0.870(1)	0.885(2)	1.494(5)	0.914(3)	1.016(4)	2.150(7)
$MAPE_x$	9.264(5)	6.404(1)	6.901(3)	9.268(6)	6.426(2)	6.930(4)	11.883(7)
Pure Endowment							
Males							
MAE_x	1.605(6)	0.927(1)	1.666(7)	1.590(5)	1.451(4)	1.039(2)	1.293(3)
$MAPE_x$	4.114(7)	2.190(1)	4.094(6)	4.064(5)	3.619(4)	2.531(2)	3.212(3)
Females							
MAE_x	1.198(6)	0.623(1)	0.651(2)	1.091(5)	0.690(3)	0.738(4)	1.556(7)
$MAPE_x$	2.615(6)	1.282(2)	1.242(1)	2.250(5)	1.408(3)	1.565(4)	3.240(7)
Life Annuity							
Males							
MAE_x	7.711(6)	5.506(2)	8.132(7)	7.637(5)	6.781(4)	5.225(1)	5.924(3)
$MAPE_x$	1.127(6)	0.785(2)	1.168(7)	1.112(5)	0.980(4)	0.748(1)	0.856(3)
Females							
MAE_x	5.484(6)	2.465(1)	2.944(2)	4.995(5)	3.254(4)	3.091(3)	6.466(7)
$MAPE_x$	0.754(6)	0.325(1)	0.386(2)	0.673(5)	0.439(4)	0.416(3)	0.873(7)

Figure 2.17 illustrates the MAE and MAPE values against age for life insurance and annuity products, respectively for males (left panels) and females (right panels) for the top four models in ranking. According to MAE and MAPE values, models M_2, M_6 and M_7 produce better insurance-related forecasts for males, while M_2, M_3 and M_5 are the top ranked models for females. For both genders, measures show that M_2 outperforms

in aggregate. However, regarding its robustness (especially for female data) and taking into account values of Table 2.7, a good insurance-related model choice should also be M_6 for males and M_3 for females. This fact is also evident in Figure 2.17, where absolute error and absolute percentage error values against age for the corresponding models lie on the lower levels for all the insurance products.

2.5 Results

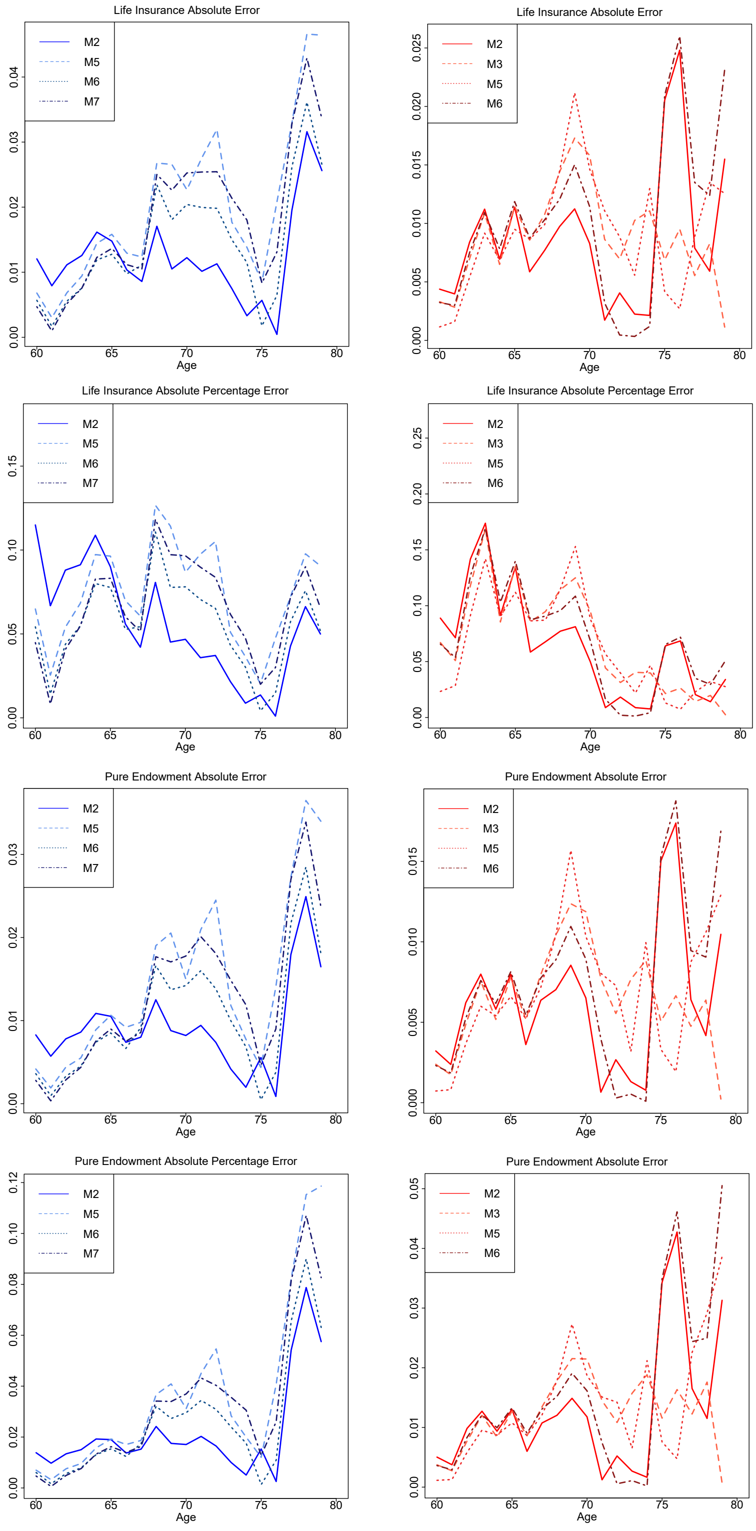
In this section, we summarize the fitting and forecasting results of this analysis and our findings are compared with the corresponding results obtained from the original papers. Our study shows that all the models capture effectively the period effects for both genders. We can also notice that the most parsimonious models M_1 and M_5 do not capture the cohort effect as it is illustrated in the right panels of male and female scatter plots of residual deviance in Figures 2.8 and 2.12. Furthermore, models M_5 and M_6 seem to be inadequate to capture the age effects, especially for females (left panels in Figures 2.12 and 2.13).

AIC and AIC(c) scores coincide to the fact that models M_7, M_6 and M_3 outperform in ordered ranking for males, while in the BIC rankings[7] M_6 is on top, M_3 follows and M_7 is third. For females, all measure values show that M_7 comes out first, M_4 second and M_2 follows. For both genders, models M_1 and M_5 have the worst criteria ranking for both genders, lacking a cohort term that must be taken into account in Greek male and female mortality modelling. Likelihood ratio results confirm the information criteria ranking, indicating that the more complicated models $M2$, M_4, M_6 and M_7 are in aggregate more suitable for modelling Greek mortality.

Mortality projections derived from the seven models are illustrated for both genders in Figure 2.15. Plotting results show that long-term forecasts from model M_1, M_2 and M_3 seem to be unreliable for both genders, since figures at age 85 are notably narrower than at age 65. In addition, model M_7 for females shows an implausible increase of mortality rates at ages 75 and 85. However, forecast accuracy measures of Table 2.6 suggests that models M_2, M_3, M_5 and M_6 produce better short-term forecasts for both genders.

Parameter uncertainty is evident in models M_2 and M_7 for both genders in higher ages (Figure 2.16). Parameter variability is also observed in model M_5 for females, while while the implausible upward trend for M_7 at age 75 and 85 raise some questions regarding the appropriateness of this model to forecast Greek female mortality.

[7] Inconsistency in male ranking results is expected, since BIC criterion penalizes stronger models with more parameters.



Life Insurance Absolute Error
M2
M5
M6
M7
Life Insurance Absolute Error
M2
M3
M5
M6
Life Insurance Absolute Percentage Error
Pure Endowment Absolute Error
Pure Endowment Absolute Percentage Error
Age

Fig. 2.17 *Cont.*

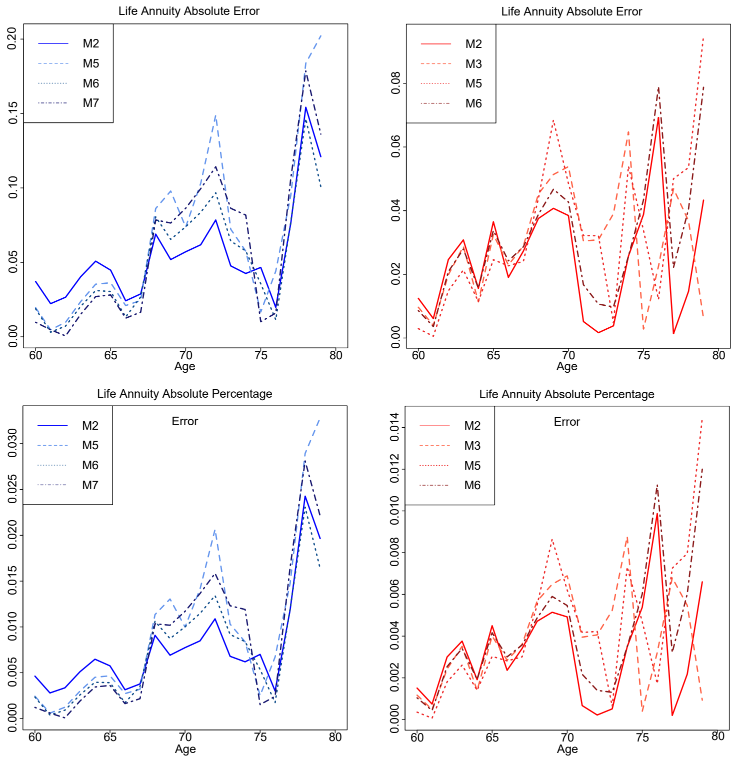


Figure 2.17 Absolute error and absolute percentage error values of life insurance and annuity products for the top four models in ranking for males (left panels) and females (right panels).

2.5.1 Comparison with Original Papers

Here, we present the commonalities and differences between estimation results of our study and the corresponding findings obtained from the original papers.

Lee and Carter (1992) modelled the mortality rates of the entire United States population for grouped ages 0–85^{+} of years 1933–1987. The same year, Carter and Lee (1992) implemented their model for males and females separately, using the SVD method to derive forecasts of the $\kappa_t^{(1)}$ time index for a full range of grouped ages 0–85^{+} of the years 1933–1988. Fitted values of α_x and $\beta_x^{(1)}$ for Greek males and females of our study show similar trends with the corresponding results obtained for both genders of the United States population data. Likewise, comparing our estimates with the corresponding Belgian results obtained from the Poisson Lee-Carter approach of Brouhns et al. (2002), we observed that their maximum likelihood estimates of α_x, $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ for ages 60–98 in the years 1960–1998 are in line with the Greek results, especially for males, where the estimates lie between the same levels.

The Haberman and Renshaw (2011) model estimates obtained from fitting ages 55–89 of years 1961–2007 for England & Wales male data. Even if estimates show similar patterns, they cannot be directly compared with our results, since authors used different model constraints in their analysis (see Haberman and Renshaw, 2011, p. 37).

The Currie (2006) model was initially fitted to selected assured lives, aged 20–90, for the years 1947–2002. Estimates of α_x are in accordance with our results for both genders. The period component of the model shows an upward trend between 1950–1975, but after the year 1980, it complies with Greek patterns.

Plat (2009) fitted his model to three different data sets of males, for the United states (ages 20–84, years 1961–2005), the England & Wales (ages 20–89, years 1961–2005) and the Netherlands (ages 20–90, years 1951–2005). Estimated parameters α_x and $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and γ_{t-x} were illustrated only for the United States mortality data. Although his α_x and $\kappa_t^{(1)}$ male estimates were based on a wide age range of data fitted onto his extended model form, they totally agree with our corresponding values exported from the reduced model form used for Greek data of ages 60–89. In the contrary, $\kappa_t^{(2)}$ parameter estimates have completely different trends for the entire common period, while the γ_{t-x} parameter values show similar patterns with the Greek males between cohort years 1890–1930 (Figure 2.4).

Cairns et al. (2006) illustrated their model using England & Wales data of males, aged 60–90, for the years 1961–2002. Their results show that $\kappa_t^{(1)}$ estimated values have a steep downward trend for the whole fitting period, while $\kappa_t^{(2)}$ values follow an opposite upward trend for the same years. Estimates of these two parameters are obviously similar with the corresponding Greek results, obtained for ages 60–89 of period 1981–2010. This is more evident for $\kappa_t^{(2)}$ parameter, where its values lie between the same levels for all countries.

The Cairns et al. (2009) "cohort" extension of the Cairns et al. (2006) model was fitted to England & Wales (1961–2004) and United States (1968–2003) male data for ages 60–89. $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ estimates for England & Wales data are in accordance with the Greek values, showing a decreasing trend for the first parameter and an upward trend for the second one, respectively. Cohort estimates lie between the same levels with the Greek ones, with an exception after birth year 1935, where Greek cohort estimates jump abruptly to higher levels. The corresponding κ_t results for the United States data are quite similar, but the γ_{t-x} parameter estimates show a steep fall around the year 1920 in comparison with the derived Greek results.

The "quadratic" extension of the Cairns et al. (2006) model was illustrated in Cairns et al. (2009) for the England & Wales (1961–2004) and the United States (1968–2003) male data for ages 60–89. Although $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and γ_{t-x} estimates for England & Wales and United States males take similar values and show the same patterns with the corresponding Greek population data, some differences are observed in the estimates $\kappa_t^{(3)}$, where a steep, upward trend between the years 1985–2003 is evident for both countries in contrast with the decreasing Greek values.

2.6 Concluding Remarks

A comparative analysis of seven stochastic mortality models of a common APC framework was conducted for Greek male and female data. The fitting behaviour of each model was examined using specific criteria and the corresponding forecasting results were presented. Fitting behaviour of each model was evaluated using AIC, AIC(c) and BIC information criteria, as well as the likelihood ratio test. Models M_3, M_6 and M_7 for males and M_2 , M_4 and M_7 for females were respectively distinguished for their fitting performance.

Although in such analyses is highly important all of the considered models should provide a good fit to historical data, it does not imply that a model which fits better the historical data does necessarily give the best forecasting results. That point was also underlined in similar studies that have been conducted for other datasets in the literature (Cairns et al., 2011). Especially for the case of Greece, a cohort effect was identified in the data that was accounted for the selection of the most appropriate mortality model.

The accuracy of the short-term forecasts was assessed by the MAE and the MAPE error values. Backtesting results show that models M_2, M_3 and M_4 for males and M_2, M_3 and M_6 for females provide with the most reliable short-term forecasts. Parameter uncertainty was also identified in some cases (more evident in M_2 for males and M_7 for females), indicating the inappropriateness of the corresponding models for long-term forecasts.

In addition, parameter estimates for Greek data were compared with the corresponding results obtained from the original papers, where each model was initially discussed and implemented, revealing several commonalities in patterns.

The main contribution of this chapter is to be the first work in comparing the fitting and the forecasting performance of the APC mortality models on Greek data, with applications in pricing insurance-related products. Unfortunately, the limited availability of historical data was an additional drawback for a more effective modelling.

Finally, animated Figures 2.18–2.24 illustrate the overall evolution of death rates for each model as an alternative, interactive way to present our results. Animated figures were created using the “animation” *R*-package. For instructions on using this package, see Xie (2013).

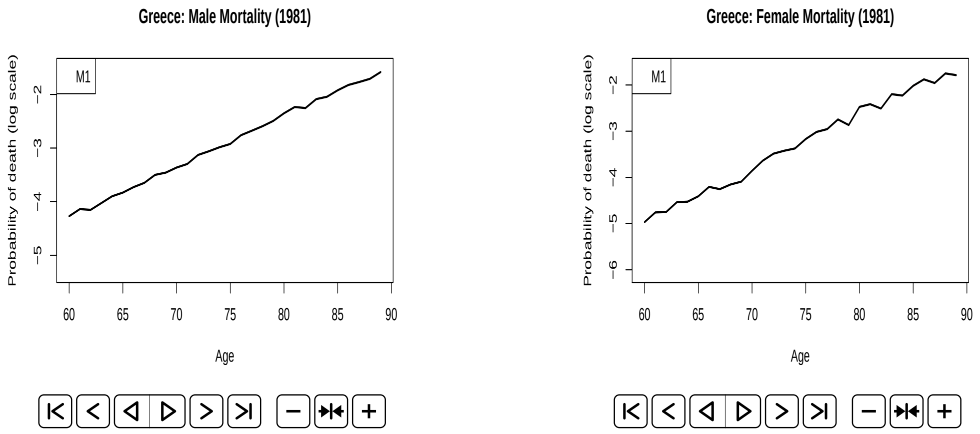


Figure 2.18 M_1 for males and females.

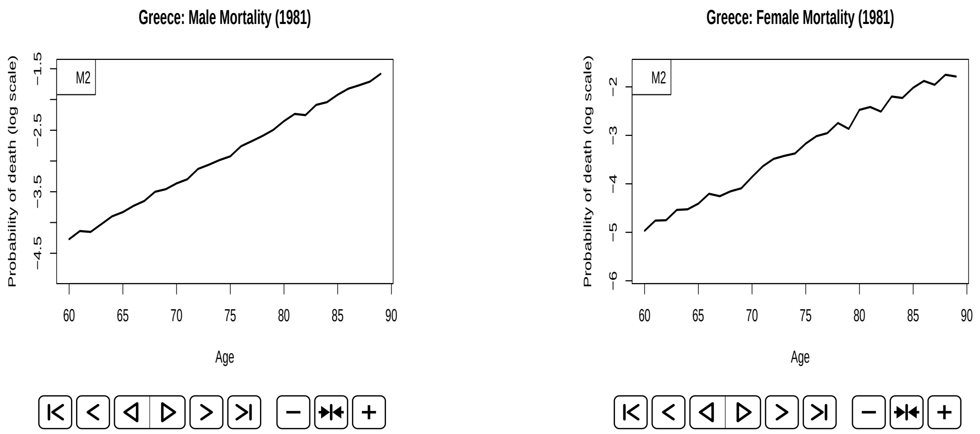


Figure 2.19 M_2 for males and females.

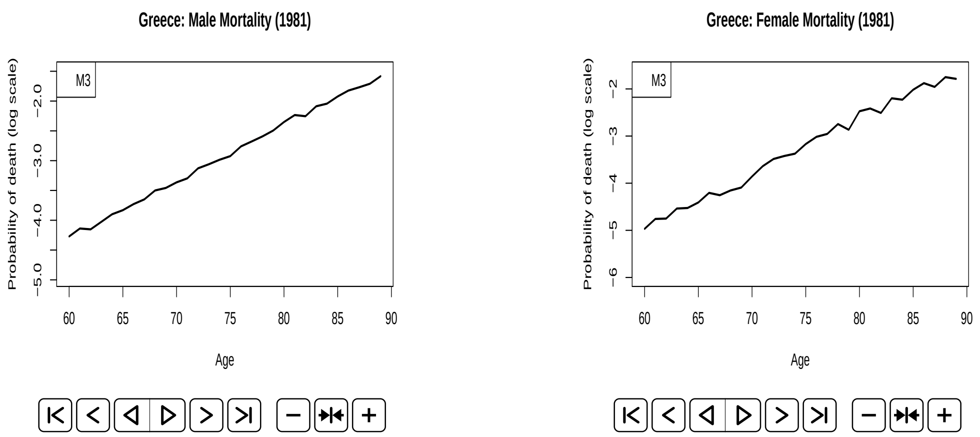


Figure 2.20 M_3 for males and females.

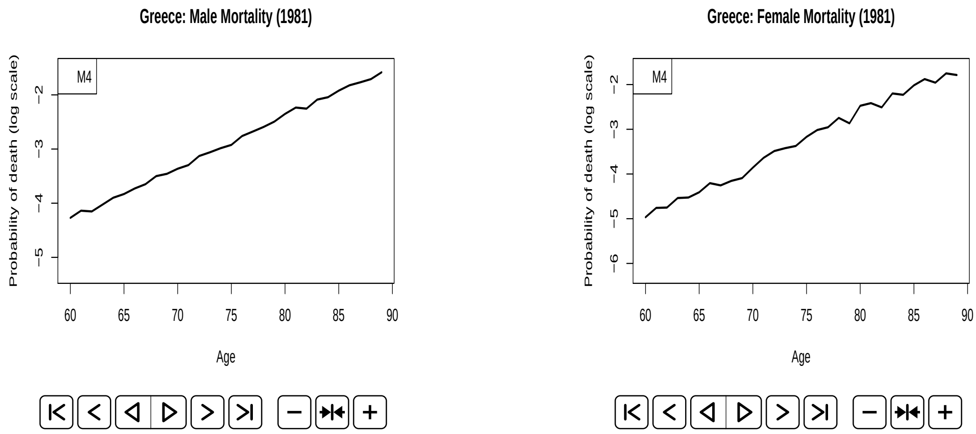


Figure 2.21 M_4 for males and females.

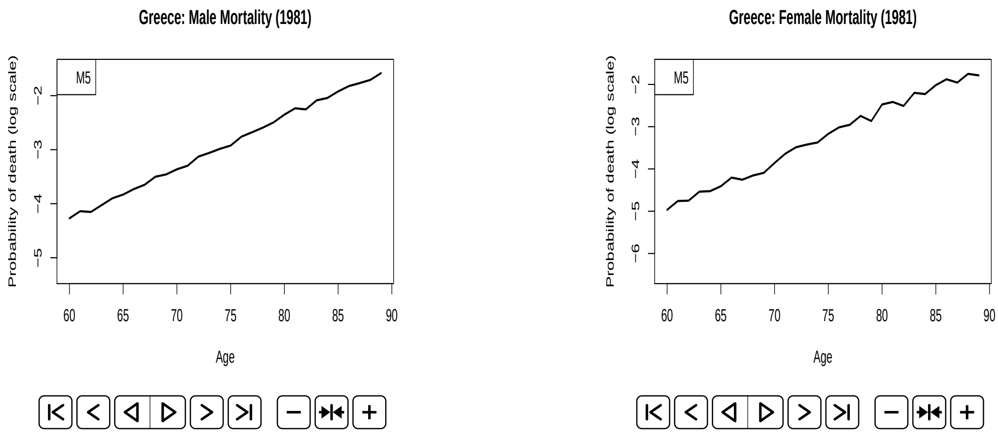


Figure 2.22 M_5 for males and females.

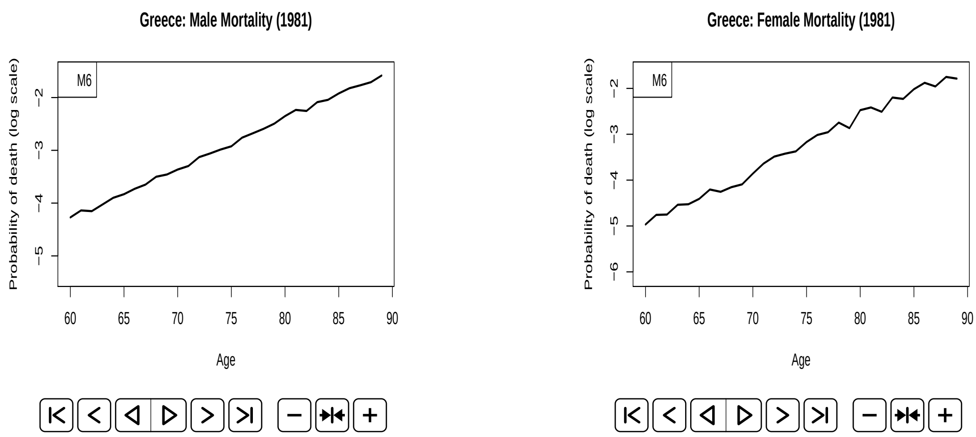


Figure 2.23 M_6 for males and females.

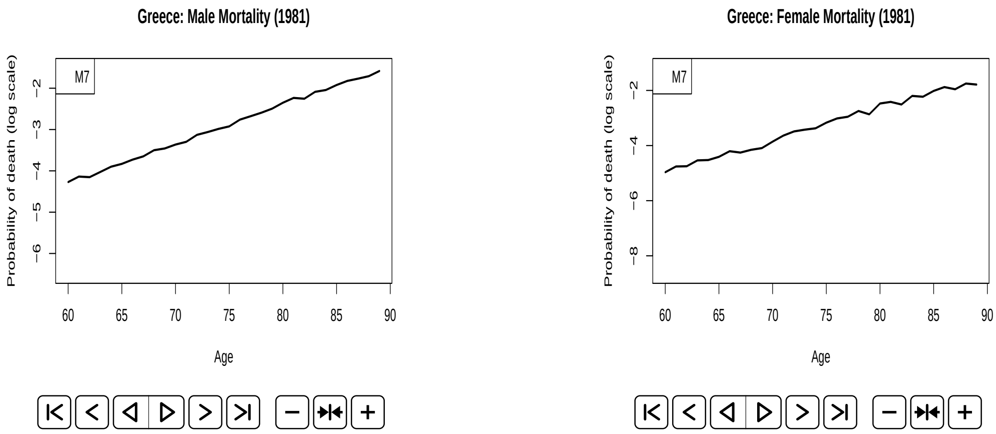


Figure 2.24 M_7 for males and females.

Chapter 3

Credibility Regression Mortality Models for Populations with Limited Data

3.1 Introduction

An issue that sometimes appears in mortality modelling is that, for some countries, there are too few data to fit. This issue affects the existing modelling methods, which inevitably base their forecasts on population datasets of a limited historical period of observations. In the literature, there are extensions of the Lee–Carter method that can be utilized when dealing with limited datasets. For instance, Li et al. (2004) extended the Lee–Carter model to be applied for Chinese and South Korean mortality data, which are available at only a few points in time and at unevenly spaced intervals. Zhao (2012) modified the Lee–Carter model by incorporating linearized cubic splines and other additive functions to approximate the model parameters and forecast mortality for short-base-period Chinese data. Also, Huang and Browne (2017) presented a stochastic modification of the CMI (Continuous Mortality Investigation) model to project mortality improvement rates for limited Chinese data using clustering analysis techniques.

Recently, some alternative modelling approaches have also been proposed as a tool in mortality forecasting. Differently from the above Lee–Carter variants and extensions, these approaches are based on credibility theory, aiming to model the period patterns of limited mortality data for a specific age, using information from a wider age span. Bühlmann (1967) established the theoretical foundation of modern credibility theory and Hachemeister (1975) introduced a credibility regression model to estimate auto-mobile bodily injury claims for various states in the USA.

Credibility regression has a long history in credibility literature, with applications mainly in non-life insurance. De Vylder (1978) proposed credibility estimators for the structural parameters in a more general regression model. Norberg (1980) proposed empirical credibility estimators under various model assumptions and established asymptotic optimality. Ledolter et al. (1991) derived a credibility method that allows for time-varying parameters in the process. Pitselis (2004) presented the relationship between claim amounts and a set of explanatory variables into a credibility regression model with cross-section and time effects, with applications for general insurance data.

Two recent contributions to modelling mortality under a credibility framework were made by Tsai and Lin (2017a, 2017b). In the first paper, they applied Bühlmann credibility to mortality data of Japan, the United Kingdom and the United States, while, in the second one, they incorporated Bühlmann credibility into the Lee and Carter (1992) model, the Cairns–Blake–Dowd model (Cairns et al., 2006) and the linear relational model of Tsai and Yang (2015) to improve forecasting performance for the Unitingdomed Kingdom dataset.

However, it has been observed that the age-specific mortality rates show a clear downward trend over time. Moreover, when we have limited mortality data experience for a specific age, but extensive data experience for the entire age range, the use of credibility regression techniques should be preferred to capture mortality trends. Our study aims to exploit the advantages of credibility regression compared with the most widely used mortality models, as an alternative to Bühlmann credibility, to forecast the mortality rates, especially for populations with limited data.

The rest of this chapter is organized as follows: Section 3.2 briefly reviews the Lee–Carter, the Cairns–Blake–Dowd and the random coefficients regression models. Section 3.3 proposes a credibility regression approach with random coefficients and a special case with fixed coefficients to model mortality rates. Section 3.4 presents the extrapolation methods used to estimate future mortality rates under the credibility regression approaches. An empirical illustration using Greek male and female data is presented in Section 3.5.1, in which forecasting performances of credibility regression, and the Lee–Carter and Cairns–Blake–Dowd methods are evaluated with the MAFE and RMSFE measures. A comparison between Bühlmann credibility and credibility regression forecasting methods is also presented in Section 3.5.3 and an application on pricing insurance-related products follows in Section 3.5.4. Finally, concluding remarks are discussed in Section 3.6.

3.2 Mortality Modelling: A Review of Methods

In this section, we briefly review the Lee–Carter model, the Cairns–Blake–Dowd model and the random coefficients regression models that will be utilized in next sections.

3.2.1 The Lee–Carter Model

In its original form, the Lee–Carter (LC) model links the natural logarithm of the observed mortality rates $Y_{t,x} = \log m(t,x)$ for age $x = x_0, \dots, x_{k-1}$ and year $t = t_0, \dots, t_{n-1}$ with the following model predictor

$$Y_{t,x} = \alpha_x^{(1)} + \alpha_x^{(2)} \kappa_t + \varepsilon_{t,x}, \tag{3.1}$$

where $\alpha_x^{(1)}$ is an age parameter that reflects the average mortality at age x, κ_t is a period parameter which indicates the general level of mortality in year t and $\alpha_x^{(2)}$ is an age parameter that indicates the deviation from the average mortality at age x, as the general level of mortality changes. The errors $\varepsilon_{t,x}$ are expected to be normally distributed, with zero mean and constant variance, reflecting specific period and age effects not captured by the model. Thus, after assuming that errors are independent and homoscedastic with zero mean, Lee and Carter (1992) suggested a close approximation to the SVD (Singular Value Decomposition) method, under the constraints $\sum_{x=x_0}^{x_{k-1}} \alpha_x^{(2)} = 1$ and $\sum_{t=t_0}^{t_{n-1}} \kappa_t = 0$, to obtain the following parameter estimates

$$\widehat{\alpha}_x^{(1)} = \frac{1}{t_{n-1} - t_0 + 1} \sum_{t=t_0}^{t_{n-1}} \log m(t,x) \ , \quad \widehat{\kappa}_t = \sum_{x=x_0}^{x_{k-1}} [\log m(t,x) - \widehat{\alpha}_x^{(1)}] \ ,$$

$$\widehat{\alpha}_x^{(2)} = \frac{\sum_{t=t_0}^{t_{n-1}} [\log m(t,x) - \widehat{\alpha}_x^{(1)}] \widehat{\kappa}_t}{\sum_{t=t_0}^{t_{n-1}} \widehat{\kappa}_t^2} \ .$$

Later on, to allow for heteroscedasticity in error variance, Brouhns et al. (2002) assumed that $D(t,x)$ follows a Poisson distribution with mean $m(t,x) \cdot E(t,x)$. Under this approach, age and period parameters are estimated by maximising the log-likelihood function of (3.1). After choosing one of the above estimation approaches, period estimates are extrapolated using time series methods. Lee and Carter (1992) suggested a random walk with a drift parameter $\widehat{\theta}$ to project period parameter for $h = 1, 2, \dots, H$ years ahead, according to $\widehat{\kappa}_{t_{n-1}+h} = \widehat{\kappa}_{t_{n-1}} + \widehat{\theta} h$. Then, projected κ_ts are utilized along with the estimates of age parameters $\alpha_x^{(1)}$ and $\alpha_x^{(2)}$ to obtain the following mortality forecasts

$$\widehat{Y}_{t_{n-1}+h,x} = \widehat{\alpha}_x^{(1)} + \widehat{\alpha}_x^{(2)} \widehat{\kappa}_{t_{n-1}+h} = \widehat{Y}_{t_{n-1},x} + (\widehat{\alpha}_x^{(2)} \widehat{\theta}) h, \text{ for } h = 1, 2, \dots, H. \tag{3.2}$$

3.2.2 The Cairns–Blake–Dowd Model

The Cairns–Blake–Dowd (CBD) model links the logit transformation of one-year probabilities of death $Y_{t,x} = \operatorname{logit} q(t,x)$ with the following model predictor

$$Y_{t,x} = \operatorname{logit} q(t,x) = \kappa_t^{(1)} + (x - \bar{x}) \kappa_t^{(2)} + \varepsilon_{t,x}, \tag{3.3}$$

where $\kappa_t^{(1)}$ is a period parameter which indicates the general level of mortality in year t and $\kappa_t^{(2)}$ is a period parameter that shows how mortality affects each age, while $\bar{x}$ is the mean age of the considered fitting age interval and $\varepsilon_{t,x}$ reflects specific effects not captured by the model and is expected to be normally distributed, with zero mean and constant variance. Again, we briefly present the estimates of the model parameters, which can be obtained by regressing $\operatorname{logit} q(t,x)$ on $(x-\bar{x})$ for each t

$$\widehat{\kappa}_t^{(1)} = \frac{1}{x_{k-1}-x_0+1}\sum_{x=x_0}^{x_{k-1}} \operatorname{logit} q(t,x) \text{ and } \widehat{\kappa}_t^{(2)} = \frac{\sum_{x=x_0}^{x_{k-1}}[\operatorname{logit} q(t,x)(x-\bar{x})]}{\sum_{x=x_0}^{x_{k-1}}(x-\bar{x})^2} .$$

Alternatively, Cairns et al. (2009) assumed that deaths follow a Poisson distribution with mean $m(t,x)\cdot E(t,x)$, where $m(t,x) = -\log[1-q(t,x)]$. Then, the CBD model parameters are obtained by maximizing the log-likelihood function of (3.3). Assuming that period estimates are independent, each one of them is extrapolated using a random walk with a drift parameter ($\widehat{\theta}_i,\ i=1,2$) and then mortality forecasts for $h=1,2,\ldots,H$ are obtained by

$$\widehat{Y}_{t_{n-1}+h,x} = (\widehat{\kappa}_{t_{n-1}}^{(1)} + \widehat{\theta}_1 h) + (x-\bar{x})(\widehat{\kappa}_{t_{n-1}}^{(2)} + \widehat{\theta}_2 h) = \widehat{Y}_{t_{n-1},x} + [\widehat{\theta}_1 + (x-\bar{x})\widehat{\theta}_2]h . \quad (3.4)$$

Remark 3.1. We can easily observe that expressions in Equations (3.2) and (3.4) are both linear functions of the forecasting horizon h, where their intercepts are equal to the fitted rates of the last observed year and their slopes are the products of the estimated age parameters with the drift terms.

3.2.3 The Random Coefficients Regression Model

Empirical data indicate that mortality in each age $x = x_0,\ldots,x_{k-1}$ decreases over time. Especially in higher ages, mortality rates have been significantly improving over the last few years. We are interested in a model structure able to capture the improvement trends and describe the mortality evolution through time. For this reason, we consider a regression structure with random coefficients, aiming to capture the underlying mortality effects that are not included in the explanatory variables.

For each age x, the regression model with random coefficients is defined by $Y_{t,x} = \beta_{1t,x} + \sum_{k=2}^{p}\beta_{kt,x}Z_{kt,x} + \varepsilon_{0t,x}$, for $t = t_0,t_1,\ldots,t_{n-1}$, where $Y_{t,x}$ is the response variable, $\beta_{kt,x},\ k=1,2,\ldots,p$ are the random coefficients and $Z_{kt,x}$ are the explanatory variables. Then, each coefficient element can be decomposed in $\beta_{kt,x} = \beta_{k,x} + \varepsilon_{kt,x}$, for all t and k, with β_k,x and $\varepsilon_{kt,x}$ being the fixed and random parts, respectively, assuming that $E(\varepsilon_{kt,x}) = 0$, $\operatorname{Var}(\varepsilon_{kt,x}) = \sigma_{k,x}^2$ for all t and $\operatorname{Cov}(\varepsilon_{kt,x},\varepsilon_{k't',x}) = 0$ for $k \neq k'$ and $t \neq t'$. The above formulation means that the unknown regression coefficients can take different values over an observed period. Actually, mortality dynamics for a specific age can vary over time, due to unknown or exogenous[1] factors. For more details on regression models with random coefficients, we refer to the works of Hildreth and Houck (1968), Hsiao (1986) and Greene (2012).

The random coefficients regression model may be reduced to a fixed coefficients model with heteroscedastic variances, defined as

$$Y_{t,x} = \beta_{1,x} + \sum_{k=2}^{p} \beta_{k,x} Z_{kt,x} + v_{t,x}, \text{ with } v_{t,x} = (\varepsilon_{0t,x} + \varepsilon_{1t,x}) + \textstyle\sum_{k=2}^{p} Z_{kt,x}\varepsilon_{kt,x}, \quad (3.5)$$

where

$$E(v_{t,x}) = 0,\ \mathrm{Var}(v_{t,x}) = (\sigma_{0,x}^2 + \sigma_{1,x}^2) + \textstyle\sum_{k=2}^{p} \sigma_{k,x}^2 Z_{kt,x}^2 \text{ and } \mathrm{Cov}(v_{t,x}, v_{t',x}) = 0, \quad (3.6)$$

for all x and t, with $t \neq t'$. We have to point out that error variances $\sigma_{0,x}^2$ and $\sigma_{1,x}^2$ cannot be identified separately, while the sum $(\sigma_{0,x}^2 + \sigma_{1,x}^2)$ can be. Therefore, without loss of generality, $\sigma_{0,x}^2$ is dropped and the above variance is simplified to $\mathrm{Var}(v_{t,x}) = \sigma_{1,x}^2 + \sum_{k=2}^{p} \sigma_{k,x}^2 Z_{kt,x}^2$. Note that variance heteroscedasticity is still present even if $\sigma_{k,x}^2 = \sigma_x^2$ for $k = 1, 2, \ldots, p$, due to the existence of squared explanatory variables $Z_{kt,x}^2$.

3.3 Credibility Regression Mortality Models

In this section, we propose a mortality modelling approach embedded, for the first time, in a credibility regression framework with random coefficients. The parameter estimation procedure is described and a special case with fixed coefficients is also provided.

3.3.1 A Credibility Regression Approach with Random Coefficients

Denote $D(t,x)$ as the observed number of deaths at age x in year t and $E(t,x)$ as the average population aged x during year t (also called as population exposure to risk). Then, age-specific mortality rates $m(t,x)$ are obtained by the ratio $D(t,x)/E(t,x)$ and one-year probabilities of death can be derived from the identity $q(t,x) = 1 - \exp[-m(t,x)]$, which is implied by the assumption of a constant force of mortality over each year of integer age and over each calendar year.

We assume that response variable $Y_{t,x}$ refers to an appropriate transform (log or logit) of a mortality measure $\left[m(t,x) \text{ or } q(t,x)\right]$ for age $x = x_0, \ldots, x_{k-1}$ of year $t = t_0, \ldots, t_{n-1}$, where variable x corresponds to consecutive integer ages (k in total) and t corresponds to consecutive calendar years (n in total). We also consider A_x as an age-related random

[1] Medical, biological, environmental or other factors that affect mortality evolution of each corresponding age over consecutive years are treated as unknown or exogenous due to the lack of specific data.

risk parameter, $\boldsymbol{Y}_x = (Y_{t_0,x}, Y_{t_1,x}, \ldots, Y_{t_{n-1},x})'$ as a mortality vector and $\boldsymbol{Z}_x$ as the design matrix of explanatory variables. We note that, in general, the design matrix could consist of various explanatory variables that reflect mortality characteristics. For instance, in a medical study, mortality may depend on various factors, such as the genetic background of an individual aged x, the life style, the nutrition, the toxicity of the environment, a possible infectious cause (bacteria, parasites, or fungi) or other socio-demographic factors that should affect mortality dynamics. Therefore, the pair that describes mortality evolution in age x is $(A_x, \boldsymbol{Y}_x)$, under the following assumptions:

(i) The pairs $(A_{x_0}, \boldsymbol{Y}_{x_0})$, $(A_{x_1}, \boldsymbol{Y}_{x_1})$, …, $(A_{x_{k-1}}, \boldsymbol{Y}_{x_{k-1}})$ are independent and $A_{x_0}, \ldots, A_{x_{k-1}}$ are independent and identically distributed.

(ii) $E(\boldsymbol{Y}_x | A_x) = \boldsymbol{Z}_x \boldsymbol{\beta}(A_x)$, where $\boldsymbol{Z}_x$ is a fixed $n \times p$ design matrix of full rank p $(< n)$ and $\boldsymbol{\beta}(A_x)$ is an unknown regression vector of length p.

(iii) $\mathrm{Cov}(\boldsymbol{Y}_x | A_x) = \mathrm{diag}\left[d_{t_0 t_0}(A_x), \ldots, d_{t_{n-1} t_{n-1}}(A_x)\right]$,
where $d_{tt}(A_x) = \sigma_1^2(A_x) + \sum_{k=2}^{p} \sigma_k^2(A_x)\, Z_{kt,x}^2$, with $\sigma_1^2(A_x) = \sigma_{01}^2(A_x) + \sigma_{11}^2(A_x)$,

or in matrix formulation as

$$\mathrm{Cov}(\boldsymbol{Y}_x | A_x) = \begin{pmatrix} \sigma_1^2(A_x) + \sum_{k=2}^{p} \sigma_k^2(A_x)\, Z_{kt_0,x}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_1^2(A_x) + \sum_{k=2}^{p} \sigma_k^2(A_x)\, Z_{kt_{n-1},x}^2 \end{pmatrix}.$$

The structural parameters are defined as follows

$$\begin{gathered} \boldsymbol{b} = E(\boldsymbol{\beta}(A_x)),\ \boldsymbol{\Phi} = \mathrm{Cov}[\boldsymbol{\beta}(A_x)],\ \boldsymbol{s}^2 = E[\boldsymbol{\sigma}^2(A_x)] = E\left[\left(\sigma_1^2(A_x), \ldots, \sigma_p^2(A_x)\right)'\right] \\ \text{and}\ \ \boldsymbol{\Delta}_x = E[\mathrm{Cov}(\boldsymbol{Y}_x | A_x)]. \end{gathered} \tag{3.7}$$

In such a regression setting, $\boldsymbol{\Delta}_x$ has to be estimated. Consequently, instead of the ordinary least squares method, regression coefficients are estimated with the generalised least squares method (*GLS*). Then, an individual estimator of $\boldsymbol{\beta}(A_x)$ can be obtained by

$$\widehat{\boldsymbol{\beta}}_x = (\boldsymbol{Z}_x' \boldsymbol{\Delta}_x^{-1} \boldsymbol{Z}_x)^{-1} \boldsymbol{Z}_x' \boldsymbol{\Delta}_x^{-1} \boldsymbol{Y}_x \text{ and } \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_x | A_x) = (\boldsymbol{Z}_x' \boldsymbol{\Delta}_x^{-1} \boldsymbol{Z}_x)^{-1}. \tag{3.8}$$

Theorem 3.2. *Under the above assumptions, the credibility estimator of* $\boldsymbol{\beta}(A_x)$ *is given by*

$$\boldsymbol{B}_x^{RC} = \boldsymbol{C}_x\, \widehat{\boldsymbol{\beta}}_x + (\boldsymbol{I} - \boldsymbol{C}_x)\, \boldsymbol{b}, \tag{3.9}$$

with

$$\boldsymbol{C}_x = \boldsymbol{\Phi}(\boldsymbol{\Xi}_x + \boldsymbol{\Phi})^{-1}, \tag{3.10}$$

where $\widehat{\boldsymbol{\beta}}_x$ *is given in* (3.8), $\boldsymbol{b}$ *and* $\boldsymbol{\Phi}$ *are defined in* (3.7), $\boldsymbol{\Xi}_x = E[Cov(\widehat{\boldsymbol{\beta}}_x|A_x)]$ *and* $\boldsymbol{I}$ *is the* $p \times p$ *identity matrix.*

Proof: The mean square error of (3.9) can be defined in terms of the norm $\|.\|_E^2$ as

$$\begin{aligned} Q &= \|\boldsymbol{\beta}(A_x) - \boldsymbol{B}_x^{RC}\|_E^2 \\ &= E\{[\boldsymbol{\beta}(A_x) - \boldsymbol{B}_x^{RC}]'[\boldsymbol{\beta}(A_x) - \boldsymbol{B}_x^{RC}]\} \qquad (3.11) \\ &= E\left[[\boldsymbol{\beta}^0(A_x)]'\boldsymbol{\beta}^0(A_x) + (\boldsymbol{\beta}_x^0)'\boldsymbol{C}_x'\,\boldsymbol{C}_x\,\boldsymbol{\beta}_x^0 - [\boldsymbol{\beta}^0(A_x)]'\,\boldsymbol{C}_x\,\boldsymbol{\beta}_x^0 - (\boldsymbol{C}_x\,\boldsymbol{\beta}_x^0)'\boldsymbol{\beta}^0(A_x)\right], \end{aligned}$$

where $\boldsymbol{\beta}^0(A_x) = \boldsymbol{\beta}(A_x) - \boldsymbol{b}$ and $\boldsymbol{\beta}_x^0 = \boldsymbol{\beta}_x - \boldsymbol{b}$. Using the product rule and differentiating (3.11) with respect to matrix $\boldsymbol{C}_x$, we have

$$\frac{\partial Q}{\partial(\boldsymbol{C}_x)} = -2E\left[\boldsymbol{\beta}^0(A_x)(\boldsymbol{\beta}_x^0)' - \boldsymbol{C}_x\,\boldsymbol{\beta}_x^0\,(\boldsymbol{\beta}_x^0)'\right]. \tag{3.12}$$

By substituting the values of $\boldsymbol{\beta}^0(A_x)$ and $\boldsymbol{\beta}_x^0$ and setting (3.12) equal to zero, we obtain

$$\begin{aligned} \boldsymbol{C}_x &= E\left\{[\boldsymbol{\beta}(A_x) - \boldsymbol{b}]\,[\boldsymbol{\beta}_x - \boldsymbol{b}]'\right\}\left\{E\left[(\boldsymbol{\beta}_x - \boldsymbol{b})(\boldsymbol{\beta}_x - \boldsymbol{b})'\right]\right\}^{-1} \\ &= \text{Cov}[\boldsymbol{\beta}(A_x), \boldsymbol{\beta}_x]\;[\text{Cov}(\boldsymbol{\beta}_x)]^{-1} \\ &= \{E\,\{\text{Cov}[\boldsymbol{\beta}(A_x), \boldsymbol{\beta}_x|A_x]\} + \text{Cov}\{E[\boldsymbol{\beta}(A_x)|A_x], E[\boldsymbol{\beta}_x|A_x]\}\}\;[\text{Cov}(\boldsymbol{\beta}_x)]^{-1} \\ &= \{0 + \text{Cov}[\boldsymbol{\beta}(A_x)]\}\,\{E[\text{Cov}(\boldsymbol{\beta}_x|A_x) + \text{Cov}[E(\boldsymbol{\beta}_x|A_x)]\}^{-1}, \end{aligned}$$

which yields (3.10). □

Then, the credibility estimator of future mortality rates $Y_{t_{n-1}+h,x}$, $h = 1, 2, \ldots, H$ may be compactly written as

$$\boldsymbol{Y}_x^{n+h} = \boldsymbol{Z}_x^{n+h}\,\boldsymbol{B}_x^{RC},$$

where $\boldsymbol{Z}_x^{n+h}$ denotes the design matrix of future periods.

3.3.2 Estimation of Structural Parameters

To estimate the structural parameters of the random coefficients credibility regression model, we can proceed similarly as in Hildreth and Houck (1968). Let $\boldsymbol{r}_x = (r_{t_0,x}, \ldots, r_{t_{n-1},x})'$ be the vector of the least squares residuals from the regression of $\boldsymbol{Y}_x$

on $\boldsymbol{Z}_x$ given A_x, which is obtained by

$$\boldsymbol{r}_x = \boldsymbol{Y}_x - \boldsymbol{Z}_x\widehat{\boldsymbol{\beta}}_x = \boldsymbol{M}_x\boldsymbol{v}_x, \tag{3.13}$$

where $\widehat{\boldsymbol{\beta}}_x = (\boldsymbol{Z}_x^{'}\boldsymbol{Z}_x)^{-1}\boldsymbol{Z}_x^{'}\boldsymbol{Y}_x$ is the least squares estimator of coefficients in ordinary regression, $\boldsymbol{M}_x = \boldsymbol{I} - \boldsymbol{Z}_x(\boldsymbol{Z}_x^{'}\boldsymbol{Z}_x)^{-1}\boldsymbol{Z}_x^{'}$ is a symmetric and idempotent matrix of order $n \times n$ and $\boldsymbol{v}_x = \boldsymbol{Y}_x - \boldsymbol{Z}_x\,\boldsymbol{\beta}(A_x)$ is the error term. Then, given A_x, the variance matrix of $\boldsymbol{r}_x$, via (3.6), becomes

$$E(\boldsymbol{r}_x\boldsymbol{r}_x^{'}|A_x) = E(\boldsymbol{M}_x\boldsymbol{v}_x\boldsymbol{v}_x^{'}\boldsymbol{M}_x|A_x), \tag{3.14}$$

from which we can get

$$E(\dot{\boldsymbol{r}}_x|A_x) = \dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x\boldsymbol{\sigma}^2(A_x), \tag{3.15}$$

where $\dot{\boldsymbol{r}}_x = (r_{t_0,x}^2, \ldots, r_{t_{n-1},x}^2)^{'}$, $\dot{\boldsymbol{M}}_x = \{m_{ts,x}^2\}_{t,s=t_0,\ldots,t_{n-1}}$ and $\dot{\boldsymbol{Z}}_x = \{Z_{kt,x}^2\}_{k=1,\ldots,p,\; t=t_0,\ldots,t_{n-1}}$ are the Hadamard products of matrices $\boldsymbol{r}_x$, $\boldsymbol{M}_x$ and $\boldsymbol{Z}_x$, respectively, while $\boldsymbol{\sigma}^2(A_x)$ is as defined in (3.7). In addition, (3.15) implies that, for given A_x, least squares residuals $\dot{\boldsymbol{r}}_x$ are regressed on $\boldsymbol{\sigma}^2(A_x)$, which yields

$$\dot{\boldsymbol{r}}_x = \dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x\boldsymbol{\sigma}^2(A_x) + \boldsymbol{e}_x = \boldsymbol{G}_x\boldsymbol{\sigma}^2(A_x) + \boldsymbol{e}_x, \tag{3.16}$$

where $\boldsymbol{G}_x = \dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x$ and $\boldsymbol{e}_x$ is a $n \times 1$ disturbance vector, such that $E(\boldsymbol{e}_x|A_x) = 0$. Hence, its variance-covariance matrix is given by

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{e}_x|A_x) &= E\{[\dot{\boldsymbol{r}}_x - E(\dot{\boldsymbol{r}}_x|A_x)][\dot{\boldsymbol{r}}_x - E(\dot{\boldsymbol{r}}_x|A_x)]^{'}|A_x\} \\ &= E(\dot{\boldsymbol{r}}_x|A_x)[E(\dot{\boldsymbol{r}}_x|A_x)]^{'} + 2E(\boldsymbol{r}_x\boldsymbol{r}_x^{'}|A_x) * E(\boldsymbol{r}_x\boldsymbol{r}_x^{'}|A_x) - E(\dot{\boldsymbol{r}}_x|A_x)[E(\dot{\boldsymbol{r}}_x|A_x)]^{'} \\ &= 2\dot{\boldsymbol{\Psi}}_x, \end{aligned} \tag{3.17}$$

where $\dot{\boldsymbol{\Psi}}_x$ represents the Hadamard product of matrix $\boldsymbol{\Psi}_x$ by itself, with

$$\boldsymbol{\Psi}_x = E(\boldsymbol{r}_x\boldsymbol{r}_x^{'}|A_x) = E(\boldsymbol{M}_x\boldsymbol{v}_x(\boldsymbol{M}_x\boldsymbol{v}_x)^{'}|A_x) = \boldsymbol{M}_xE(\boldsymbol{v}_x\boldsymbol{v}_x^{'}|A_x)\boldsymbol{M}_x = \boldsymbol{M}_x\boldsymbol{\Delta}_x\boldsymbol{M}_x.$$

Then, if σ_k^2s are known, the *GLS* estimator of $\boldsymbol{\sigma}^2(A_x)$ in (3.16) is obtained by minimising the criterion function $[\dot{\boldsymbol{r}}_x - \boldsymbol{G}_x\boldsymbol{\sigma}^2(A_x)]^{'}(2\dot{\boldsymbol{\Psi}}_x)^{-1}[\dot{\boldsymbol{r}}_x - \boldsymbol{G}_x\boldsymbol{\sigma}^2(A_x)]$, which gives

$$\widehat{\boldsymbol{\sigma}}_x^2 = (\boldsymbol{G}_x^{'}\dot{\boldsymbol{\Psi}}_x^{-1}\boldsymbol{G}_x)^{-1}\boldsymbol{G}_x^{'}\dot{\boldsymbol{\Psi}}_x^{-1}\dot{\boldsymbol{r}}_x\,. \tag{3.18}$$

However, estimators of $\boldsymbol{\beta}(A_x)$ in (3.8) and $\boldsymbol{\sigma}^2(A_x)$ in (3.18) are non-operational, because the variance-covariance matrices $\boldsymbol{\Delta}_x$ and $2\dot{\boldsymbol{\Psi}}_x$ are functions of unknown variances. Therefore, operational estimators of $\boldsymbol{\beta}(A_x)$ and $\boldsymbol{\sigma}^2(A_x)$ can be obtained by replacing unknown matrices with estimators $\widehat{\boldsymbol{\Delta}}_x$ and $\widehat{2\dot{\boldsymbol{\Psi}}}_x$, respectively. A least squares estimator of the unknown variances $\boldsymbol{\sigma}^2(A_x)$ is directly obtained from (3.16) as follows

$$\begin{aligned}\widehat{\boldsymbol{\sigma}}_x^2 &= (\boldsymbol{G}_x'\boldsymbol{G}_x)^{-1}\boldsymbol{G}_x'\dot{\boldsymbol{r}}_x \\ &= [(\dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x)'(\dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x)]^{-1}(\dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x)'\dot{\boldsymbol{r}}_x \\ &= (\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{M}}_x^2\dot{\boldsymbol{Z}}_x)^{-1}\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{M}}_x\dot{\boldsymbol{r}}_x\,,\end{aligned} \tag{3.19}$$

where equality $\dot{\boldsymbol{M}}_x' = \dot{\boldsymbol{M}}_x$ holds true, since $(\boldsymbol{M}_x * \boldsymbol{M}_x)' = \boldsymbol{M}_x * \boldsymbol{M}_x$ for a symmetric matrix $\boldsymbol{M}_x$.

Remark 3.3. In the actuarial literature, there are many other types of estimators for variance in (3.16). For instance, Hildreth and Houck (1968) suggested the unbiased estimator $\widehat{\boldsymbol{\sigma}}_x^{2\,(alt1)} = (\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x)^{-1}\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{r}}_x$, while Rao (1973) proposed the so-called "Minimum Norm Quadratic Unbiased Estimator" (MINQUE), which is given by $\widehat{\boldsymbol{\sigma}}_x^{2\,(alt2)} = (\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{M}}_x\dot{\boldsymbol{Z}}_x)^{-1}\dot{\boldsymbol{Z}}_x'\dot{\boldsymbol{M}}_x\dot{\boldsymbol{r}}_x$.

The random coefficients (RC) credibility estimator of $\boldsymbol{\beta}(A_x)$, which is denoted as $\widehat{\boldsymbol{B}}_x^{RC} = (\widehat{B}_{1x}^{RC},\ldots,\widehat{B}_{px}^{RC})'$, is given by

$$\widehat{\boldsymbol{B}}_x^{RC} = \widehat{\boldsymbol{C}}_x\,\widehat{\widehat{\beta}}_x + (\boldsymbol{I}-\widehat{\boldsymbol{C}}_x)\,\widehat{\boldsymbol{b}}\,, \tag{3.20}$$

where

$$\widehat{\widehat{\beta}}_x = (\boldsymbol{Z}_x'\widehat{\boldsymbol{\Delta}}_x^{-1}\boldsymbol{Z}_x)^{-1}\boldsymbol{Z}_x'\widehat{\boldsymbol{\Delta}}_x^{-1}\boldsymbol{Y}_x$$

and $\widehat{\boldsymbol{\Delta}}_x = \operatorname{diag}\left(\widehat{\delta}_{t_0t_0}^x,\ldots,\widehat{\delta}_{t_{n-1},t_{n-1}}^x\right)$, with $\widehat{\delta}_{tt}^x = \widehat{s}_1^2 + \sum_{k=2}^{p}\widehat{s}_k^2 Z_{kt,x}^2$, $t = t_0,\ldots,t_{n-1}$, obtained according to (3.7), by using the mean of the estimated variances in (3.19). Future mortality estimates follow from

$$\widehat{\boldsymbol{Y}}_x^{n+h} = \boldsymbol{Z}_x^{n+h}\widehat{\boldsymbol{B}}_x^{RC} = \boldsymbol{Z}_x^{n+h}\widehat{\boldsymbol{C}}_x\,\widehat{\widehat{\beta}}_x + \boldsymbol{Z}_x^{n+h}(\boldsymbol{I}-\widehat{\boldsymbol{C}}_x)\,\widehat{\boldsymbol{b}}, \quad h = 1,2,\ldots,H, \tag{3.21}$$

where $\widehat{\boldsymbol{C}}_x = \widehat{\boldsymbol{\Phi}}(\widehat{\boldsymbol{\Xi}}_x + \widehat{\boldsymbol{\Phi}})^{-1}$, $x = x_0,\ldots,x_{k-1}$, is the corresponding credibility factor. We suggest the following estimators for parameters $\boldsymbol{b}$, $\boldsymbol{\Xi}_x$ and $\boldsymbol{\Phi}$ to obtain De Vylder's (1978) optimality (minimum variance within the class of unbiased estimators)

$$\widehat{\boldsymbol{b}} = (\sum_{x=x_0}^{x_{k-1}}\widehat{\boldsymbol{C}}_x)^{-1}\sum_{x=x_0}^{x_{k-1}}\widehat{\boldsymbol{C}}_x\,\widehat{\widehat{\beta}}_x\,, \tag{3.22}$$

$$\widehat{\boldsymbol{\Xi}}_x = \frac{1}{x_{k-1}-x_0+1}\sum_{x'=x_0}^{x_{k-1}}(\boldsymbol{Z}_x'\widehat{\boldsymbol{\Delta}}_{x'}^{-1}\boldsymbol{Z}_x)^{-1} \tag{3.23}$$

and

$$\widehat{\boldsymbol{\Phi}} = \frac{1}{x_{k-1} - x_0} \sum_{x=x_0}^{x_{k-1}} \widehat{\boldsymbol{C}}_x (\widehat{\widehat{\beta}}_x - \widehat{\boldsymbol{b}})(\widehat{\widehat{\beta}}_x - \widehat{\boldsymbol{b}})' \,. \tag{3.24}$$

Note that the estimators of $\widehat{\boldsymbol{\Phi}}$ and $\widehat{\boldsymbol{b}}$ are implicit functions of the parameter to be estimated and should be calculated iteratively, by imposing $(\widehat{\boldsymbol{\Phi}} + \widehat{\boldsymbol{\Phi}}^{'})/2 = 0$ to retain symmetry after each iteration.

3.3.3 Credibility Regression with Fixed Coefficients as Special Case

In the case of fixed regression's coefficients, the previous model reduces to a special case of Hachemeister's (1975) model with no weighs, i.e., $\boldsymbol{W}_x = \boldsymbol{I}$. In particular, some weights may appear in each regression line of A_x. For instance, population exposures $E(t,x)$, for $t = t_0, \ldots, t_{n-1}$ can be used as weights. In this case, we have the standard regression case of Hachemeister's model. To proceed, we follow the same Assumptions (i) and (ii) as in the random coefficients case, but covariance matrix in Assumption (iii) is simplified to $\mathrm{Cov}(\boldsymbol{Y}_x|A_x) = \sigma^2(A_x)\boldsymbol{W}_x$, where $\boldsymbol{W}_x$ is a fixed $n \times n$ positive definite diagonal matrix, with weights $\boldsymbol{W}_x = \mathrm{diag}\,[E(t_0,x), \ldots, E(t_{n-1},x)]$. The structural parameters are now defined as

$$\boldsymbol{b} = E[\boldsymbol{\beta}(A_x)],\ \boldsymbol{U} = \mathrm{Cov}[\boldsymbol{\beta}(A_x)] \text{ and } s^2 = E[\sigma^2(A_x)] \tag{3.25}$$

and the ordinary least squares estimator of the coefficients vector $\boldsymbol{\beta}(A_x)$ is given by

$$\widehat{\boldsymbol{\beta}}_x = (\boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Z}_x)^{-1} \boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Y}_x \,. \tag{3.26}$$

The variance-covariance matrix is obtained by $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_x|A_x) = \sigma^2(A_x)(\boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Z}_x)^{-1}$, while its expected value is given by

$$E[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_x|A_x)] = E[\sigma^2(A_x)(\boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Z}_x)^{-1}] = s^2(\boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Z}_x)^{-1}. \tag{3.27}$$

Based on the above assumptions, the credibility estimator $\widehat{\boldsymbol{B}}_x^{FC} = (\widehat{B}_{1x}^{FC}, \ldots, \widehat{B}_{px}^{FC})'$ of $\boldsymbol{\beta}(A_x)$ for the fixed coefficients (FC) model is given by

$$\widehat{\boldsymbol{B}}_x^{FC} = \widehat{\boldsymbol{K}}_x \, \widehat{\boldsymbol{\beta}}_x + (\boldsymbol{I} - \widehat{\boldsymbol{K}}_x) \, \widehat{\boldsymbol{b}} \,, \tag{3.28}$$

where $\widehat{\boldsymbol{K}}_x = \widehat{\boldsymbol{U}}[\widehat{s^2}(\boldsymbol{Z}_x' \boldsymbol{W}_x^{-1} \boldsymbol{Z}_x)^{-1} + \widehat{\boldsymbol{U}}]^{-1}$ is the estimated credibility factor. Similarly, for the derivation of (3.28), we refer to Bühlmann and Gisler (2005). To recapture De Vylder's (1978) optimality, we use the following estimators

$$\widehat{s^2} = \frac{1}{(x_{k-1} - x_0 + 1)(t_{n-1} - t_0 + 1 - p)} \sum_{x=x_0}^{x_{k-1}} (\boldsymbol{Y}_x - \boldsymbol{Z}_x \widehat{\boldsymbol{\beta}}_x)' \boldsymbol{W}_x^{-1} (\boldsymbol{Y}_x - \boldsymbol{Z}_x \widehat{\boldsymbol{\beta}}_x) \,, \tag{3.29}$$

$$\widehat{\boldsymbol{U}} = \frac{1}{x_{k-1} - x_0} \sum_{x=x_0}^{x_{k-1}} \widehat{\boldsymbol{K}}_x (\widehat{\boldsymbol{\beta}}_x - \widehat{\boldsymbol{b}})(\widehat{\boldsymbol{\beta}}_x - \widehat{\boldsymbol{b}})', \tag{3.30}$$

$$\widehat{\boldsymbol{b}} = (\sum_{x=x_0}^{x_{k-1}} \widehat{\boldsymbol{K}}_x)^{-1} \sum_{x=x_0}^{x_{k-1}} \widehat{\boldsymbol{K}}_x \, \widehat{\boldsymbol{\beta}}_x \,. \tag{3.31}$$

Again, the estimators of $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{b}}$ should be calculated iteratively, imposing $(\widehat{\boldsymbol{U}} + \widehat{\boldsymbol{U}}')/2 = 0$ after each iteration.

3.4 Extrapolation Methods for Estimating Future Mortality Rates

In this section, we fit the random coefficients (RC) and the fixed coefficients (FC) credibility regression models to mortality rates for age $x = x_0, \ldots, x_{k-1}$ of year $t = t_0, \ldots, t_{n-1}$. For both models, the fitted rates up to year t_{n-1} can be compactly written as $\widehat{\boldsymbol{Y}}_x = \boldsymbol{Z}_x \widehat{\boldsymbol{\beta}}_x$. As we noted before, design matrix $\boldsymbol{Z}_x$ could consist of various independent variables that reflect risk factors for any given age x, but due to lack of specific data, we assume that $\boldsymbol{Y}_x$s for each given age x, depend only on the period effects of each calendar year, i.e., $\boldsymbol{Z}_x = \boldsymbol{Z}$. However, if specific data are available, for instance in case of life insurance datasets, then more explanatory variables can be incorporated in the regression model. Henceforth, we consider the same design matrix for all $\boldsymbol{Y}_x$s, i.e.,

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 2 & \ldots & n \end{pmatrix}' .$$

3.4.1 Standard Extrapolation Method (SEM)

Based on current fitting data of the response variable $\widehat{\boldsymbol{Y}_x} = (Y_{t_0,x}, Y_{t_1,x}, \ldots, Y_{t_{n-1},x})'$, mortality rates for one-year ahead are estimated by

$$\widehat{Y}_{t_{n-1}+1,x} = \widehat{B}^c_{1x} + \widehat{B}^c_{2x} \, (t_{n-1} - t_0 + 2), \tag{3.32}$$

where $c =$ RC or FC. Similarly, estimates of future mortality rates for age $x = x_0, \ldots, x_{k-1}$ are given by extrapolating one-year ahead estimates of (3.32) to $\widehat{Y}_{t_{n-1}+h,x} = \widehat{B}^c_{1x} + \widehat{B}^c_{2x} \, (t_{n-1} - t_0 + 1 + h)$, for $h = 2, 3, \ldots, H$, where the credibility estimators $\widehat{\boldsymbol{B}}^c_x = (\widehat{B}^c_{1x}, \widehat{B}^c_{2x})'$ are obtained by (3.20) for the RC or (3.28) for the FC model. Hence, under this method, future estimates are based on the mortality data of the initial fitting span $[t_0, t_{n-1}]$.

3.4.2 Other Extrapolation Methods

In practice, two additional methods can also be used to extrapolate mortality rates over a given forecasting horizon $h = 1, 2, \ldots, H$. Thus, for each one of the RC and

FC models, one-year ahead estimates $\widehat{Y}_{t_{n-1}+1,x}$ can be embedded to the existing fitting span, with $Y_{t_0,x}$ simultaneously excluded from it, so that the fitting year span is moved forward by one year each time to $[t_1, t_{n-1}+1]$, $[t_2, t_{n-1}+2]$, $[t_3, t_{n-1}+3]$, ... Then, after repeating the estimation procedure, we can consecutively obtain $\widehat{Y}_{t_{n-1}+2,x}$, $\widehat{Y}_{t_{n-1}+3,x}, \widehat{Y}_{t_{n-1}+4,x}, \ldots, \widehat{Y}_{t_{n-1}+H,x}$. Under this "moving extrapolation method (MEM)", future estimates are based on more recent mortality trends.

Alternatively, one-year ahead estimates $\widehat{Y}_{t_{n-1}+1,x}$ can be embedded to the existing fitting span, without removing $Y_{t_0,x}$, so that the fitting year span is extended by one year each time to $[t_0, t_{n-1}+1]$, $[t_0, t_{n-1}+2]$, $[t_0, t_{n-1}+3]$, ... Hence, in each estimation step, credibility regression models are fitted on a continuously extended response variable, to obtain $\widehat{Y}_{t_{n-1}+2,x}$, $\widehat{Y}_{t_{n-1}+3,x}$, $\widehat{Y}_{t_{n-1}+4,x}, \ldots, \widehat{Y}_{t_{n-1}+H,x}$. Under this "extended extrapolation method (EEM)", future mortality trends are based on both the initial mortality rates and the recent ones that have been obtained after each estimation step. Similar practical approaches have also been adopted by Luan (2015). The numerical results in the following section justify that all methods can be efficiently applied in actuarial practice.

3.5 Empirical Illustration

In this section, the Lee–Carter (LC), the Cairns–Blake–Dowd (CBD) and the credibility regression models are fitted on Greek mortality data. Then, forecasting results are evaluated using the mean absolute forecast error (MAFE) and the root mean of squared forecast error (RMSFE) measures. Greek data have a limited number of historical mortality observations (1981–2013), which are available on the Human Mortality Database (2017), structured by year, age and gender.

Furthermore, in life insurance datasets similar limitations frequently exist. Credibility regression can efficiently capture the underlying data trends, especially in cases where there is limited mortality experience for a specific age, but extensive experience for the entire age range (the case of Greek data). Of course, credibility regression methods can also be used for larger datasets.

Mortality evolution for the period 1981–2010 in Greece is illustrated in Figures 3.1 and 3.2 for $\log m(t,x)$ and $\operatorname{logit} q(t,x)$, respectively. Both mortality measures show a downward trend for discrete ages $x = 40,\ 60,\ 80$ of males (left panels of Figures 3.1 and 3.2) and females (middle panels of Figures 3.1 and 3.2). In addition, for both genders, average mortality decline shows a clear downward trend over time (right panels of 3.1 and 3.2).

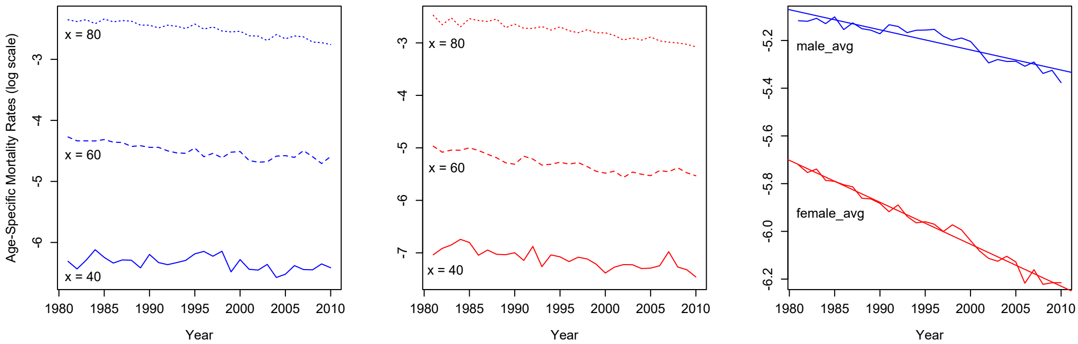


Figure 3.1 Observed $\log m(t,x)$ of the period 1981–2010 in Greece, for males (left) and females (middle) at the age of 40, 60 and 80. Average male and female $\log m(t,x)$ values over ages 15–84 are illustrated in (right), where straight lines show the corresponding trends in mortality decline.

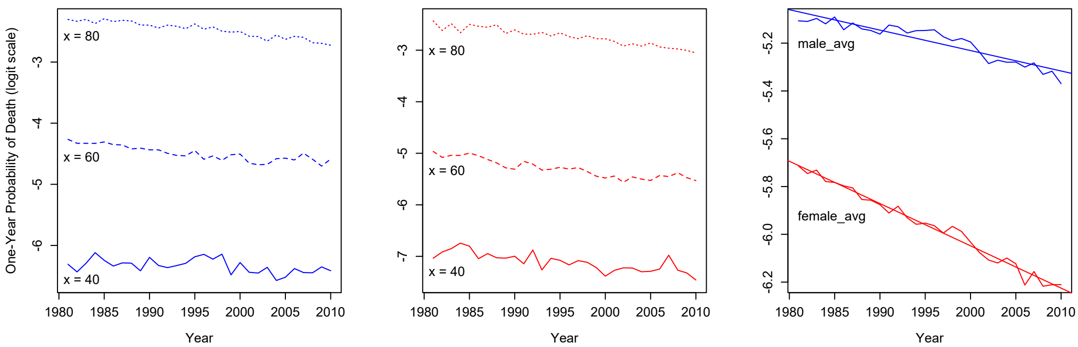


Figure 3.2 Observed $\text{logit}\, q(t,x)$ of the period 1981–2010 in Greece, for males (left) and females (middle) at the age of 40, 60 and 80. Average male and female $\text{logit}\, q(t,x)$ values over ages 15–84 are illustrated in (right), where straight lines show the corresponding trends in mortality decline.

3.5.1 Forecasting Results

For the numerical illustration that follows, we used the empirical age-specific mortality rates $m(t,x)$ from 1981 to 2010, for males and females at the ages of 15 to 84. This age span choice is in accordance with similar studies (Tsai and Lin 2017a, 2017b) as it corresponds to the age of a young adult up to the overall level of life expectancy in developed countries. To ensure robustness, relative to changes in the fitting range of data, we used two age and three period spans to extract forecasts for a 10-year $(H = 10)$ forecasting horizon, presented in Table 3.1. In particular, for the FC model, we used $\boldsymbol{W}_x = \boldsymbol{I}$ as weights. The credibility regression mortality methods, as well as the LC and the CBD mortality models, were developed into the R (R Core Team, 2017) statistical software, using our own routines. Exceptionally, for the Poisson LC and CBD fitting methods, we used the "LifeMetrics" R package[2].

[2]This software, which is not part of CRAN, is available from http://www.macs.hw.ac.uk/~andrewc/lifemetrics/.

Table 3.1 Selected fitting and forecasting periods.

Fitting Ages	Fitting Period	Forecasting Period
$[x_0, x_{k-1}]$	$[t_0, t_{n-1}]$	$[t_{n-1}+1, t_{n-1}+H]$
$[15, 84]$	$[1981, 2000]$	$[2001, 2010]$
$[15, 84]$	$[1986, 2000]$	$[2001, 2010]$
$[15, 84]$	$[1991, 2000]$	$[2001, 2010]$
$[55, 84]$	$[1981, 2000]$	$[2001, 2010]$
$[55, 84]$	$[1986, 2000]$	$[2001, 2010]$
$[55, 84]$	$[1991, 2000]$	$[2001, 2010]$

To retain linearity over each corresponding fitting period, the logarithmic transform $Y_{t,x} = \log m(t,x)$ was used for the age-specific mortality rates and the logit transform $Y_{t,x} = \operatorname{logit} q(t,x) = \log \dfrac{q(t,x)}{1-q(t,x)}$ for the one-year probabilities of death. Forecast errors were then evaluated over the 10-year forecasting horizon using MAFE and RMSFE measures[3], where smaller values indicate a better forecasting performance. Averaged (avg) MAFE and RMSFE values are obtained by using

$$MAFE_{avg} = \frac{1}{H \times (x_{k-1} - x_0 + 1)} \sum_{h=1}^{H} \sum_{x=x_0}^{x_{k-1}} |\widehat{m}(t_{n-1}+h,x) - m(t_{n-1}+h,x)| \times 100 \tag{3.33}$$

and

$$RMSFE_{avg} = \sqrt{\frac{1}{H \times (x_{k-1} - x_0 + 1)} \sum_{h=1}^{H} \sum_{x=x_0}^{x_{k-1}} \left[\widehat{m}(t_{n-1}+h,x) - m(t_{n-1}+h,x)\right]^2} \times 100\,. \tag{3.34}$$

Similarly, in the case of using $Y_{t,x} = \operatorname{logit} q(t,x)$ as response variable, $m(t,x)$ should be replaced by $q(t,x)$ in above formulas. Forecast accuracy results at percentage (%) scales are evaluated over the period $[2001, 2010]$. MAFE and RMSFE values for fitting ages $[15, 84]$, using $Y_{t,x} = \log m(t,x)$ are illustrated in Table 3.2 (a) and (b), respectively, while the corresponding values for ages $[55, 84]$ with $Y_{t,x} = \operatorname{logit} q(t,x)$ are presented in Table 3.3 (a) and (b), respectively. Note that CBD model is included only for comparisons in fitting ages $[55, 84]$, as it has been particularly designed for higher ages.

[3]For instance, use of MAFE is demonstrated in the modelling comparison study of Shang et al. (2011), while RMSFE in Hansen (2013) and Van Berkum et al. (2016).

Table 3.2 MAFE and RMSFE values of forecasts over the period $[2001, 2010]$ for ages $[15, 84]$.

(a) MAFE Values

$MAFE_{[15,84]}$		Lee–Carter		Random Coefficients (RC)			Fixed Coefficients (FC)		
Fitting Period	Gender	LC	LC-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
$[1981, 2000]$	Male	0.1513	0.1569	0.1338	0.1205	0.1322	0.1352	0.1256	0.1361
	Female	0.0831	0.0861	0.0702	0.0740	0.0711	0.0691	0.0657	0.0690
$[1986, 2000]$	Male	0.1684	0.1514	0.1175	0.1196	0.1158	0.1203	0.1221	0.1206
	Female	0.0625	0.0799	0.0650	0.0696	0.0758	0.0608	0.0651	0.0613
$[1991, 2000]$	Male	0.1468	0.1681	0.1275	0.1257	0.1280	0.1288	0.1289	0.1289
	Female	0.0763	0.0959	0.0705	0.0678	0.0750	0.0622	0.0663	0.0669
Average		0.1147(7)	0.1231(8)	0.0974(5)	0.0962(3)	0.0997(6)	0.0961(2)	0.0956(1)	0.0971(4)

(b) RMSFE Values

$RMSFE_{[15,84]}$		Lee–Carter		Random Coefficients (RC)			Fixed Coefficients (FC)		
Fitting Period	Gender	LC	LC-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
$[1981, 2000]$	Male	0.3165	0.3220	0.2661	0.2349	0.2629	0.2716	0.2511	0.2745
	Female	0.1791	0.1825	0.1398	0.1594	0.1457	0.1376	0.1365	0.1374
$[1986, 2000]$	Male	0.3543	0.3200	0.2257	0.2265	0.2204	0.2362	0.2364	0.2375
	Female	0.1307	0.1742	0.1410	0.1509	0.1700	0.1264	0.1385	0.1288
$[1991, 2000]$	Male	0.3180	0.4010	0.2478	0.2457	0.2470	0.2570	0.2551	0.2516
	Female	0.1694	0.2415	0.1580	0.1511	0.1707	0.1302	0.1438	0.1476
Average		0.2447(7)	0.2735(8)	0.1964(4)	0.1948(3)	0.2028(6)	0.1932(1)	0.1936(2)	0.1962(5)

For both genders, accuracy results for ages $[15, 84]$ and $[55, 84]$ indicate that, for each fitting period, credibility regression methods outperform the LC and CBD models for both error measures. Average values over the whole period are given in the last rows of each measure's subtable.

More precisely, for ages $[15, 84]$, the FC-MEM and FC-SEM produce the smallest average MAFE and RMSFE, while for ages $[55, 84]$, RC-MEM performs better in average under both measures, which indicates that forecasts for higher ages are based on more recent mortality trends. Moreover, we observe that errors are getting evidently larger, when shortening the age fitting span to $[55, 84]$. This is due to the fact that both $|\widehat{m}(t_{n-1}+h,x) - m(t_{n-1}+h,x)|$ in (3.33) and $\left[\widehat{m}(t_{n-1}+h,x) - m(t_{n-1}+h,x)\right]^2$ in (3.34) are generally increasing with age x. Therefore, $MAFE_{avg}$ and $RMSFE_{avg}$ for ages $[55, 84]$ are larger than those for $[15, 84]$.

We note that, for our comparison, we used the Lee–Carter (1992) and Cairns–Blake–Dowd (2006) models, which incorporate only age and period effects. Models with cohort parameters were intentionally excluded from our analysis to be consistent with the age-period structure of the proposed credibility regression methods that model the period dynamics of mortality across the ages.

Table 3.3 MAFE and RMSFE values of forecasts over the period $[2001, 2010]$ for ages $[55, 84]$.

(a) MAFE Values

$MAFE_{[55,84]}$		Mortality Models				Random Coefficients (RC)			Fixed Coefficients (FC)		
Fitting Period	Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
$[1981, 2000]$	Male	0.3191	0.3322	0.2924	0.3247	0.2885	0.2642	0.2846	0.2871	0.2673	0.2870
	Female	0.1884	0.1933	0.1694	0.1884	0.1624	0.1458	0.1611	0.1629	0.1448	0.1627
$[1986, 2000]$	Male	0.2928	0.3186	0.2682	0.2988	0.2506	0.2547	0.2494	0.2544	0.2581	0.2541
	Female	0.1577	0.1769	0.1618	0.1708	0.1287	0.1377	0.1344	0.1289	0.1351	0.1288
$[1991, 2000]$	Male	0.3091	0.3622	0.2790	0.3348	0.2483	0.2461	0.2464	0.2538	0.2493	0.2525
	Female	0.1723	0.2126	0.1659	0.1868	0.1324	0.1350	0.1363	0.1363	0.1382	0.1361
Average		0.2399(8)	0.2660(10)	0.2228(7)	0.2507(9)	0.2018(3)	0.1973(1)	0.2020(4)	0.2039(6)	0.1988(2)	0.2035(5)

(b) RMSFE Values

$RMSFE_{[55,84]}$		Mortality Models				Random Coefficients (RC)			Fixed Coefficients (FC)		
Fitting Period	Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
$[1981, 2000]$	Male	0.4616	0.4848	0.3904	0.4467	0.4041	0.3644	0.3963	0.4065	0.3786	0.4061
	Female	0.2795	0.2842	0.2221	0.2512	0.2260	0.1996	0.2213	0.2304	0.2010	0.2299
$[1986, 2000]$	Male	0.4320	0.4872	0.3551	0.4073	0.3506	0.3522	0.3419	0.3631	0.3653	0.3618
	Female	0.2340	0.2699	0.2165	0.2244	0.1805	0.1940	0.1895	0.1803	0.1897	0.1800
$[1991, 2000]$	Male	0.4671	0.6129	0.3698	0.4625	0.3484	0.3423	0.3389	0.3660	0.3501	0.3616
	Female	0.2652	0.3721	0.2202	0.2510	0.1866	0.1888	0.1912	0.1961	0.1930	0.1954
Average		0.3566(9)	0.4185(10)	0.2957(7)	0.3405(8)	0.2827(4)	0.2736(1)	0.2799(3)	0.2904(6)	0.2796(2)	0.2891(5)

3.5.2 Credibility Effects on Mortality Modelling

In the preceding section, we used the proposed credibility regression methods to estimate the actual mortality trend for a specific age, by weighting the mortality trend for this age and the mean trend over a wider group of ages that encompasses much more information. Figure 3.3 illustrates the downward trend of the actual (observed) $\operatorname{logit} q(t,x)$ for Greek males (left panel) and females (right panel), aged 55, 65 and 75 over the period 1981–2010.

The intuition behind using credibility regression is that the proposed methods could potentially lead us to more accurate estimates for the intercept and the slope of the mortality curve for a given age $x = x_0, \ldots, x_{k-1}$. To assure this, we used the absolute forecast errors by age (AFE_x) to compare the trend (intercept and slope) of the $\operatorname{logit} q(2000+h,x)$, $h = 1, \ldots, 10$ between the actual rates and the rates produced from the best performing models for both genders over years $[2001, 2010]$, with and without credibility, for pension ages $[65, 84]$, fitted for $[1981, 2000]$. For each model, AFE_x can be obtained by $AFE_x = |\operatorname{logit}\widehat{q}(2000+h,x) - \operatorname{logit} q(2000+h,x)| \times 100$.

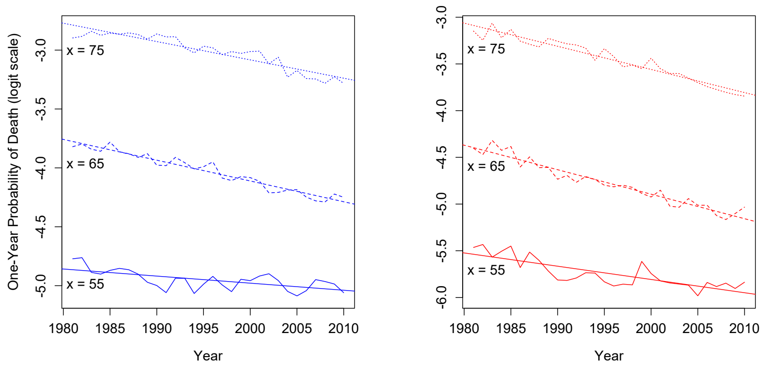


Figure 3.3 Observed logit $q(t,x)$ of the period 1981–2010 in Greece, for males (left) and females (right) at the age of 55, 65 and 75.

Figure 3.4 displays the AFE_x comparison results, which indicate that, almost for all ages, credibility regression methods (dot lines) perform better than the LC (solid lines) and CBD (dashed lines) models. An alternative way to see how close the credibility forecasts are to the the actual mortality trend, Figure 3.5 illustrates the intercept and the slope of the actual rates and the forecasted ones for some ages, under the best performing methods (based on AFE$_x$) with credibility (FC-MEM for males and RC-MEM for females) and without credibility (LC, CBD).

The trend lines for the RC-MEM and FC-MEM forecasts can be easily extracted using the ordinary least squares method. Recall that, the intercept and the slope for the LC and CBD models is given by Equations (3.2) and (3.4), respectively (Remark 3.1), while for the credibility method RC by (3.20) and for FC by (3.28). The illustrated results in Figure 3.5 indicate that intercepts and slopes of the FC-MEM (for males) and RC-MEM (for females) lines are closer to the actual ones, which set the best starting point for the forecasts.

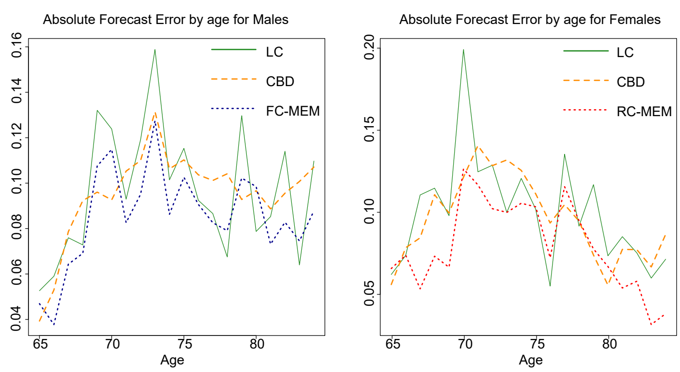
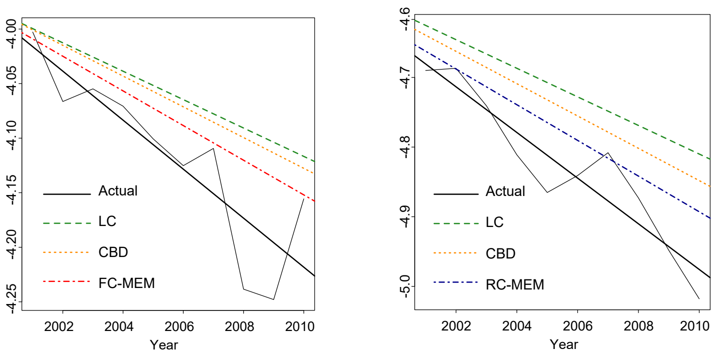


Figure 3.4 AFE values against age of logit $q(2000+h,x)$, $h=1,\ldots,10$ between the actual rates and the rates produced from the best performing models with and without credibility for males (left) and females (right) over $[2001,2010]$, fitted to pension ages $[65,84]$ for years $[1981,2000]$.



Trend / Model	Intercept	Slope	Trend / Model	Intercept	Slope
Actual	−4.0028	−0.0225	Actual	−4.6515	−0.0328
LC	−3.9957	−0.0130	LC	−4.6084	−0.0205
CBD	−3.9959	−0.0141	CBD	−4.6198	−0.0232
FC-MEM	−4.0023	−0.0159	RC-MEM	−4.6403	−0.0256

Figure 3.5 Intercept and slope estimates of logit $q(2000+h,x)$ for $h=1,\ldots,10$ and ages $x=66$ for males and $x=67$ for females, with credibility (dot-dashed lines for FC-MEM and RC-MEM) and without credibility (dashed lines for LC and dot lines for CBD). Solid lines show the actual mortality and its trend.

3.5.3 Comparison with the Bühlmann Credibility Approach

Tsai and Lin (2017a) proposed a Bühlmann credibility approach to forecast mortality rates for both genders in Japan, the United Kingdom and the United States. This model can be directly obtained from the more general regression model, presented in Section 3.3.3, if we set $\boldsymbol{Z}_x = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}'$ and $\boldsymbol{W}_x = \boldsymbol{I}$ for $x = x_0, \ldots, x_{k-1}$. Then, from (3.26), $\boldsymbol{\beta}_x$ is equal to $\overline{Y}_x$ and the model parameters, which are scalars now, can be estimated by

$$\widehat{s^2} = \frac{1}{(x_{k-1}-x_0+1)(t_{n-1}-t_0)} \sum_{x=x_0}^{x_{k-1}} \sum_{t=t_0}^{t_{n-1}} \left(Y_{t,x} - \overline{Y}_x\right) , \tag{3.35}$$

$$\widehat{b} = \frac{1}{x_{k-1}-x_0+1} \sum_{x=x_0}^{x_{k-1}} \overline{Y}_x = \frac{1}{(x_{k-1}-x_0+1)(t_{n-1}-t_0+1)} \sum_{x=x_0}^{x_{k-1}} \sum_{t=t_0}^{t_{n-1}} Y_{t,x} = \overline{Y} , \tag{3.36}$$

$$\widehat{U} = \frac{1}{x_{k-1}-x_0} \sum_{x=x_0}^{x_{k-1}} \left(\overline{Y}_x - \overline{Y}\right) - \frac{\widehat{s^2}}{t_{n-1}-t_0+1} , \tag{3.37}$$

$$\widehat{K} = \left[(t_{n-1}-t_0+1)\,\widehat{U}\right] \left[\widehat{s^2} + (t_{n-1}-t_0+1)\,\widehat{U}\right]^{-1} . \tag{3.38}$$

The Bühlmann credibility estimates for one year ahead can be obtained by

$$\widehat{Y}_{t_{n-1}+1,x} = \widehat{K}\,\overline{Y}_x + (1-\widehat{K})\,\overline{Y}, \quad \text{for } x = x_0,\ldots,x_{k-1}\,. \tag{3.39}$$

In contrast to the credibility regression approaches, which aim to capture the downward trend of $m(t,x)$s over t, for the Bühlmann credibility approach to be applied, this downward trend must be eliminated. For this reason, Tsai and Lin (2017a) applied the Bühlmann credibility model on the time series of mortality rate changes rather than the mortality rate levels, i.e., $Y_{t,x} = \log m(t,x) - \log m(t-1,x)$, for $t_1,\ldots,t_{n-1}$. Then, they proposed two strategies for estimating $Y_{t+h,x}$, $h = 2,\ldots,H$. The first strategy expands fitting window (EW) by one year, similarly with the EEM regression method, described in Section 3.4 and the second one moves fitting window (MW) by one year, similarly with the MEM regression method. In what follows, we compare the forecasting performance between the Bühlmann and the credibility regression methods on Greek data. To be consistent with the Bühlmann modelling framework of Tsai and Lin (2017a), age fitting spans [21,85] and [56,85] were selected and forecast errors were also evaluated under the averaged MAPFE values, which is defined by

$$MAPFE_{avg} = \frac{1}{H\times(x_{k-1}-x_0+1)}\sum_{h=1}^{H}\sum_{x=x_0}^{x_{k-1}}\frac{|\widehat{m}(t_{n-1}+h,x)-m(t_{n-1}+h,x)|}{|m(t_{n-1}+h,x)|}\times 100.$$

Error values for each gender were evaluated by fitting $Y_{t,x}$s for periods [1982,2000], [1986,2000], and [1990,2000]. Comparison of averaged MAFE, RMSFE and MAPFE[4] results between Bühlmann and credibility regression methods is given for both genders in Table 3.4 (a) – (c) for ages [21,85] and Table 3.5 (a) – (c) for ages [56,85].

The results indicate that credibility regression methods produce the smallest MAFE, RMSFE and MAPFE values for the majority of the selected fitting periods for both age spans. More precisely, the FC-MEM method has the best average performance according to MAFE and MAPFE values for ages [21,85], while the RC-MEM method seems to be more appropriate to capture future mortality trends for older ages [56,85]. We note that the smallest values in average are produced by different regression methods, depending on which measure is used. Such inconsistencies are expected due to the nature of MAFE, RMSFE and MAPFE formulas. That was also pointed out by Tsai and Yang (2015).

[4] To be distinguished, MAFE and RMSFE averaged error values are rounded to four decimal points, while, for MAPFE values, two decimal points are enough.

Table 3.4 MAFE, RMSFE and MAPFE values of forecast errors over the period $[2001, 2010]$ for ages $[21, 85]$.

(a) MAFE Values

$MAFE_{[21,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
$[1982, 2000]$	Male	0.2348	0.2334	0.1404	0.1287	0.1379	0.1444	0.1361	0.1460
	Female	0.0930	0.0931	0.0816	0.0882	0.0843	0.0791	0.0780	0.0790
$[1986, 2000]$	Male	0.2170	0.2294	0.1329	0.1321	0.1306	0.1364	0.1358	0.1373
	Female	0.0918	0.0919	0.0782	0.0852	0.0909	0.0741	0.0805	0.0747
$[1990, 2000]$	Male	0.2355	0.2258	0.1399	0.1392	0.1369	0.1434	0.1422	0.1423
	Female	0.0954	0.0933	0.0836	0.0839	0.0879	0.0798	0.0818	0.0802
Average		0.1613(8)	0.1612(7)	0.1094(2)	0.1096(4)	0.1114(6)	0.1095(3)	0.1091(1)	0.1099(5)

(b) RMSFE Values

$RMSFE_{[21,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
$[1982, 2000]$	Male	0.4980	0.4948	0.2633	0.2342	0.2566	0.2756	0.2564	0.2799
	Female	0.1795	0.1795	0.1613	0.1884	0.1730	0.1540	0.1581	0.1541
$[1986, 2000]$	Male	0.4584	0.4861	0.2447	0.2387	0.2386	0.2564	0.2532	0.2591
	Female	0.1767	0.1772	0.1633	0.1781	0.1999	0.1484	0.1643	0.1502
$[1990, 2000]$	Male	0.4997	0.4765	0.2578	0.2574	0.2472	0.2704	0.2666	0.2668
	Female	0.1849	0.1802	0.1640	0.1761	0.1767	0.1567	0.1674	0.1570
Average		0.3329(8)	0.3324(7)	0.2091(1)	0.2122(5)	0.2153(6)	0.2103(2)	0.2110(3)	0.2112(4)

(c) MAPFE Values

$MAPFE_{[21,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
$[1982, 2000]$	Male	11.90	11.86	11.97	11.24	11.80	11.95	11.43	12.00
	Female	13.75	13.76	11.66	11.83	11.69	11.66	11.54	11.66
$[1986, 2000]$	Male	11.30	11.71	12.05	10.76	11.73	11.86	10.72	11.90
	Female	13.71	13.72	11.52	11.82	11.89	11.56	11.73	11.55
$[1990, 2000]$	Male	11.93	11.60	10.81	9.81	10.60	10.57	9.71	10.52
	Female	13.83	13.77	11.84	11.79	12.08	12.05	11.83	11.99
Average		12.73(7)	12.74(8)	11.64(6)	11.21(2)	11.63(5)	11.61(4)	11.16(1)	11.60(3)

Table 3.5 MAFE, RMSFE and MAPFE values of forecast errors over the period [2001, 2010] for ages [56, 85].

(a) MAFE Values

$MAFE_{[56,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
[1982, 2000]	Male	0.3599	0.3503	0.3272	0.3012	0.3210	0.3262	0.3036	0.3255
	Female	0.1686	0.1623	0.1717	0.1633	0.1735	0.1711	0.1595	0.1709
[1986, 2000]	Male	0.3233	0.3430	0.2893	0.2958	0.2886	0.2946	0.2991	0.2937
	Female	0.1481	0.1539	0.1534	0.1601	0.1617	0.1495	0.1573	0.1511
[1990, 2000]	Male	0.3745	0.3641	0.2958	0.2934	0.2937	0.2999	0.2954	0.2973
	Female	0.1670	0.1646	0.1617	0.1616	0.1613	0.1601	0.1625	0.1615
Average		0.2569(8)	0.2564(7)	0.2332(3)	0.2293(1)	0.2333(4)	0.2336(6)	0.2296(2)	0.2334(5)

(b) RMSFE Values

$RMSFE_{[56,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
[1982, 2000]	Male	0.5411	0.5261	0.4670	0.4213	0.4524	0.4700	0.4305	0.4679
	Female	0.2358	0.2282	0.2368	0.2242	0.2366	0.2389	0.2202	0.2381
[1986, 2000]	Male	0.4852	0.5159	0.4065	0.4138	0.3987	0.4221	0.4224	0.4185
	Female	0.2120	0.2178	0.2151	0.2235	0.2271	0.2089	0.2192	0.2107
[1990, 2000]	Male	0.5636	0.5472	0.4139	0.4130	0.4072	0.4291	0.4184	0.4195
	Female	0.2338	0.2307	0.2243	0.2246	0.2236	0.2217	0.2257	0.2232
Average		0.3786(8)	0.3777(7)	0.3273(4)	0.3201(1)	0.3243(3)	0.3318(6)	0.3227(2)	0.3297(5)

(c) MAPFE Values

$MAPFE_{[56,85]}$		Bühlmann Methods		Regression Methods – RC			Regression Methods – FC		
Fitting Period	Gender	EW	MW	SEM	MEM	EEM	SEM	MEM	EEM
[1982, 2000]	Male	9.53	9.34	9.48	9.17	9.54	9.29	8.97	9.31
	Female	9.93	9.72	9.98	9.81	10.36	9.65	9.43	9.69
[1986, 2000]	Male	8.82	9.20	8.78	8.97	8.99	8.61	8.85	8.66
	Female	9.23	9.45	9.14	9.42	9.62	8.84	9.26	8.98
[1990, 2000]	Male	9.82	9.61	8.85	8.74	9.00	8.62	8.74	8.78
	Female	9.88	9.81	9.49	9.33	9.46	9.32	9.37	9.48
Average		9.54(8)	9.52(7)	9.29(5)	9.24(4)	9.50(6)	9.06(1)	9.10(2)	9.15(3)

3.5.4 Application in Insurance-Related Products

In this section, we apply the mortality forecasts obtained from the Lee–Carter, the Cairns–Blake–Dowd and the credibility regression models to calculate life premiums, reflecting the appropriateness of each model in pricing applications. Denote $A^{1}_{t_{n-1}+1,x:\overline{K|}}$ as the fully discrete life insurance premium, payable at the end of the year of death, if it occurs within a term of K years and $A_{t_{n-1}+1,x:\overline{K|}}^{\quad\;\;1}$ as the pure endowment, payable at the end of K years in case of being alive. Both products are issued to an insured aged x in year $t_{n-1}+1$. Net premiums (NP) are obtained (see Bozikas and Pitselis, 2018) by

$$A_{\underset{t_{n-1}+1,x:\overline{K|}}{1}} = \sum_{k=0}^{K-1} {}_k p_{t_{n-1}+1,x} \cdot q(t_{n-1}+1+k, x+k) \cdot (1+i)^{-(k+1)} \,, \tag{3.40}$$

$$A_{\underset{t_{n-1}+1,x:\overline{K|}}{}\overset{1}{}} = {}_K p_{t_{n-1}+1,x} \cdot (1+i)^{-K} \,, \tag{3.41}$$

where ${}_k p_{t_{n-1}+1,x}$ denotes the k-year survival probability for age x in year $t_{n-1}+1$, while its estimate is given by ${}_k\widehat{p}_{t_{n-1}+1,x} = \widehat{p}_{t_{n-1}+1,x} \cdot \ldots \cdot \widehat{p}_{t_{n-1}+1+k-1,x+k-1}$, $k = 1, \ldots, K-1$ and similarly for ${}_K\widehat{p}_{t_{n-1}+1,x}$, where i is the interest rate and ${}_0\widehat{p}_{t_{n-1}+1,x} = 1$. In addition, to see the performance on a life annuity product, typically used for pension applications, denote $\ddot{a}_{t_{n-1}+1,x:\overline{k|}}$ as the discrete life annuity-due at age x in year $t_{n-1}+1$, payable annually for up to K years. Its actuarial present value (APV) can be obtained by

$$\ddot{a}_{t_{n-1}+1,x:\overline{K|}} = \sum_{k=0}^{K-1} {}_k p_{t_{n-1}+1,x} \cdot (1+i)^{-k} \,. \tag{3.42}$$

Then, we apply the estimated mortality rates obtained from the LC, CBD and credibility methods, fitted to 1981–2000 rates, to calculate the NPs and APVs for ages 55–74 with $K = 10$, assuming $i = 4\%$. The errors between forecasted values and those produced from the observed mortality rates for the years 2001–2010 are evaluated using MAFE and RMSFE, which are defined by

$$MAFE_{avg} = \frac{1}{20} \sum_{x=55}^{74} \left| \widehat{A}_{\underset{2001,x:\overline{10|}}{1}} - A_{\underset{2001,x:\overline{10|}}{1}} \right| \times 100 \,, \tag{3.43}$$

$$RMSFE_{avg} = \sqrt{\frac{1}{20} \sum_{x=55}^{74} \left(\widehat{A}_{\underset{2001,x:\overline{10|}}{1}} - A_{\underset{2001,x:\overline{10|}}{1}} \right)^2} \times 100 \,. \tag{3.44}$$

Similarly, MAFE and RMSFE formulas are adjusted for pure endowment or annuity products by replacing $A_{\underset{t_{n-1}+1,x:\overline{K|}}{1}}$ with $A_{\underset{t_{n-1}+1,x:\overline{K|}}{}\overset{1}{}}$ or $\ddot{a}_{t_{n-1}+1,x:\overline{K|}}$ in Equations (3.43) and (3.44). Table 3.6 presents the averaged error values in ranking order for a 10 year forecasted life insurance, pure endowment and life annuity for both genders, aged 55–74 in 2001–2010. In addition, Figure 3.6 illustrates the absolute forecast error values against each corresponding age (AFE$_x$) for the top LC, CBD, RC and FC credibility regression methods for males and females, according to Table 3.6 values. For each model, AFE$_x$ is obtained from $AFE_x = \left| \widehat{A}_{\underset{x,2001:\overline{10|}}{1}} - A_{\underset{x,2001:\overline{10|}}{1}} \right| \times 100$.

According to MAFE and RMSFE values for both genders and insurance products in Table 3.6, credibility regression models produce better insurance-related forecasts in comparison with the LC and CBD modelling methods. We can easily observe that for each gender, error measures coincide in the same ranking order for all insurance products. In particular, measures show that credibility regression methods under a moving fitting span outperform LC and CBD methods in aggregate, with FC-MEM being dominant and RC-MEM following. This fact is also evident in Figure 3.6, where absolute error values against age for the MEM regression models lie on the lower levels for all the insurance products. Nevertheless, the FC-SEM should also be a good modelling choice for pricing insurance-related products.

Table 3.6 MAFE and RMSFE values (ranking order in brackets) for a 10-year forecasted life insurance, a pure endowment and a life annuity for males and females of ages 55–74 during 2001–2010.

(a) Life Insurance										
$MAFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	1.6019(8)	1.5640(7)	1.7151(10)	1.6794(9)	1.5000(6)	1.4169(2)	1.4924(5)	1.4735(3)	1.3932(1)	1.4741(4)
Female	1.0264(6)	1.0269(7)	1.2141(10)	1.1079(9)	1.0262(5)	0.9317(2)	1.0346(8)	0.9898(3)	0.8840(1)	0.9910(4)
$RMSFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	1.8423(8)	1.8043(7)	1.9401(10)	1.9089(9)	1.7125(6)	1.6143(2)	1.7043(5)	1.6871(3)	1.5989(1)	1.6875(4)
Female	1.2320(8)	1.2294(7)	1.4023(10)	1.2918(9)	1.2133(5)	1.0965(2)	1.2215(6)	1.1744(3)	1.0494(1)	1.1756(4)
(b) Pure Endowment										
$MAFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	1.1439(8)	1.1139(7)	1.2417(10)	1.2044(9)	1.0722(6)	1.0153(2)	1.0681(5)	1.0512(3)	0.9942(1)	1.0518(4)
Female	0.7181(7)	0.7192(8)	0.8894(10)	0.7923(9)	0.7340(5)	0.6717(2)	0.7463(6)	0.7026(3)	0.6297(1)	0.7038(4)
$RMSFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	1.3274(8)	1.2975(7)	1.4104(10)	1.3786(9)	1.2347(6)	1.1659(2)	1.2303(5)	1.2150(3)	1.1535(1)	1.2154(4)
Female	0.8745(8)	0.8717(7)	1.0310(10)	0.9319(9)	0.8774(5)	0.7968(2)	0.8889(6)	0.8440(3)	0.7552(1)	0.8451(4)
(c) Life Annuity										
$MAFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	5.4466(8)	5.2602(7)	6.3032(10)	5.7857(9)	5.1642(6)	4.9260(2)	5.1561(5)	5.0331(3)	4.7893(1)	5.0369(4)
Female	2.9140(7)	2.9361(8)	4.4471(10)	3.5932(9)	3.1527(5)	2.9479(2)	3.2151(6)	2.9656(3)	2.7024(1)	2.9721(4)
$RMSFE_{avg}$	**Mortality Models**				**Random Coefficients (RC)**			**Fixed Coefficients (FC)**		
Gender	LC	LC-Poisson	CBD	CBD-Poisson	*SEM*	*MEM*	*EEM*	*SEM*	*MEM*	*EEM*
Male	6.6138(7)	6.4342(8)	7.2583(10)	6.8729(9)	6.1786(6)	5.8730(2)	6.1608(5)	6.0681(3)	5.7919(1)	6.0704(4)
Female	3.7013(7)	3.7218(8)	5.1510(10)	4.3300(9)	3.9187(5)	3.6344(2)	3.9846(6)	3.7084(3)	3.3878(1)	3.7155(4)

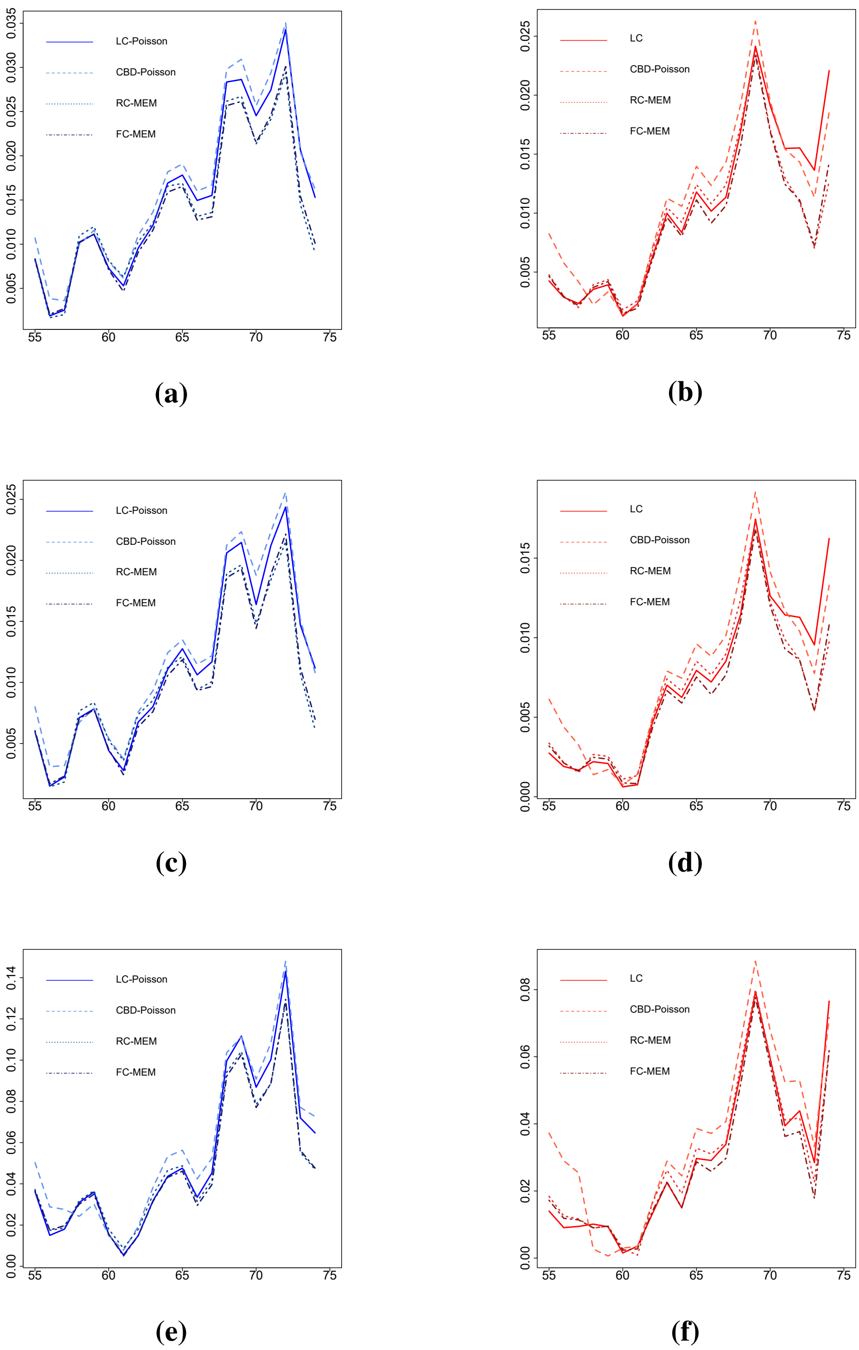


Figure 3.6 AFE values against age of life insurance and annuity products for the top LC, CBD and credibility regression models for males (left panels) and females (right panels): (a) life insurance AFEs for males; (b) life insurance AFEs for females; (c) pure endowment AFEs for males; (d) pure endowment AFEs for females; (e) life annuity AFEs for males; and (f) life annuity AFEs for females.

3.6 Concluding Remarks

Credibility regression techniques seem to be of special interest and particularly useful for mortality datasets of a relatively short historical period of observations (limited data), as they can efficiently capture the underlying mortality trend for a given age, using all the information gained from populations of other ages. This chapter proposes mortality modelling approaches embedded, for the first time, in a credibility regression framework. In our illustration on Greek data, credibility regression approaches resulted in better forecasts for both genders (in terms of MAFE and RMSFE measures), compared to the Lee–Carter and Cairns–Blake–Dowd models, as well as the Bühlmann credibility approach (Tsai and Lin, 2017a). Their performance was also evaluated on insurance-related applications.

Specifically, in Section 3.3, we proposed a credibility regression mortality framework with random coefficients and a special case with fixed coefficients. To estimate future mortality rates, we presented extrapolation methods for each credibility approach in Section 3.4. The applicability of our modelling approaches was comparatively illustrated on Greek male and female data in Section 3.5, accompanied with an explanation of the credibility effects in mortality modelling and a pricing application on insurance-related products. From our analysis, we concluded that, in aggregate, credibility modelling methods performed better than the LC and CBD methods. Forecasting accuracy results indicate that, for the whole age fitting span, fixed coefficients credibility methods performed better on average, while, for higher ages, the RC-MEM should also be a good choice. In addition, the FC-MEM performed a bit better in aggregate on pricing insurance-related products.

Furthermore, we noted that FC-SEM credibility forecasts were closer to observed rates for the same periods, when we used population exposure to risk as weights, i.e., $\boldsymbol{W}_x = \mathrm{diag}\left[E(t_0,x),\ldots,E(t_{n-1},x)\right]$, for $x = x_0,\ldots,x_{k-1}$, but weighted regression is restricted for use only under the SEM , as $E(t,x)$s are practically unknown for the upcoming years. Additionally, during the estimation procedure for the random regression models, we observed that, if we use the MINQUE estimator (Remark 3.3) instead of (3.19), error values for all the credibility modelling methods become even smaller for both genders.

For the sake of comparability, the Bühlmann credibility approach (Tsai and Lin, 2017a) was applied on our dataset in Section 3.5, where the credibility regression methods resulted to better forecasts based on MAFE, RMSFE and MAPFE measures. In addition, credibility regression methods had a very good forecasting performance, when we applied them to the datasets of other countries for a limited selected fitting period (1980–2000), such as Belgium, Finland, Norway, Ireland, Slovakia and New Zealand.

Chapter 4

A Hierarchical Credibility Regression Mortality Model for Multiple Populations

4.1 Introduction

In the previous chapter, we described how credibility regression can be used to model the mortality trends of a population, especially in cases where there is limited mortality experience for a specific age, but extensive experience for the entire age range. Even if credibility regression can yield desirable results when it is applied on mortality rates of a single population, a question arises when we deal with datasets of multiple populations. Since Wilson (2001) observed a global convergence in mortality, many multi-population mortality models have been developed to account for any relationship between multiple populations (see Introduction, Section 1.3).

Hierarchical credibility can accommodate classification schemes, where multi-population mortality data can be represented in a hierarchical form. Jewell (1975) extended Bühlmann credibility into a hierarchical structure. Under this specification, Tsai and Wu (2018) modelled the mortality rates for multiple populations. Furthermore, Jewell's model was generalized to a hierarchical credibility regression model with random parameters on two levels in Sundt (1979) and on multiple levels in Sundt (1980). This chapter extends the credibility regression mortality approach of the previous chapter (Bozikas and Pitselis, 2019) to its multi-level hierarchical counterpart to derive multi-population mortality forecasts, using a different approach than Sundt (1980) for the estimation of model parameters.

The rest of this chapter is organized as follows. Section 4.2 reviews two widely used Lee-Carter extensions for multiple populations. Section 4.3 presents the multi-level credibility mortality model, with a detailed description of parameters estimation. Section 4.4 gives a numerical illustration on projecting age specific mortality rates for males and females using data from three countries in northern Europe, the Ireland, the Finland and the Norway. Concluding remarks are presented in Section 4.5.

4.2 Multi-Population Mortality Models: A Review of Methods

In this section, we briefly review the most popular and widely applied multi-population mortality models, the joint-k model extension for multiple populations presented in Carter and Lee (1992) and the augmented common factor model proposed by Li and Lee (2005). We denote $D(t,xgc)$ as the observed number of deaths and $E(t,xgc)$ as the average population (exposure to risk) in calendar year $t=t_1,\ldots,t_n$, for age $x=x_1,\ldots,x_k$, gender $g=1,\ldots,G$ $(G=2,$ i.e. male $=1,$ female $=2)$ and country $c=1,\ldots,C$. Then, the corresponding age-specific mortality rates $m(t,xgc)$ are obtained by the ratio $D(t,xgc)/E(t,xgc)$.

4.2.1 The joint-K model

Carter and Lee (1992) proposed a model structure for multiple populations, in which mortality is jointly driven by a single period parameter K_t as follows

$$\log m_{t,xgc} = \alpha_{xgc} + \beta_{xgc}K_t + \varepsilon_{t,xgc}, \tag{4.1}$$

where α_{xgc} and β_{xgc} are defined as in the original Lee-Carter model, but now, K_t is the common period parameter of the mortality level in year t for gender g and country c. Errors $\varepsilon_{t,xgc}$ are assumed independent and identically distributed. Then, constraints $\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=x_1}^{x_k}\beta_{xgc}=1$ and $\sum_{t=t_1}^{t_n}K_t=0$ lead to the following parameter estimates

$$\widehat{\alpha}_{xgc} = \frac{1}{t_n-t_1+1}\sum_{t=t_1}^{t_n}\log m_{t,xgc}, \quad \widehat{K}_t = \sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=x_1}^{x_k}[\log\widehat{m}_{t,xgc}-\widehat{\alpha}_{xgc}]$$

and

$$\widehat{\beta}_{xgc} = \frac{\sum_{t=t_1}^{t_n}[\log m_{t,xgc}-\widehat{\alpha}_{xgc}]\widehat{K}_t}{\sum_{t=t_1}^{t_n}\widehat{K}_t^2} .$$

The common period parameter $\widehat{K}_t$ is assumed to follow a random walk with a drift θ, $\widehat{K}_t=\widehat{K}_{t-1}+\theta+\varepsilon_t$, where the time trend errors ε_t are again assumed to be independent and identically distributed, and independent of the model errors $\varepsilon_{t,xgc}$. The drift parameter is estimated by $\widehat{\theta}=\frac{1}{t_n-t_1}\sum_{t=t_1+1}^{t_n}(\widehat{K}_t-\widehat{K}_{t-1})=\frac{\widehat{K}_{t_n}-\widehat{K}_{t_1}}{n-1}$ and then it used to project period estimates $\widehat{K}_{t_n+h}=\widehat{K}_{t_n}+\widehat{\theta}h$. Thus, projected mortality rates for $h=1,2,\ldots$ years ahead, for age x, gender g and country c are given by

$$\log\widehat{m}_{t_n+h,xgc} = \widehat{\alpha}_{xgc}+\widehat{\beta}_{xgc}\widehat{K}_{t_n+h} = \log\widehat{m}_{t_n,xgc}+(\widehat{\beta}_{xgc}\widehat{\theta})h . \tag{4.2}$$

4.2.2 The augmented common factor model

To avoid long-run divergence in mean mortality forecasts for multiple countries, Li and Lee (2005) modified the original Lee-Carter model by setting a common age parameter B_x and the same period parameter K_t for all populations as follows

$$\log m_{t,xgc} = \alpha_{xgc} + B_x K_t + \varepsilon_{t,xgc}. \tag{4.3}$$

Again, we use two constraints $\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=x_1}^{x_k} w_{gc}B_x = 1$, and $\sum_{t=t_1}^{t_n} K_t = 0$, where w_{gc} is the weight for gender g in country c, set to be proportional to the total number of populations, i.e., $w_{gc} = 1/(GC)$. The model parameters are estimated by

$$\widehat{\alpha}_{xgc} = \frac{1}{t_n - t_1 + 1}\sum_{t=t_1}^{t_n} \log m_{t,xgc}, \quad \widehat{K}_t = \sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=x_1}^{x_k} w_{gc}\left[\log m_{t,xgc} - \widehat{\alpha}_{xgc}\right]$$

and

$$\widehat{B}_x = \frac{\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{t=t_1}^{t_n} w_{gc}\left[\log m_{t,xgc} - \widehat{\alpha}_{xgc}\right]\widehat{K}_t}{\sum_{t=t_1}^{t_n}\widehat{K}_t^2} .$$

To include the individual differences in the trends, Li and Lee (2005) suggested an additional factor $\alpha'_{xgc}\kappa'_{t,gc}$ to form the augmented common factor model

$$\log m_{t,xgc} = \alpha_{xgc} + B_x\, K_t + \alpha'_{xgc}\kappa'_{t,gc} + \varepsilon_{t,xgc} .$$

Assuming the extra constraint $\sum_{x=x_1}^{x_k}\alpha'_{xgc} = 1$, the additional parameters are estimated as

$$\widehat{\kappa}'_{t,gc} = \sum_{x=x_1}^{x_k}\left[\log m_{t,xgc} - \widehat{\alpha}_{xgc} - \widehat{B}_x\widehat{K}_t\right]$$

and

$$\widehat{\alpha}'_{xgc} = \frac{\sum_{t=t_1}^{t_n}\left[\log m_{t,xgc} - \widehat{\alpha}_{xgc} - \widehat{B}_x\widehat{K}_t\right]\widehat{\kappa}'_{t,gc}}{\sum_{t=t_1}^{t_n}\widehat{\kappa}'^{2}_{t,gc}} .$$

Both period parameters $\widehat{K}_t$ and $\widehat{\kappa}'_{t,gc}$ follow a random walk model with a drift, given by $\widehat{K}_t = \widehat{K}_{t-1} + \theta + \varepsilon_t$ and $\widehat{\kappa}'_{t,gc} = \widehat{\kappa}'_{t-1,gc} + \theta'_{gc} + \varepsilon_{t,gc}$, respectively, where time trend errors ε_t and $\varepsilon_{t,gc}$ are assumed to be independent and identically distributed, and independent of the model error $\varepsilon_{t,xgc}$. The drift parameters can be estimated by

$$\widehat{\theta} = \frac{1}{t_n - t_1}\sum_{t=t_1+1}^{t_n}\left(\widehat{K}_t - \widehat{K}_{t-1}\right) = \frac{\widehat{K}_{t_n} - \widehat{K}_{t_1}}{n-1}$$

and

$$\widehat{\theta}'_{gc} = \frac{1}{t_n - t_1}\sum_{t=t_1+1}^{t_n}\left(\widehat{\kappa}'_{t,gc} - \widehat{\kappa}'_{t-1,gc}\right) = \frac{\widehat{\kappa}'_{t_n,gc} - \widehat{\kappa}'_{t_1,gc}}{n-1} .$$

Then, projected mortality rates for $h = 1, 2, \ldots$ years ahead, for age x, gender g and country c are obtained by

$$\begin{aligned} \log \widehat{m}_{t_n+h,xgc} &= \widehat{\alpha}_{xgc} + \widehat{B}_x \widehat{K}_{t_{n+h}} + \widehat{\alpha}'_{xgc} \widehat{\kappa}'_{t_{n+h},gc} \\ &= \log \widehat{m}_{t_n,xgc} + (\widehat{B}_x \widehat{\theta} + \widehat{\alpha}'_{xgc} \widehat{\theta}'_{gc}) h \,. \end{aligned} \tag{4.4}$$

Table 4.1 lists the model structure of the original Lee-Carter (1992) model (LC) for a single population, as well as its two extensions for multiple populations, the joint-K model ($\mathrm{LC_{joK}}$) and the augmented common factor model ($\mathrm{LC_{acf}}$).

Table 4.1 Structure overview of multi-population Lee-Carter models.

Model	Structure	Original Papers
LC	$\log m_{t,xgc} = \alpha_{xgc} + \beta_{xgc} \cdot \kappa_{t,gc} + \varepsilon_{t,xgc}$	Lee and Carter (1992)
$\mathrm{LC_{joK}}$	$\log m_{t,xgc} = \alpha_{xgc} + \beta_{xgc} \cdot K_t + \varepsilon_{t,xgc}$	Carter and Lee (1992)
$\mathrm{LC_{acf}}$	$\log m_{t,xgc} = \alpha_{xgc} + B_x \cdot K_t + \alpha'_{xgc} \cdot \kappa'_{t,gc} + \varepsilon_{t,xgc}$	Li and Lee (2005)

Remark 4.1. We can easily observe that expressions (4.2) and (4.4) are linear functions of the forecasting horizon h, where their intercept equal to the fitted rates of the last observed year and their slope is the product of the estimated age parameters with the drift terms.

4.3 Credibility Mortality Modelling for Multiple Populations

In this section, we present a multi-level hierarchical credibility regression (henceforth HCR) approach to model mortality for multiple populations. Beginning from the bottom, mortality data are first classified by age (Level 1). Ages are then grouped by gender (Level 2) and genders by country (Level 3) to constitute a multi-country credibility structure. Each level is also associated with a random risk variable that reflects specific characteristics for each country, gender and age. More precisely, country level is associated with a random variable Θ_c, $c = 1, \ldots, C$, gender level is associated with Θ_{gc}, $g = 1, \ldots, G_c$, and age level by Θ_{xgc}, $x = x_1, \ldots, x_{k_{gc}}$ (see Figure 4.1).

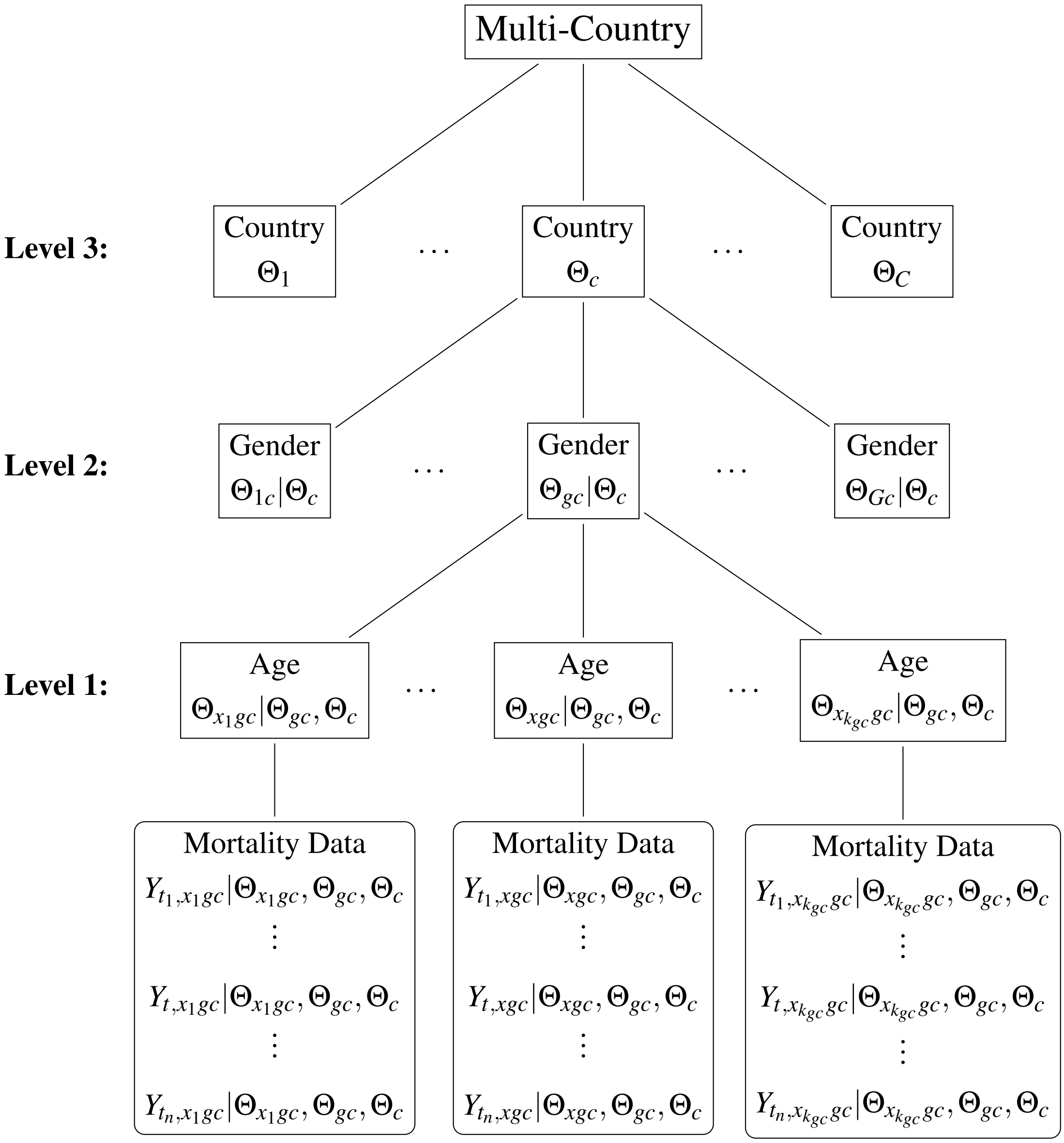


Figure 4.1 A multi-level hierarchical credibility structure.

4.3.1 Assumptions and Notation

We assume that the observable variable $Y_{t,xgc}$ corresponds to the log transform of mortality rates $m_{t,x}$ for $t = t_1, \dots, t_n$, $x = x_1, \dots, x_{k_{gc}}$, $g = 1, \dots, G_c$ and $c = 1, \dots, C$. We denote $\boldsymbol{Y}_{xgc} = (Y_{t_1,xgc}, \dots, Y_{t_n,xgc})'$ as the mortality vector for age x, gender g, country c and $\boldsymbol{Z}_{xgc}$ as the corresponding design matrix of explanatory variables. We note that in general, design matrix could consist of various explanatory variables that reflect mortality characteristics. However, due to lack of data related with other mortality factors, we assume that for each age, gender and country, mortality rates are only affected by time trends. Thus, for all x, g and c the observable variables and the design matrices for Level 1 (age), Level 2 (gender) and Level 3 (country) are denoted as

$$\boldsymbol{Y}_{xgc} = (\log m_{t_1,xgc}, \ldots, \log m_{t_n,xgc})', \;\; \boldsymbol{Z}_{xgc} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & n \end{pmatrix}', \quad \text{(Level 1)}, \tag{4.5}$$

$$\boldsymbol{Y}_{gc} = (\boldsymbol{Y}'_{x_1gc}, \ldots, \boldsymbol{Y}'_{x_{k_{gc}}gc})', \;\; \boldsymbol{Z}_{gc} = (\boldsymbol{Z}'_{x_1gc}, \ldots, \boldsymbol{Z}'_{x_{k_{gc}}gc})', \quad \text{(Level 2)} \tag{4.6}$$

and

$$\boldsymbol{Y}_c = (\boldsymbol{Y}'_{1c}, \ldots, \boldsymbol{Y}'_{G_cc})', \;\; \boldsymbol{Z}_c = (\boldsymbol{Z}'_{1c}, \ldots, \boldsymbol{Z}'_{G_cc})', \quad \text{(Level 3)}. \tag{4.7}$$

Then, country c consists of the set variables $\{\Theta_{xgc}, \Theta_{gc}, \Theta_c, \boldsymbol{Y}_c\}$ for $x = x_1, \ldots, x_{k_{gc}}$, $g = 1, \ldots, G_c$, gender g consists of the variables $\{\Theta_{xgc}, \Theta_{gc}, \boldsymbol{Y}_{gc}\}$ for $x = x_1, \ldots, x_{k_{gc}}$ and age x consists of $\{\Theta_{xgc}, \boldsymbol{Y}_{xgc}\}$ under the following assumptions:

(A1) The countries are independent.

(A2) For each $c = 1, \ldots, C$ and for given Θ_c, the genders $\{\Theta_{xgc}, \Theta_{gc}, \boldsymbol{Y}_{gc}\}$ are conditionally independent.

(A3) For each $c = 1, \ldots, C$ and $g = 1, \ldots, G_c$ and for given (Θ_{gc}, Θ_c), the ages $\{\Theta_{xgc}, \boldsymbol{Y}_{xgc}\}$ are conditionally independent.

(A4) For each $c = 1, \ldots, C$, $g = 1, \ldots, G_c$, $x = x_1, \ldots, x_{k_{gc}}$ and for given $(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$, the mortality observations $\boldsymbol{Y}_{xgc}$ are conditionally independent.

(A5) All the vectors $(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$ for $c = 1, \ldots, C$, $g = 1, \ldots, G_c$ and $x = x_1, \ldots, x_{k_{gc}}$ are identically distributed.

(A6) $E(\boldsymbol{Y}_{xgc}|\Theta_{xgc}, \Theta_{gc}, \Theta_c) = \boldsymbol{Z}_{xgc}\, \boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$, for $t = t_1, ..., t_n$, where $\boldsymbol{Z}_{xgc}$ is a $n \times 2$ design matrix and $\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$ is the coefficients vector.

(A7) $\mathrm{Cov}(\boldsymbol{Y}_{xgc}|\Theta_{xgc}, \Theta_{gc}, \Theta_c) = \sigma^2(\Theta_{xgc}, \Theta_{gc}, \Theta_c)\, \boldsymbol{W}_{xgc}^{-1}$, where $\boldsymbol{W}_{xgc}$ is a fixed positive definite diagonal $(n \times n)$ matrix with known elements $w_{t,xgc}$ for $c = 1, \ldots, C$, $g = 1, \ldots, G_c$, $x = x_1, \ldots, x_{k_{gc}}$ and $t = t_1, \ldots, t_n$.

The structural parameters of the hierarchical credibility regression model are given by

(S1) $\boldsymbol{\beta} = E[\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)]$,

(S2) $s^2 = E[\sigma^2(\Theta_{xgc}, \Theta_{gc}, \Theta_c)]$,

(S3) $\boldsymbol{A} = E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)|\Theta_{gc}, \Theta_c]\}$,

(S4) $\boldsymbol{U} = E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)|\Theta_c]\}$,

(S5) $\boldsymbol{\Psi} = \mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$.

The regression coefficients for Level 1 are given by

$$\widehat{\boldsymbol{\beta}}_{xgc} = (\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Y}_{xgc}, \quad \text{(Level 1)}, \tag{4.8}$$

with conditional covariance

$$\text{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c) = \sigma^2(\Theta_{xgc},\Theta_{gc},\Theta_c)(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\ , \tag{4.9}$$

while for the other levels we have

$$\widehat{\boldsymbol{\beta}}_{gc} = (\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc})^{-1}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc}, \quad \text{(Level 2)}, \tag{4.10}$$

$$\widehat{\boldsymbol{\beta}}_{c} = (\sum_{g=1}^{Gc}\boldsymbol{K}_{gc})^{-1}\sum_{g=1}^{Gc}\boldsymbol{K}_{gc}\widehat{\boldsymbol{\beta}}_{gc}, \quad \text{(Level 3)} \tag{4.11}$$

and for the higher multi-country level

$$\widehat{\boldsymbol{\beta}} = (\sum_{c=1}^{C}\boldsymbol{K}_{c})^{-1}\sum_{c=1}^{C}\boldsymbol{K}_{c}\widehat{\boldsymbol{\beta}}_{c}\ . \tag{4.12}$$

The corresponding credibility factors for each level are respectively defined by

$$\boldsymbol{K}_{xgc} = \boldsymbol{A}[\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}]^{-1}, \quad \text{(Level 1)}, \tag{4.13}$$

$$\boldsymbol{K}_{gc} = \boldsymbol{U}[\boldsymbol{U}+(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc})^{-1}\boldsymbol{A}]^{-1}, \quad \text{(Level 2)} \tag{4.14}$$

and

$$\boldsymbol{K}_{c} = \boldsymbol{\Psi}[\boldsymbol{\Psi}+(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc})^{-1}\boldsymbol{U}]^{-1}, \quad \text{(Level 3)}. \tag{4.15}$$

Theorem 4.2. *Based on the above assumptions and notation, we obtain the following expressions for the conditional expectations*

$$E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{xgc},\Theta_{gc},\Theta_c] = \boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c), \tag{4.16}$$

$$E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c] = \boldsymbol{\beta}(\Theta_{gc},\Theta_c), \tag{4.17}$$

$$E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_c] = \boldsymbol{\beta}(\Theta_c), \tag{4.18}$$

$$E[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c] = \boldsymbol{\beta}(\Theta_c), \tag{4.19}$$

$$E[\boldsymbol{\beta}(\Theta_c)|\Theta_c] = \boldsymbol{\beta}(\Theta_c), \tag{4.20}$$

$$E[\boldsymbol{\beta}(\Theta_c)] = \boldsymbol{\beta}, \tag{4.21}$$

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c] = \boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c), \tag{4.22}$$

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{gc},\Theta_c] = \boldsymbol{\beta}(\Theta_{gc},\Theta_c), \tag{4.23}$$

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_c] = \boldsymbol{\beta}(\Theta_c), \tag{4.24}$$

$$E(\widehat{\boldsymbol{\beta}}_{xgc}) = \boldsymbol{\beta}, \tag{4.25}$$

$$E(\widehat{\boldsymbol{\beta}}_{gc}) = \boldsymbol{\beta}, \tag{4.26}$$

$$E(\widehat{\boldsymbol{\beta}}_{c}) = \boldsymbol{\beta}. \tag{4.27}$$

Proof: Expressions (4.16)-(4.21) are notations. Expression (4.22) can be easily proved using assumptions (A6) and (4.8)

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c] = (\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\, E(\boldsymbol{Y}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c)$$

$$= (\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\,\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)$$

$$= \boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c).$$

From (4.17), we can easily get (4.23)

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{gc},\Theta_c] = E\left[E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c]|\Theta_{gc},\Theta_c\right]$$

$$= E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c]$$

$$= \boldsymbol{\beta}(\Theta_{gc},\Theta_c),$$

while from (4.19) we can prove (4.24)

$$E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_c] = E\Big[E[\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{gc},\Theta_c]|\Theta_c\Big]$$

$$= E[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c] \;=\; \boldsymbol{\beta}(\Theta_c).$$

Similarly, the proof of (4.25) is direct from (4.21) and (4.24)

$$E(\widehat{\boldsymbol{\beta}}_{xgc}) = E\Big[E(\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_c)\Big]$$

$$= E[\boldsymbol{\beta}(\Theta_c)] \;=\; \boldsymbol{\beta},$$

while (4.26) follows from (4.10) and (4.25)

$$E(\widehat{\boldsymbol{\beta}}_{gc}) = (\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1} \sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc} E(\widehat{\boldsymbol{\beta}}_{xgc}) \;=\; \boldsymbol{\beta}.$$

Finally, (4.27) can be directly proved using (4.11) and (4.26)

$$E(\widehat{\boldsymbol{\beta}}_{c}) = (\sum_{g=1}^{Gc} \boldsymbol{K}_{gc})^{-1} \sum_{g=1}^{Gc} \boldsymbol{K}_{gc} E(\widehat{\boldsymbol{\beta}}_{gc}) \;=\; \boldsymbol{\beta}.$$

□

Theorem 4.3. *Based on Theorem 4.2, we can prove the following expressions for the conditional covariances*

$$Cov[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = \delta_{cc'}[\delta_{gg'}(\delta_{xx'}\boldsymbol{A}+\boldsymbol{U})+\boldsymbol{\Psi}], \tag{4.28}$$

$$Cov[\boldsymbol{\beta}(\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = \delta_{cc'}(\delta_{gg'}\boldsymbol{U}+\boldsymbol{\Psi}), \tag{4.29}$$

$$Cov[\boldsymbol{\beta}(\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = \delta_{cc'}\boldsymbol{\Psi}, \tag{4.30}$$

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}_{xgc},\hat{\boldsymbol{\beta}}_{x'g'c'}) &= Cov(\hat{\boldsymbol{\beta}}_{x'g'c'},\hat{\boldsymbol{\beta}}_{xgc}) \\ &= \delta_{cc'}\{\delta_{gg'}[\delta_{xx'}(\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1})+\boldsymbol{U}]+\boldsymbol{\Psi}\}, \end{aligned} \tag{4.31}$$

$$Cov(\hat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{x'g'c'}) = \delta_{cc'}\{\delta_{gg'}[(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\,\boldsymbol{A}+\,\boldsymbol{U}]+\boldsymbol{\Psi}\}, \tag{4.32}$$

$$Cov(\hat{\boldsymbol{\beta}}_{xgc}, \hat{\boldsymbol{\beta}}_{g'c'}) = \delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\ (\sum_{x'=x_1}^{x_{k_{gc}}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U}] + \boldsymbol{\Psi}\}, \tag{4.33}$$

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{g'c'}) &= \delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\ (\sum_{x'=x_1}^{x_{k_{gc}}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U}] + \boldsymbol{\Psi}\} \\ &= \delta_{cc'}\{\delta_{gg'}[(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\boldsymbol{A} + \boldsymbol{U}] + \boldsymbol{\Psi}\}, \end{aligned} \tag{4.34}$$

$$Cov(\widehat{\boldsymbol{\beta}}_{c}, \widehat{\boldsymbol{\beta}}_{x'g'c'}) = \delta_{cc'}[(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\boldsymbol{U} + \boldsymbol{\Psi}], \tag{4.35}$$

$$Cov(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{c'}) = \delta_{cc'}[\boldsymbol{U}\ (\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'})^{-1} + \boldsymbol{\Psi}], \tag{4.36}$$

$$Cov(\hat{\boldsymbol{\beta}}_{c}, \hat{\boldsymbol{\beta}}_{g'c'}) = \delta_{cc'}[(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\ \boldsymbol{U} + \boldsymbol{\Psi}], \tag{4.37}$$

$$Cov(\hat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{c'}) = \delta_{cc'}[\boldsymbol{U}\ (\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'})^{-1} + \boldsymbol{\Psi}], \tag{4.38}$$

$$Cov(\hat{\boldsymbol{\beta}}_{c}, \hat{\boldsymbol{\beta}}_{c'}) = \delta_{cc'}[\boldsymbol{U}\ (\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'})^{-1} + \boldsymbol{\Psi}] = \delta_{cc'}[(\sum_{c=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\boldsymbol{U} + \boldsymbol{\Psi}], \tag{4.39}$$

$$Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{c'}) = (\sum_{c=1}^{C} \boldsymbol{K}_{c})^{-1}\boldsymbol{\Psi}, \tag{4.40}$$

$$Cov(\hat{\boldsymbol{\beta}}_{c}, \hat{\boldsymbol{\beta}}) = \boldsymbol{\Psi}(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'})^{-1}, \tag{4.41}$$

$$Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = (\sum_{c=1}^{C} \boldsymbol{K}_{c})^{-1}\boldsymbol{\Psi} = \boldsymbol{\Psi}(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'})^{-1}. \tag{4.42}$$

Proof: For the proof of Theorem 4.3, we use the following properties of transpose matrices.

- The transpose of the inverse of a square matrix $\boldsymbol{M}$ equals the inverse of the transpose of the same matrix, i.e. $(\boldsymbol{M}^{-1})' = (\boldsymbol{M}')^{-1}$,
- The transpose of the inverse sum of matrices is the inverse sum of transposes, that is $\left[\left(\sum \boldsymbol{M}\right)^{-1}\right]' = \left[\left(\sum \boldsymbol{M}'\right)\right]^{-1}$.

Hence, (4.28) is proved as follows

$$\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\}$$

$$+ \mathrm{Cov}\{E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\}$$

$$= E\Big\{\mathrm{Cov}\{E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\}$$

$$+E\Big\{E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\} + \delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$$

$$= \delta_{cc'}\delta_{gg'}E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c]\}$$

$$+ E\Big\{E\Big\{\mathrm{Cov}\Big\{E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{xgc},\Theta_{gc},\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{xgc},\Theta_{gc},\Theta_c]|\Theta_{gc},\Theta_c\Big\}|\Theta_c\Big\}\Big\}$$

$$+ E\Big\{E\Big\{E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{xgc},\Theta_{gc},\Theta_c]|\Theta_{gc},\Theta_c\}|\Theta_c\Big\}\Big\} + \delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$$

$$= \delta_{cc'}\delta_{gg'}E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c]\}$$

$$+ \delta_{cc'}\delta_{gg'}\delta_{xx'}E\Big\{E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\}$$

$$+0+ \delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$$

$$= \delta_{cc'}\delta_{gg'}\boldsymbol{U} + \delta_{cc'}\delta_{gg'}\delta_{xx'}\boldsymbol{A} + \delta_{cc'}\boldsymbol{\Psi}$$

$$= \delta_{cc'}[\delta_{gg'}(\delta_{xx'}\boldsymbol{A} + \boldsymbol{U}) + \boldsymbol{\Psi}].$$

Similarly, we can obtain (4.29) and (4.30) as

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] &= E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\} \\
&\quad + \mathrm{Cov}\{E[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\} \\
&= E\Big\{\mathrm{Cov}\{E[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\} \\
&\quad + E\Big\{E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\} \\
&\quad + \delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)] \\
&= \delta_{cc'}\delta_{gg'}E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c]\}+0+\ \delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)] \\
&= \delta_{cc'}(\delta_{gg'}\boldsymbol{U}+\boldsymbol{\Psi}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] &= E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c),\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\}+\ \mathrm{Cov}\{E[\boldsymbol{\beta}(\Theta_c)|\Theta_c],E[\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c]\} \\
&= 0+\delta_{cc'}\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)] \\
&= \delta_{cc'}\boldsymbol{\Psi}.
\end{aligned}$$

Expression (4.31) can be proved by

$$\begin{aligned}
\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{x'g'c'}) &= E[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c)]+\mathrm{Cov}[E(\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_c],E(\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_c)] \\
&= E\left\{\mathrm{Cov}\left[E(\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{gc},\Theta_c),E(\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c)\right]|\Theta_c\right\} \\
&\quad + E\left\{E\left[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{gc},\Theta_c)|\Theta_c\right]\right\}+\delta_{cc'}\ \mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)] \\
&= \delta_{cc'}\delta_{gg'}E\{\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc},\Theta_c)|\Theta_c]\} \\
&\quad +E\Big\{E\{\mathrm{Cov}[E(\widehat{\boldsymbol{\beta}}_{xgc}|\Theta_{xgc},\Theta_{gc},\Theta_c),E(\widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{xgc},\Theta_{gc},\Theta_c)|\Theta_{gc},\Theta_c]|\Theta_c\}\Big\}
\end{aligned}$$

$$+ E\left\{E\left\{E\left[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{x'g'c'}|\Theta_{xgc}, \Theta_{gc}, \Theta_c)|\Theta_{gc}, \Theta_c\right]|\Theta_c\right\}\right\} + \delta_{cc'}\ \mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$$

$$= \delta_{cc'}\delta_{gg'}\ \boldsymbol{U}$$

$$+\ \delta_{cc'}\delta_{gg'}\delta_{xx'}\ \boldsymbol{A}$$

$$+\ E\left\{E\left\{E\left[\sigma^2(\Theta_{xgc}, \Theta_{gc}, \Theta_c)(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}|\Theta_{gc}, \Theta_c\right]|\Theta_c\right\}\right\}$$

$$+\ \delta_{cc'}\ \mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)]$$

$$= \delta_{cc'}\delta_{gg'}\boldsymbol{U} + \delta_{cc'}\delta_{gg'}\delta_{xx'}\boldsymbol{A} + \delta_{cc'}\delta_{gg'}\delta_{xx'}\ s^2\ (\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1} + \delta_{cc'}\boldsymbol{\Psi}$$

$$= \delta_{cc'}\left\{\delta_{gg'}\left[\delta_{xx'}(\boldsymbol{A} + s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}) + \boldsymbol{U}\right] + \boldsymbol{\Psi}\right\}$$

$$= \delta_{cc'}\left\{\delta_{gg'}\left[\delta_{xx'}(\boldsymbol{A} + s^2(\boldsymbol{Z}'_{x'g'c'}\boldsymbol{W}_{x'g'c'}\boldsymbol{Z}_{x'g'c'})^{-1}) + \boldsymbol{U}\right] + \boldsymbol{\Psi}\right\} = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{x'g'c'}, \widehat{\boldsymbol{\beta}}_{xgc}).$$

Expression (4.31) can also be proved by using the covariance property $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{x'g'c'}, \widehat{\boldsymbol{\beta}}_{xgc}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{x'g'c'})'$ and taking into account the symmetry of matrices $\boldsymbol{A}, \boldsymbol{U}$ and $\boldsymbol{\Psi}$. For (4.32) and (4.33) we have

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{x'g'c'}) = \mathrm{Cov}\left[\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{x'g'c'}\right]$$

$$= \left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left[\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{x'g'c'})\right]$$

$$= \left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left\{\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\ \delta_{cc'}\{\delta_{gg'}[\delta_{xx'}(\boldsymbol{A} + s^2(\boldsymbol{Z}'_{xg}\boldsymbol{W}_{xg}\boldsymbol{Z}_{xg})^{-1}) + \boldsymbol{U}] + \boldsymbol{\Psi}\}\right\}$$

$$= \left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left[\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\ \delta_{cc'}\delta_{gg'}\delta_{xx'}(\boldsymbol{A} + s^2(\boldsymbol{Z}'_{xg}\boldsymbol{W}_{xg}\boldsymbol{Z}_{xg})^{-1})\right]$$

$$+\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\ \delta_{cc'}\delta_{gg'}\boldsymbol{U}\right)+\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\ \delta_{cc'}\boldsymbol{\Psi}\right)$$

$$=\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\Big[\delta_{cc'}\delta_{gg'}\ \boldsymbol{A}\ [\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}]^{-1}(\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xg}\boldsymbol{W}_{xg}\boldsymbol{Z}_{xg})^{-1})\Big]$$

$$+\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\left[\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}(\delta_{cc'}\delta_{gg'}\boldsymbol{U}+\delta_{cc'}\boldsymbol{\Psi})\right]$$

$$=\delta_{cc'}\left\{\delta_{gg'}\left[\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\boldsymbol{A}+\ \boldsymbol{U}\right]+\boldsymbol{\Psi}\right\},$$

$$\operatorname{Cov}(\hat{\boldsymbol{\beta}}_{xgc},\hat{\boldsymbol{\beta}}_{g'c'})=\operatorname{Cov}\left[\widehat{\boldsymbol{\beta}}_{xgc},\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{x'g'c'}\right)^{-1}\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{x'g'c'}\widehat{\boldsymbol{\beta}}_{x'g'c'}\right]$$

$$=E\left[(\widehat{\boldsymbol{\beta}}_{xgc}-\boldsymbol{\beta})\sum_{x'=x_1}^{x_{k_{gc}}}(\widehat{\boldsymbol{\beta}}_{x'g'c'}-\boldsymbol{\beta})'\boldsymbol{K}'_{x'g'c'}\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}\right]$$

$$=\sum_{x=x_1}^{x_{k_{gc}}}E\left[(\widehat{\boldsymbol{\beta}}_{xgc}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{x'g'c'}-\boldsymbol{\beta})'\right]\boldsymbol{K}'_{x'g'c'}\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\left[\sum_{x'=x_1}^{x_{k_{gc}}}\operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{x'g'c'}\right)\boldsymbol{K}'_{x'g'c}\right]\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\left[\sum_{x'=x_1}^{x_{k_{gc}}}\delta_{cc'}\left\{\delta_{gg'}\left[\delta_{xx'}(\boldsymbol{A}+s^2(\boldsymbol{Z}'_{x'g'c'}\boldsymbol{W}_{x'g'c'}\boldsymbol{Z}_{x'g'c'})^{-1})+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\}\boldsymbol{K}'_{x'g'c}\right]\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\left[\delta_{cc'}\delta_{gg'}[\boldsymbol{A}+s^2(\boldsymbol{Z}'_{x'g'c'}\boldsymbol{W}_{x'g'c'}\boldsymbol{Z}_{x'g'c'})^{-1}]\boldsymbol{K}'_{x'g'c}\right]\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$+\left(\delta_{cc'}\delta_{gg'}\boldsymbol{U}\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c}\right)\left(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}+\left(\delta_{cc'}\boldsymbol{\Psi}\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c}\right)\left(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\delta_{cc'}\delta_{gg'}\boldsymbol{A}\left(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}\delta_{cc'}\delta_{gg'}\boldsymbol{U}+\delta_{cc'}\boldsymbol{\Psi}$$

$$=\delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\,(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'})^{-1}+\boldsymbol{U}]+\boldsymbol{\Psi}\}.$$

Expression (4.33) can also be directly proved by (4.32) as follows

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{xgc},\hat{\boldsymbol{\beta}}_{g'c'})&=\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{g'c'},\hat{\boldsymbol{\beta}}_{xgc})'\\&=\delta_{cc'}\left\{\delta_{gg'}\left[\left[\left(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}_{x'g'c'}\right)^{-1}\boldsymbol{A}\right]'+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\}\\&=\delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\,(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'})^{-1}+\boldsymbol{U}]+\boldsymbol{\Psi}\}.\end{aligned}$$

Expression (4.34) is obtained as

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{g'c'})&=\mathrm{Cov}\left[\left(\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\right)^{-1}\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{g'c'}\right]\\&=\left(\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\right)^{-1}\left[\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc},\widehat{\boldsymbol{\beta}}_{g'c'})\right]\\&=\left(\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\right)^{-1}\left\{\sum_{x=x_1}^{x_{kgc}}\boldsymbol{K}_{xgc}\ \delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\,(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'})^{-1}+\boldsymbol{U}]+\boldsymbol{\Psi}\}\right\}\\&=\delta_{cc'}\{\delta_{gg'}[\boldsymbol{A}\,(\sum_{x'=x_1}^{x_{kgc}}\boldsymbol{K}'_{x'g'c'})^{-1}+\boldsymbol{U}]+\boldsymbol{\Psi}\}.\end{aligned}$$

Symmetry of (4.34) can be proved from

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{g'c'},\widehat{\boldsymbol{\beta}}_{gc})=\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc},\widehat{\boldsymbol{\beta}}_{g'c'})'=\delta_{cc'}\left\{\delta_{gg'}\left[\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{x'g'c'}\right)^{-1}\boldsymbol{A}+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\},$$

taking into account the symmetry of matrices $\boldsymbol{A},\boldsymbol{U}$ and $\boldsymbol{\Psi}$ or by using (4.32) as

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{g'c'})=\mathrm{Cov}\left[\widehat{\boldsymbol{\beta}}_{gc},\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{x'g'c'}\right)^{-1}\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{x'g'c'}\widehat{\boldsymbol{\beta}}_{x'g'c'}\right]$$

$$=E\left[(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})\sum_{x'=x_1}^{x_{k_{gc}}}(\widehat{\boldsymbol{\beta}}_{x'g'c'}-\boldsymbol{\beta})'\boldsymbol{K}'_{x'g'c'}\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}\right]$$

$$=\sum_{x'=x_1}^{x_{k_{gc}}}E\left[(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{x'g'c'}-\boldsymbol{\beta})'\right]\boldsymbol{K}'_{x'g'c'}\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\left[\sum_{x'=x_1}^{x_{k_{gc}}}\mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_{gc},\widehat{\boldsymbol{\beta}}_{x'g'c'}\right)\boldsymbol{K}'_{x'g'c'}\right]\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\left[\sum_{x'=x_1}^{x_{k_{gc}}}\delta_{cc'}\left\{\delta_{gg'}\left[\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\boldsymbol{A}+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\}\boldsymbol{K}'_{x'g'c'}\right]\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\delta_{cc'}\left\{\delta_{gg'}\left[\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\boldsymbol{A}+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\}\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)\left(\sum_{x'=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{x'g'c'}\right)^{-1}$$

$$=\delta_{cc'}\left\{\delta_{gg'}\left[\left(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\right)^{-1}\boldsymbol{A}+\boldsymbol{U}\right]+\boldsymbol{\Psi}\right\}.$$

From expression (4.32), we can get (4.35) as

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{c},\widehat{\boldsymbol{\beta}}_{x'g'c'})=\mathrm{Cov}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\widehat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{x'g'c'}\right]$$

$$= \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{x'g'c'})$$

$$= \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{x'g'c'})$$

$$= \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big)^{-1} \Big\{\boldsymbol{K}_{gc}\,\delta_{cc'}\Big[(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\boldsymbol{A} + \boldsymbol{U}\Big] + \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\boldsymbol{\Psi}\Big)\Big\}$$

$$= \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big)^{-1} \Big\{\boldsymbol{U}\Big[\boldsymbol{U} + \Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A}\Big]^{-1} \delta_{cc'}\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A} + \boldsymbol{U}\Big]\Big\}$$

$$+ \Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big)^{-1} \delta_{cc'}\Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\boldsymbol{\Psi}\Big)$$

$$= \delta_{cc'}\Big[(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\boldsymbol{U} + \boldsymbol{\Psi}\Big].$$

Expression (4.36) can be proved using

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{x'g'c'}, \widehat{\boldsymbol{\beta}}_{c}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{c}, \widehat{\boldsymbol{\beta}}_{x'g'c'})' = \delta_{cc'}\Big[\boldsymbol{U}(\sum_{g=1}^{G_c} \boldsymbol{K}'_{gc})^{-1} + \boldsymbol{\Psi}\Big],$$

exploiting symmetry of matrices $\boldsymbol{A}, \boldsymbol{U}$ and $\boldsymbol{\Psi}$ or alternatively by using (4.33)

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{c'}) = \mathrm{Cov}\Big[\widehat{\boldsymbol{\beta}}_{xgc}, \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}_{g'c'}\Big)^{-1} \sum_{g'=1}^{G_c} \boldsymbol{K}_{g'c'}\hat{\boldsymbol{\beta}}_{g'c'}\Big]$$

$$= E\Big[(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta}) \sum_{g'=1}^{G_c} (\widehat{\boldsymbol{\beta}}_{g'c'} - \boldsymbol{\beta})' \boldsymbol{K}'_{g'c'} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'}\Big)^{-1}\Big]$$

$$= \sum_{g'=1}^{G_c} E\Big[(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{g'c'} - \boldsymbol{\beta})'\Big] \boldsymbol{K}'_{g'c'} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'}\Big)^{-1}$$

$$= \sum_{g'=1}^{G_c} \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{g'c'}) \boldsymbol{K}'_{g'c'} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right)^{-1}$$

$$= \left[\sum_{g'=1}^{G_c} \delta_{cc'} \{ \delta_{gg'} [\boldsymbol{A} \, (\sum_{x'=x_1}^{x_{kgc}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U}] + \boldsymbol{\Psi} \} \boldsymbol{K}'_{g'c'} \right] \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right)^{-1}$$

$$= \left\{ \delta_{cc'} \left[\boldsymbol{A} \, (\sum_{x'=x_1}^{x_{kgc}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U} \right] \boldsymbol{K}'_{g'c'} + \delta_{cc'} \, \boldsymbol{\Psi} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right) \right\} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right)^{-1}$$

$$= \delta_{cc'} \, \boldsymbol{U} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right)^{-1} + \delta_{cc'} \, \boldsymbol{\Psi}$$

$$= \delta_{cc'} \left[\boldsymbol{U} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \right)^{-1} + \boldsymbol{\Psi} \right].$$

Expression (4.37) is obtained by (4.34) as follows

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}}_{g'c'}) = \mathrm{Cov}\left[\left(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \right)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \widehat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{g'c'} \right]$$

$$= \left(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \right)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{g'c'})$$

$$= \left(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \right)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \delta_{cc'} \left\{ \delta_{gg'} \left[\left(\sum_{x=x_1}^{x_{kgc}} \boldsymbol{K}_{xgc} \right)^{-1} \boldsymbol{A} + \boldsymbol{U} \right] + \boldsymbol{\Psi} \right\}$$

$$= \left(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \right)^{-1} \boldsymbol{U} \left[\boldsymbol{U} + \left(\sum_{x=x_1}^{x_{kgc}} \boldsymbol{K}_{xgc} \right)^{-1} \boldsymbol{A} \right]^{-1} \delta_{cc'} \left[\left(\sum_{x=x_1}^{x_{kgc}} \boldsymbol{K}_{xgc} \right)^{-1} \boldsymbol{A} + \boldsymbol{U} \right]$$

$$+ \left(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \right)^{-1} \sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \delta_{cc'} \boldsymbol{\Psi}$$

$$= \delta_{cc'} \Big[\Big(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc} \Big)^{-1} \boldsymbol{U} + \boldsymbol{\Psi} \Big].$$

and from $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{g'c'}, \widehat{\boldsymbol{\beta}}_{c}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{c}, \widehat{\boldsymbol{\beta}}_{g'c'})' = \delta_{cc'} \Big[\boldsymbol{U} \, \Big(\sum_{g=1}^{G_c} \boldsymbol{K}'_{gc} \Big)^{-1} + \boldsymbol{\Psi} \Big]$, expression (4.38) yields from

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \widehat{\boldsymbol{\beta}}_{c'}) = \delta_{cc'} \Big[\boldsymbol{U} \, \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1} + \boldsymbol{\Psi} \Big],$$

or by using (4.34)

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{gc}, \hat{\boldsymbol{\beta}}_{c'}) = \mathrm{Cov}\Bigg[\widehat{\boldsymbol{\beta}}_{gc}, \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}_{g'c'} \Big)^{-1} \sum_{g'=1}^{G_c} \boldsymbol{K}_{g'c'} \hat{\boldsymbol{\beta}}_{g'c'} \Bigg]$$

$$= E\Bigg[(\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta}) \sum_{g'=1}^{G_c} (\widehat{\boldsymbol{\beta}}_{g'c'} - \boldsymbol{\beta})' \boldsymbol{K}'_{g'c'} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1} \Bigg]$$

$$= \sum_{g'=1}^{G_c} E\Big[(\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{g'c'} - \boldsymbol{\beta})' \Big] \boldsymbol{K}'_{g'c'} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1}$$

$$= \Bigg[\sum_{g'=1}^{G_c} \mathrm{Cov}\Big(\widehat{\boldsymbol{\beta}}_{gc}, \widehat{\boldsymbol{\beta}}_{g'c'} \Big) \boldsymbol{K}'_{g'c'} \Bigg] \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1}$$

$$= \Bigg[\sum_{g'=1}^{G_c} \delta_{cc'} \{ \delta_{gg'} [\boldsymbol{A} \, (\sum_{x'=x_1}^{x_{k_{gc}}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U}] + \boldsymbol{\Psi} \} \boldsymbol{K}'_{g'c'} \Bigg] \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1}$$

$$= \Bigg\{ \delta_{cc'} \Big[\boldsymbol{A} \, (\sum_{x'=x_1}^{x_{k_{gc}}} \boldsymbol{K}'_{x'g'c'})^{-1} + \boldsymbol{U} \Big] \boldsymbol{K}'_{g'c'} + \delta_{cc'} \, \boldsymbol{\Psi} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big) \Bigg\} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1}$$

$$= \delta_{cc'} \, \boldsymbol{U} \Big(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'} \Big)^{-1} + \delta_{cc'} \, \boldsymbol{\Psi}$$

$$= \delta_{cc'}\left[\boldsymbol{U}\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}+\boldsymbol{\Psi}\right].$$

From (4.38), we can prove (4.39)

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_c,\hat{\boldsymbol{\beta}}_{c'}) = \mathrm{Cov}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\widehat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{c'}\right]$$

$$= \left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc},\hat{\boldsymbol{\beta}}_{c'})$$

$$= \left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\delta_{cc'}\left[\boldsymbol{U}\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}+\boldsymbol{\Psi}\right]$$

$$= \delta_{cc'}\left[\boldsymbol{U}\left(\sum_{g'=1}^{H_c}\boldsymbol{K}'_{g'c'}\right)^{-1}+\boldsymbol{\Psi}\right]$$

and if we use $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{c'},\widehat{\boldsymbol{\beta}}_{c}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{c},\widehat{\boldsymbol{\beta}}_{c'})' = \delta_{cc'}\left[\left(\sum_{g'=1}^{H_c}\boldsymbol{K}_{g'c'}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]$, the symmetry of (4.39) is proved from

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c,\widehat{\boldsymbol{\beta}}_{c'}) = \delta_{cc'}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right],$$

or alternatively using (4.37)

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_c,\hat{\boldsymbol{\beta}}_{c'}) = \mathrm{Cov}\left[\widehat{\boldsymbol{\beta}}_c,\left(\sum_{g'=1}^{G_c}\boldsymbol{K}_{g'c'}\right)^{-1}\sum_{g'=1}^{G_c}\boldsymbol{K}_{g'c'}\hat{\boldsymbol{\beta}}_{g'c'}\right]$$

$$= E\left[(\widehat{\boldsymbol{\beta}}_c-\boldsymbol{\beta})\sum_{g'=1}^{G_c}(\widehat{\boldsymbol{\beta}}_{g'c'}-\boldsymbol{\beta})'\boldsymbol{K}'_{g'c'}\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}\right]$$

$$= \sum_{g'=1}^{G_c}E\left[(\widehat{\boldsymbol{\beta}}_c-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{g'c'}-\boldsymbol{\beta})'\right]\boldsymbol{K}'_{g'c'}\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}$$

$$= \left[\sum_{g'=1}^{G_c} \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_{g'c'}\right)\boldsymbol{K}'_{g'c'}\right]\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}$$

$$= \left[\sum_{g'=1}^{G_c} \delta_{cc'}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]\boldsymbol{K}'_{g'c'}\right]\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}$$

$$= \delta_{cc'}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)\left(\sum_{g'=1}^{G_c}\boldsymbol{K}'_{g'c'}\right)^{-1}$$

$$= \delta_{cc'}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right].$$

Similarly, (4.40) can be obtained using (4.39) as follows

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{c'}) = \mathrm{Cov}\left[\left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\sum_{c=1}^{C}\boldsymbol{K}_c\widehat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}}_{c'}\right]$$

$$= \left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\left[\sum_{c=1}^{C}\boldsymbol{K}_c\,\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}}_{c'})\right]$$

$$= \left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\left\{\sum_{c=1}^{C}\boldsymbol{K}_c\,\delta_{cc'}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]\right\}$$

$$= \left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\left\{\boldsymbol{K}_c\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]\right\}$$

$$= \left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\left\{\boldsymbol{\Psi}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]^{-1}\left[\left(\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\right)^{-1}\boldsymbol{U}+\boldsymbol{\Psi}\right]\right\}$$

$$= \left(\sum_{c=1}^{C}\boldsymbol{K}_c\right)^{-1}\boldsymbol{\Psi},$$

while (4.41) can be obtained as

$$\operatorname{Cov}(\hat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}}) = \operatorname{Cov}\left[\widehat{\boldsymbol{\beta}}_c, \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}\right)^{-1} \sum_{c'=1}^{C} \boldsymbol{K}_{c'} \widehat{\boldsymbol{\beta}}_{c'}\right]$$

$$= E\left[(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) \sum_{c'=1}^{C} (\widehat{\boldsymbol{\beta}}_{c'} - \boldsymbol{\beta})' \boldsymbol{K}'_{c'} \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}\right]$$

$$= \sum_{c'=1}^{C} E\left[(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{c'} - \boldsymbol{\beta})'\right] \boldsymbol{K}'_{c'} \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}$$

$$= \left[\sum_{c'=1}^{C} \operatorname{Cov}\left(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_{c'}\right) \boldsymbol{K}'_{c'}\right] \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}$$

$$= \left[\sum_{c'=1}^{C} \delta_{cc'} \left[\boldsymbol{U} \left(\sum_{g'=1}^{H_c} \boldsymbol{K}'_{g'c'}\right)^{-1} + \boldsymbol{\Psi}\right] \boldsymbol{K}'_{c'}\right] \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}$$

$$= \left[\sum_{c'=1}^{C} \delta_{cc'} \left[\boldsymbol{U} \left(\sum_{g'=1}^{H_c} \boldsymbol{K}'_{g'c'}\right)^{-1} + \boldsymbol{\Psi}\right] \left[\boldsymbol{U} \left(\sum_{g'=1}^{G_c} \boldsymbol{K}'_{g'c'}\right)^{-1} + \boldsymbol{\Psi}\right]^{-1} \boldsymbol{\Psi}\right] \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}$$

$$= \boldsymbol{\Psi} \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1},$$

or alternatively using property

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}) = \operatorname{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_c)' = \boldsymbol{\Psi} \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}.$$

Finally, (4.42) can be proved from (4.40)

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}) = \operatorname{Cov}\left[\widehat{\boldsymbol{\beta}}, \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}\right)^{-1} \sum_{c'=1}^{C} \boldsymbol{K}_{c'} \widehat{\boldsymbol{\beta}}_{c'}\right]$$

$$= E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sum_{c'=1}^{C} (\widehat{\boldsymbol{\beta}}_{c'} - \boldsymbol{\beta})' \boldsymbol{K}'_{c'} \left(\sum_{c'=1}^{C} \boldsymbol{K}'_{c'}\right)^{-1}\right]$$

$$= \sum_{c'=1}^{C} E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{c'} - \boldsymbol{\beta})'\right] \boldsymbol{K}_{c'}' \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}$$

$$= \left[\sum_{c'=1}^{C} \mathrm{Cov}\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_{c'}\right) \boldsymbol{K}_{c'}'\right] \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}$$

$$= \left[\sum_{c'=1}^{C} \left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \boldsymbol{\Psi} \boldsymbol{K}_{c'}'\right] \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}$$

$$= \left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \boldsymbol{\Psi} \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right) \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}$$

$$= \left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \boldsymbol{\Psi},$$

while the symmetric expression of (4.41) is given by

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}) = \mathrm{Cov}\left[\left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \sum_{c=1}^{C} \boldsymbol{K}_c\, \widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}\right]$$

$$= \left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \left[\sum_{c=1}^{C} \boldsymbol{K}_c\, \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \hat{\boldsymbol{\beta}})\right]$$

$$= \left(\sum_{c=1}^{C} \boldsymbol{K}_c\right)^{-1} \left[\sum_{c=1}^{C} \boldsymbol{K}_c\, \boldsymbol{\Psi} \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}\right]$$

$$= \boldsymbol{\Psi} \left(\sum_{c'=1}^{C} \boldsymbol{K}_{c'}'\right)^{-1}.$$

□

4.3.2 Estimation of Parameters

In this section, we present the linear credibility estimators for each level (country-gender-age) of our hierarchical regression model. For Level 3 (country), we consider

$$\boldsymbol{Y}_c^{Cred} = \boldsymbol{Z}_c\, \boldsymbol{\beta}_c^{Cred}(\Theta_c), \tag{4.43}$$

where $\boldsymbol{Y}_c^{Cred} = (\boldsymbol{Y}_1^{Cred'}, \ldots, \boldsymbol{Y}_C^{Cred'})'$ and

$$\boldsymbol{\beta}_c^{Cred}(\Theta_c) = \boldsymbol{K}_c\, \widehat{\boldsymbol{\beta}}_c + (\boldsymbol{I} - \boldsymbol{K}_c)\, \boldsymbol{\beta}\ , \tag{4.44}$$

with $\boldsymbol{K}_c$ defined by (4.15) and $\boldsymbol{I}$ the 2×2 identity matrix.
For Level 2 (gender), we have

$$\boldsymbol{Y}_{gc}^{Cred} = \boldsymbol{Z}_{gc}\, \boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c), \tag{4.45}$$

where $\boldsymbol{Y}_{gc}^{Cred} = (\boldsymbol{Y}_{1c}^{Cred'}, \ldots, \boldsymbol{Y}_{G_c c}^{Cred'})'$ and

$$\boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c) = \boldsymbol{K}_{gc}\, \widehat{\boldsymbol{\beta}}_{gc} + (\boldsymbol{I} - \boldsymbol{K}_{gc})\, \boldsymbol{\beta}_c^{Cred}(\Theta_c)\ , \tag{4.46}$$

with $\boldsymbol{K}_{gc}$ obtained from (4.14). For Level 1 (age), estimator is defined by

$$\boldsymbol{Y}_{xgc}^{Cred} = \boldsymbol{Z}_{xgc}\, \boldsymbol{\beta}_{xgc}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c), \tag{4.47}$$

where $\boldsymbol{Y}_{xgc}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c) = (Y_{t_1,xgc}^{Cred}, \ldots, Y_{t_n,xgc}^{Cred})'$ and

$$\boldsymbol{\beta}_{xgc}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c) = \boldsymbol{K}_{xgc}\, \widehat{\boldsymbol{\beta}}_{xgc} + (\boldsymbol{I} - \boldsymbol{K}_{xgc})\, \boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c)\ , \tag{4.48}$$

with $\boldsymbol{K}_{xgc}$ as it is given by (4.13).

The following theorem reviews the optimal projection results, which imply the best linear estimators in hierarchical credibility regression [De Vylder (1976), Sundt (1979; 1980)]. For the proof, we refer to Bühlmann and Gisler (2005).

Theorem 4.4. *Let $\boldsymbol{\beta}_c^{Cred}(\Theta_c)$ be the linear estimator of $\boldsymbol{\beta}(\Theta_c)$, $\boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c)$ be the linear estimator of $\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)$ and $\boldsymbol{\beta}_{xgc}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$ be the linear estimator of $\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)$. Then, a) $\boldsymbol{\beta}_c^{Cred}(\Theta_c)$ is the best linear credibility estimator of $\boldsymbol{\beta}(\Theta_c)$ if it satisfies*

$$E[\boldsymbol{\beta}_c^{Cred}(\Theta_c)] = E[\boldsymbol{\beta}(\Theta_c)] \tag{4.49}$$

and

$$Cov[\boldsymbol{\beta}_c^{Cred}(\Theta_c), \hat{\boldsymbol{\beta}}_{x'g'c'}] = Cov[\boldsymbol{\beta}(\Theta_c), \hat{\boldsymbol{\beta}}_{x'g'c'}], \tag{4.50}$$

b) $\boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c)$ is the best linear credibility estimator of $\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)$ if it satisfies

$$E[\boldsymbol{\beta}_{gc}^{Cred}(\Theta_{gc}, \Theta_c)] = E[\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)] \tag{4.51}$$

and

$$Cov[\boldsymbol{\beta}^{Cred}_{gc}(\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = Cov[\boldsymbol{\beta}(\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}], \tag{4.52}$$

c) $\boldsymbol{\beta}^{Cred}_{xgc}(\Theta_{xgc},\Theta_{gc},\Theta_c)$ *is the best linear credibility estimator of* $\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)$ *if it satisfies*

$$E[\boldsymbol{\beta}^{Cred}_{xgc}(\Theta_{xgc},\Theta_{gc},\Theta_c)] = E[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c)] \tag{4.53}$$

and

$$Cov[\boldsymbol{\beta}^{Cred}_{xgc}(\Theta_{xgc},\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}] = Cov[\boldsymbol{\beta}(\Theta_{xgc},\Theta_{gc},\Theta_c),\hat{\boldsymbol{\beta}}_{x'g'c'}]. \tag{4.54}$$

The next lemma gives the conditions for the best linear credibility estimators of regression coefficients for our hierarchical model.

Lemma 4.5. *Based on* (4.49)-(4.54)*, the best linear estimators for Level 3 (country), Level 2 (gender) and Level 1 (age) are given by*

$$a)\ \boldsymbol{\beta}^{Cred}_{c}(\Theta_c) = \boldsymbol{K}_c\,\widehat{\boldsymbol{\beta}}_c + (\boldsymbol{I}-\boldsymbol{K}_c)\,\boldsymbol{\beta} \quad (Level\ 3), \tag{4.55}$$

$$b)\ \boldsymbol{\beta}^{Cred}_{gc}(\Theta_{gc},\Theta_c) = \boldsymbol{K}_{gc}\,\widehat{\boldsymbol{\beta}}_{gc} + (\boldsymbol{I}-\boldsymbol{K}_{gc})\,\boldsymbol{\beta}^{Cred}_{c}(\Theta_c) \quad (Level\ 2), \tag{4.56}$$

$$c)\ \boldsymbol{\beta}^{Cred}_{xgc}(\Theta_{xgc},\Theta_{gc},\Theta_c) = \boldsymbol{K}_{xgc}\,\widehat{\boldsymbol{\beta}}_{xgc} + (\boldsymbol{I}-\boldsymbol{K}_{xgc})\,\boldsymbol{\beta}^{Cred}_{gc}(\Theta_{gc},\Theta_c) \quad (Level\ 1), \tag{4.57}$$

where $\boldsymbol{K}_c$*,* $\boldsymbol{K}_{gc}$ *and* $\boldsymbol{K}_{xgc}$ *are defined in* (4.15)*,* (4.14) *and* (4.13)*, respectively.*

Proof: a) For country level it is sufficient to show that (4.55) satisfies (4.49) and (4.50) of Theorem (4.4). Expectation unbiasedness holds from (4.27) and (4.21), as follows

$$\begin{aligned} E\left[\boldsymbol{\beta}^{Cred}(\Theta_c)\right] &= E\left[\boldsymbol{K}_c\,\widehat{\boldsymbol{\beta}}_c + (\boldsymbol{I}-\boldsymbol{K}_c)\,\boldsymbol{\beta}\right] \\ &= \boldsymbol{K}_c\,E\left[\widehat{\boldsymbol{\beta}}_c\right] + (\boldsymbol{I}-\boldsymbol{K}_c)\,\boldsymbol{\beta} \\ &= \boldsymbol{\beta} = E\left[\boldsymbol{\beta}(\Theta_c)\right]. \end{aligned} \tag{4.58}$$

The covariance condition is proved using expression (4.35) and (4.30).

$$\mathrm{Cov}[\boldsymbol{\beta}(\Theta_c)^{Cred}, \widehat{\boldsymbol{\beta}}_{x'g'c'}) = \mathrm{Cov}\left[\boldsymbol{K}_c\, \widehat{\boldsymbol{\beta}}_c + (\boldsymbol{I} - \boldsymbol{K}_c)\, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}_{x'g'c'}\right]$$

$$= \boldsymbol{K}_c\ \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_{x'g'c'}) + (\boldsymbol{I} - \boldsymbol{K}_c)\ \mathrm{Cov}(\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}_{x'g'c'})$$

$$= \boldsymbol{K}_c\ \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_{x'g'c'}) + 0$$

$$= \boldsymbol{\Psi}[\boldsymbol{\Psi} + (\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\boldsymbol{U}]^{-1}\delta_{cc'}[(\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1}\boldsymbol{U} + \boldsymbol{\Psi}]$$

$$= \delta_{cc'}\boldsymbol{\Psi} = \mathrm{Cov}[\boldsymbol{\beta}(\Theta_c), \hat{\boldsymbol{\beta}}_{x'g'c'}]. \tag{4.59}$$

b) Similarly, for gender level it is sufficient to show that (4.56) satisfies (4.51) and (4.52). The expectation condition can be proved by (4.26), (4.58), (4.25) and (4.23) as

$$E\left[\boldsymbol{\beta}^{Cred}(\Theta_{gc}, \Theta_c)\right] = E\left[\boldsymbol{K}_{gc}\widehat{\boldsymbol{\beta}}_{gc} + (\boldsymbol{I} - \boldsymbol{K}_{gc})\ \boldsymbol{\beta}^{Cred}(\Theta_c)\right]$$

$$= \boldsymbol{K}_{gc} E\left[\widehat{\boldsymbol{\beta}}_{gc}\right] + (\boldsymbol{I} - \boldsymbol{K}_{gc})\ E\left[\boldsymbol{\beta}^{Cred}(\Theta_c)\right]$$

$$= \boldsymbol{\beta} = E(\widehat{\boldsymbol{\beta}}_{xgc}) = E\left[E(\widehat{\boldsymbol{\beta}}_{xgc} | \Theta_{gc}, \Theta_c)\right] = E\left[\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)\right]. \tag{4.60}$$

Then, from (4.32), (4.61) and (4.29) we get

$$\mathrm{Cov}[\boldsymbol{\beta}(\Theta_{gc}, \Theta_c)^{Cred}, \widehat{\boldsymbol{\beta}}_{x'g'c'}] =$$

$$= \mathrm{Cov}[\boldsymbol{K}_{gc}\widehat{\boldsymbol{\beta}}_{gc} + (\boldsymbol{I} - \boldsymbol{K}_{gc})\ \boldsymbol{\beta}^{Cred}(\Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]$$

$$= \boldsymbol{K}_{gc}\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \widehat{\boldsymbol{\beta}}_{x'g'c'}) + (\boldsymbol{I} - \boldsymbol{K}_{gc})\ \mathrm{Cov}[\boldsymbol{\beta}^{Cred}(\Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]$$

$$= \boldsymbol{U}[\boldsymbol{U} + (\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\boldsymbol{A}]^{-1}\{\delta_{cc'}\{\delta_{gg'}[(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\ \boldsymbol{A} +\ \boldsymbol{U}] + \boldsymbol{\Psi}\}\}$$

$$+ \{\boldsymbol{I} - \boldsymbol{U}[\boldsymbol{U} + (\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc})^{-1}\boldsymbol{A}]^{-1}\}\ \delta_{cc'}\boldsymbol{\Psi}$$

$$= \delta_{cc'}(\delta_{gg'}\boldsymbol{U} + \boldsymbol{\Psi}) = \text{Cov}[\boldsymbol{\beta}(\Theta_{gc}, \Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]. \tag{4.61}$$

c) Finally, for the estimator (4.57) of age level, we have to prove (4.51) and (4.52). For the expectation condition, we use (4.25), (4.60) and (S1) as follows

$$E\left[\boldsymbol{\beta}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)\right] = E\left[\boldsymbol{K}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc} + (\boldsymbol{I} - \boldsymbol{K}_{xgc})\ \boldsymbol{\beta}^{Cred}(\Theta_{gc}, \Theta_c)\right]$$

$$= \boldsymbol{K}_{xgc}E\left[\widehat{\boldsymbol{\beta}}_{xgc}\right] + (\boldsymbol{I} - \boldsymbol{K}_{xgc})\ E\left[\boldsymbol{\beta}^{Cred}(\Theta_{gc}, \Theta_c)\right]$$

$$= \boldsymbol{\beta} = E\left[\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c)\right],$$

while for the covariance, the proof is given by using (4.31), (4.61) and (4.28)

$$\text{Cov}[\boldsymbol{\beta}^{Cred}(\Theta_{xgc}, \Theta_{gc}, \Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]$$

$$= \text{Cov}[\boldsymbol{K}_{xgc}\ \widehat{\boldsymbol{\beta}}_{xgc} + (\boldsymbol{I} - \boldsymbol{K}_{xgc})\ \boldsymbol{\beta}^{Cred}(\Theta_{gc}, \Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]$$

$$= \boldsymbol{K}_{xgc}\text{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{x'g'c'}) + (\boldsymbol{I} - \boldsymbol{K}_{xgc})\ \text{Cov}[\boldsymbol{\beta}^{Cred}(\Theta_{gc}, \Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}]$$

$$= \boldsymbol{K}_{xgc}\delta_{cc'}\{\delta_{gg'}[\delta_{xx'}(\boldsymbol{A} + s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}) + \boldsymbol{U}] + \boldsymbol{\Psi}\}$$

$$+ (\boldsymbol{I} - \boldsymbol{K}_{xgc})\ \delta_{cc'}(\delta_{gg'}\boldsymbol{U} + \boldsymbol{\Psi})$$

$$= \delta_{cc'}[\delta_{gg'}(\delta_{xx'}\boldsymbol{A} + \boldsymbol{U}) + \boldsymbol{\Psi}] = \text{Cov}[\boldsymbol{\beta}(\Theta_{xgc}, \Theta_{gc}, \Theta_c), \widehat{\boldsymbol{\beta}}_{x'g'c'}].$$

□

Theorem 4.6. *The unbiased estimators of the structural parameters s^2, $\boldsymbol{A}$, $\boldsymbol{U}$ and $\boldsymbol{\Psi}$ are given by the following formulas*

$$\widehat{s^2} = \frac{1}{\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}(n-2)}\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}(\boldsymbol{Y}_{xgc} - \boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc})'\boldsymbol{W}_{xgc}(\boldsymbol{Y}_{xgc} - \boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc}), \tag{4.62}$$

$$\widehat{\boldsymbol{A}} = \frac{1}{\sum_{c=1}^{C}\sum_{g=1}^{G_c}(x_{k_{gc}} - x_1)}\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}(\widehat{\boldsymbol{\beta}}_{xgc} - \widehat{\boldsymbol{\beta}}_{gc})(\widehat{\boldsymbol{\beta}}_{xgc} - \widehat{\boldsymbol{\beta}}_{gc})', \tag{4.63}$$

$$\widehat{\boldsymbol{U}} = \frac{1}{\sum_{c=1}^{C}(G_c - 1)} \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}(\widehat{\boldsymbol{\beta}}_{gc} - \widehat{\boldsymbol{\beta}}_{c})(\widehat{\boldsymbol{\beta}}_{gc} - \widehat{\boldsymbol{\beta}}_{c})', \tag{4.64}$$

$$\widehat{\boldsymbol{\Psi}} = \frac{1}{(C-1)} \sum_{c=1}^{C} \boldsymbol{K}_{c}(\widehat{\boldsymbol{\beta}}_{c} - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{c} - \widehat{\boldsymbol{\beta}})'. \tag{4.65}$$

Proof: For the proof of expression (4.62) let

$$S^{\star} = E\left[(\boldsymbol{Y}_{xgc} - \boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc})'\boldsymbol{W}_{xgc}(\boldsymbol{Y}_{xgc} - \boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc})\right]$$

$$= E\Big[\boldsymbol{Y}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Y}_{xgc} - \boldsymbol{Y}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc}$$

$$-\ (\widehat{\boldsymbol{\beta}}_{xgc})'\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Y}_{xgc}\ +\ (\widehat{\boldsymbol{\beta}}_{xgc})'\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\widehat{\boldsymbol{\beta}}_{xgc}\Big]$$

$$= E\left[\boldsymbol{Y}'_{xgc}\left(\boldsymbol{W}_{xgc} - \boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\right)\boldsymbol{Y}_{xgc}\right]$$

$$= E\left\{tr\left[\boldsymbol{Y}'_{xgc}\left(\boldsymbol{W}_{xgc} - \boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\right)\boldsymbol{Y}_{xgc}\right]\right\},$$

as matrix multiplication of above matrices gives a number equal to its trace. Also, we have that $tr(ABC) = tr(BCA)$, since trace is invariant under cyclic permutations of random matrices $A_{(m\times n)}, B_{(n\times n)}, C_{(n\times m)}$. Therefore,

$$S^{\star} = tr\left[\left(\boldsymbol{W}_{xgc} - \boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\right)E\left(\boldsymbol{Y}_{xgc}\boldsymbol{Y}'_{xgc}\right)\right],$$

where

$$E\left(\boldsymbol{Y}_{xgc}\boldsymbol{Y}'_{xgc}\right) = s^2\ \boldsymbol{W}^{-1}_{xgc} + \boldsymbol{Z}_{xgc}\boldsymbol{\beta}\ \boldsymbol{\beta}'\boldsymbol{Z}'_{xgc}\ .$$

By recalling the linearity of trace, we take

$$\begin{aligned} S^{\star} &= tr\left(s^2\boldsymbol{W}'_{xgc}\boldsymbol{W}_{xgc}\right) + tr\left(\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\ \boldsymbol{\beta}\ \boldsymbol{\beta}'\boldsymbol{Z}'_{xgc}\right) \\ &-\ tr\left[\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}s^2\ \boldsymbol{W}^{-1}_{xgc}\right] \\ &-\ tr\left[\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\ \boldsymbol{\beta}\ \boldsymbol{\beta}'\boldsymbol{Z}'_{xgc}\right] \end{aligned}$$

$$= s^2\ tr\Big(\boldsymbol{W}'_{xgc}\boldsymbol{W}_{xgc}) + tr\left(\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\ \boldsymbol{\beta}\ \boldsymbol{\beta}'\boldsymbol{Z}'_{xgc}\right)$$

$$- \; s^2 \, tr\Big[(\boldsymbol{Z}_{xgc}'\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\boldsymbol{Z}_{xgc}'\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc}\Big]$$
$$- \; tr\Big(\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc} \, \boldsymbol{\beta} \, \boldsymbol{\beta}' \boldsymbol{Z}_{xgc}'\Big)$$

$$= s^2 \, \Big[tr(\boldsymbol{I}_n) - tr(\boldsymbol{I}_2)\Big]$$

$$= s^2 \, (n-2).$$

Hence,

$$E(\widehat{s^2}) = \frac{1}{\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}(n-2)} \sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} S^\star = s^2.$$

For the proof of (4.63), we have

$$E[\widehat{\boldsymbol{A}}^\star] = E\left[\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}(\widehat{\boldsymbol{\beta}}_{xgc} - \widehat{\boldsymbol{\beta}_{gc}})(\widehat{\boldsymbol{\beta}}_{xgc} - \hat{\boldsymbol{\beta}_{gc}})'\right]$$

$$= E\left\{\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big[(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta}) - (\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})\Big]\Big[(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta}) - (\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})\Big]'\right\}$$

$$= E\Bigg[\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta})' - (\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})'$$

$$- (\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{xgc} - \boldsymbol{\beta})' + (\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{gc} - \boldsymbol{\beta})'\Bigg]$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{xgc}) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{xgc}, \widehat{\boldsymbol{\beta}}_{gc}) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \widehat{\boldsymbol{\beta}}_{xgc}) + \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc}, \widehat{\boldsymbol{\beta}}_{gc})\Big]$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big[\boldsymbol{A} + s^2(\boldsymbol{Z}_{xgc}'\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1} + \boldsymbol{U} + \boldsymbol{\Psi}\Big]$$

$$- \sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big[\boldsymbol{A}\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}'\Big)^{-1} + \boldsymbol{U} + \boldsymbol{\Psi}\Big]$$

$$-\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big[(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc})^{-1}\boldsymbol{A}+\boldsymbol{U}+\boldsymbol{\Psi}\Big]$$

$$+\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big[\boldsymbol{A}\Big(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}'_{xgc}\Big)^{-1}+\boldsymbol{U}+\boldsymbol{\Psi}\Big]$$

$$=\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big\{\Big[\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}+\boldsymbol{U}+\boldsymbol{\Psi}\Big]-\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A}+\boldsymbol{U}+\boldsymbol{\Psi}\Big]\Big\}$$

$$=\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{A}\Big[\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\Big]^{-1}\Big[\boldsymbol{A}+s^2(\boldsymbol{Z}'_{xgc}\boldsymbol{W}_{xgc}\boldsymbol{Z}_{xgc})^{-1}\Big]$$

$$-\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big(\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A}$$

$$=\sum_{c=1}^{C}\sum_{g=1}^{G_c}\sum_{x=x_1}^{x_{k_{gc}}}\boldsymbol{A}-\sum_{c=1}^{C}\sum_{g=1}^{G_c}\boldsymbol{A}=\sum_{c=1}^{C}\sum_{g=1}^{G_c}(x_{k_{gc}}-x_1)\boldsymbol{A}.$$

Similarly, (4.64) is proved as follows

$$E[\widehat{\boldsymbol{U}}^{\star}]=E\Big[\sum_{c=1}^{C}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}(\widehat{\boldsymbol{\beta}}_{gc}-\hat{\boldsymbol{\beta}}_{c})(\widehat{\boldsymbol{\beta}}_{gc}-\hat{\boldsymbol{\beta}}_{c})'\Big]$$

$$=E\Big\{\sum_{c=1}^{C}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}\Big[(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})-(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})\Big]\Big[(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})-(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})\Big]'\Big\}$$

$$=E\Big[\sum_{c=1}^{C}\sum_{g=1}^{G_c}\boldsymbol{K}_{gc}(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})'-(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})'$$

$$-(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{gc}-\boldsymbol{\beta})'+(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{c}-\boldsymbol{\beta})'\Big]$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc},\widehat{\boldsymbol{\beta}}_{gc}) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{gc},\widehat{\boldsymbol{\beta}_c}) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}_c},\widehat{\boldsymbol{\beta}}_{gc}) + \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c,\widehat{\boldsymbol{\beta}_c})\Big]$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big\{\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A} + \boldsymbol{U} + \boldsymbol{\Psi}\Big] - \Big[\boldsymbol{U}\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}^{'}\Big)^{-1} + \boldsymbol{\Psi}\Big]$$

$$- \Big[\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}\Big)^{-1}\boldsymbol{U} + \boldsymbol{\Psi}\Big] + \Big[\boldsymbol{U}\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}^{'}\Big)^{-1} + \boldsymbol{\Psi}\Big]\Big\}$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big\{\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A} + \boldsymbol{U} + \boldsymbol{\Psi}\Big] - \Big[\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}\Big)^{-1}\boldsymbol{U} + \boldsymbol{\Psi}\Big]\Big\}$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A} + \boldsymbol{U}\Big] - \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{K}_{gc}\Big[\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}\Big)^{-1}\boldsymbol{U}\Big]$$

$$= \sum_{c=1}^{C}\sum_{g=1}^{G_c} \boldsymbol{U}[\boldsymbol{U} + \Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A}]^{-1}\Big[\Big(\sum_{x=x_1}^{x_{k_{gc}}} \boldsymbol{K}_{xgc}\Big)^{-1}\boldsymbol{A} + \boldsymbol{U}\Big] - \sum_{c=1}^{C} \boldsymbol{U}$$

$$= \sum_{c=1}^{C} (G_c - 1)\boldsymbol{U},$$

and finally, (4.65) is obtained as follows

$$E[\widehat{\boldsymbol{\Psi}}^{\star}] = E\Big[\sum_{c=1}^{C} \boldsymbol{K}_c(\widehat{\boldsymbol{\beta}}_c - \widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_c - \widehat{\boldsymbol{\beta}})^{'}\Big]$$

$$= E\Big\{\sum_{c=1}^{C} \boldsymbol{K}_c\Big[(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big]\Big[(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\Big]^{'}\Big\}$$

$$= E\Big[\sum_{c=1}^{C} \boldsymbol{K}_c(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})^{'} - (\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{'}$$

$$- (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})' + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\Big]$$

$$= \sum_{c=1}^{C} \boldsymbol{K}_c \Big[\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}_c) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_c, \widehat{\boldsymbol{\beta}}) - \mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_c) + \mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}})\Big]$$

$$= \sum_{c=1}^{C} \boldsymbol{K}_c \Big[\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}\Big)^{-1} \boldsymbol{U} + \boldsymbol{\Psi}\Big] - \sum_{c=1}^{C} \boldsymbol{K}_c \Big[\Big(\sum_{c=1}^{C} \boldsymbol{K}_c\Big)^{-1} \boldsymbol{\Psi}\Big]$$

$$= \sum_{c=1}^{C} \boldsymbol{\Psi}\Big[\boldsymbol{\Psi} + (\sum_{g=1}^{G_c} \boldsymbol{K}_{gc})^{-1} \boldsymbol{U}\Big]^{-1} \Big[\Big(\sum_{g=1}^{G} \boldsymbol{K}_{gc}\Big)^{-1} \boldsymbol{U} + \boldsymbol{\Psi}\Big] - \sum_{c=1}^{C} \boldsymbol{K}_c \Big[\Big(\sum_{c=1}^{C} \boldsymbol{K}_c\Big)^{-1} \boldsymbol{\Psi}\Big]$$

$$= (C-1)\boldsymbol{\Psi}.$$

4.3.3 Estimation of Future Mortality Rates

In this Section, we fit the $\log m_{t,xgc}$ rates for year $t = t_1, \ldots, t_n$, age $x = x_1, \ldots, x_{k_{gc}}$, gender $g = 1, \ldots, G_c$ and country $c = 1, \ldots, C$ to the HCR model. Hence, fitted rates up to year t_n can be compactly written as $\widehat{\boldsymbol{Y}}_{xgc} = \boldsymbol{Z}_{xgc}\, \widehat{\boldsymbol{\beta}}_{xgc}$ for $x = x_1, \ldots, x_{k_{gc}},\ g = 1, \ldots, G_c,\ c = 1, \ldots, C$ and mortality rates for one-year ahead are estimated by

$$Y^{Cred}_{t_n+1,xgc} = \log \widehat{m}^{Cred}_{t_n+1,xgc} = \beta^{Cred}_{1,xgc} + \beta^{Cred}_{2,xgc}\ (n+1). \tag{4.66}$$

Then, we can employ two methods to extrapolate mortality rates over a given forecasting horizon $h = 1, 2, \ldots, H$.

4.3.4 Method 1: Initial Fitting Span (IF)

The estimates of future mortality rates for ages $x = x_1, \ldots, x_{k_{gc}}$ are given by extrapolating one-year ahead estimates (4.66) to $\log \widehat{m}^{Cred}_{t_n+h,xgc} = \widehat{\beta}^{Cred}_{1,xgc} + \widehat{\beta}^{Cred}_{2,xgc}\ (n+h)$, for $h = 2, \ldots, H$. Under this method, forecasts are based on the mortality data of the initial fitting span.

4.3.5 Method 2: Moving Fitting Span (MF)

In actuarial practice, we can use the MF method to estimate future mortality rates. One-year ahead estimates $\log \widehat{m}^{Cred}_{t_n+1,xgc}$ are embedded to the existing fitting span and $\log m_{t_1,xgc}$ is simultaneously excluded from it, such that the fitting year span is moved by one year each time to $[t_2, t_n+1]$, $[t_3, t_n+2]$, $[t_4, t_n+3]$ and on. Then, after estimating structural parameters, we can consecutively obtain $\log \widehat{m}^{Cred}_{t_n+2,xgc}, \ldots, \log \widehat{m}^{Cred}_{t_n+H,xgc}$. Similar practical methods have also been applied by Tsai and Zhang (2019). We note that under the MF method forecasts are based on more recent mortality trends, while for both extrapolation methods, one-year ahead estimates $\log \widehat{m}^{Cred}_{t_n+1,xgc}$ are the same.

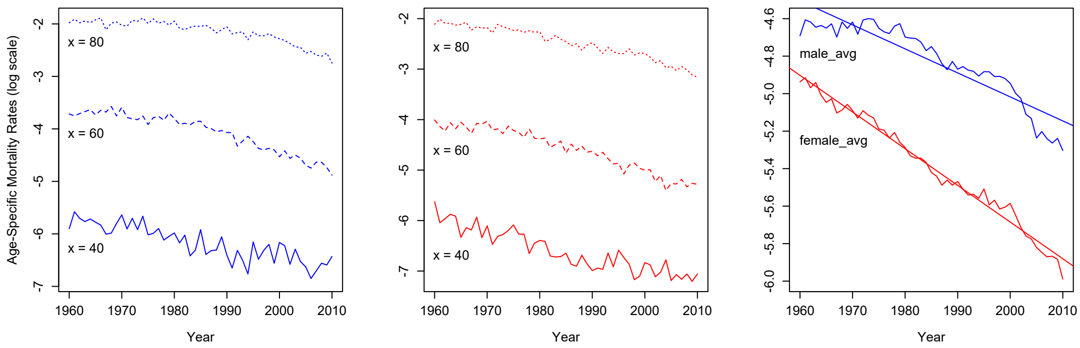
4.4 Empirical Illustration

In this section, we evaluate the LC_{joK}, the LC_{acf} model and the proposed credibility model on mortality data for both genders of Ireland, Finland and Norway. Mortality data are obtained from the Human Mortality Database (HMD, 2017), including calendar years from 1960 to 2010. Mortality evolution over the period 1960-2010 in Ireland, Finland and Norway is illustrated in Figures 4.2a, 4.2b and 4.2c, respectively, where the observed rates show a downward trend for discrete ages $x = 40,\ 60,\ 80$ for males in the left panels and females in the middle panels. In addition, average mortality decline shows a steep downward trend over time in the right panels of Figures 4.2a, 4.2b and 4.2c for both genders of each country.

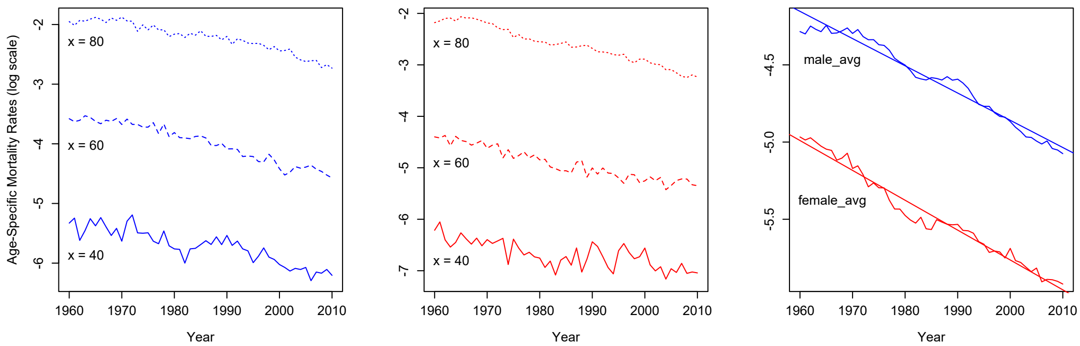
For the numerical illustration that follows, we set $w_{t,xgc} = 1$ and $k_{gc} = k$ ages, implying no weights and the same number of fitted ages for any gender g and country c. Then, the $\log m_{t,xgc}$ rates are fitted on selected periods of totally $n = 41$ ($[t_1, t_n] = [1960, 2000]$) years and $k = 65$ ($[x_1, x_k] = [20, 84]$) ages. We consider two forecasting periods $[t_n + 1, t_n + H]$, $h = 1\cdots, H$, for an $H = 10$ years span (2001-2010) and an $H = 20$ years span (1991-2010). Also, for robustness (relative to changes in the fitting period of data) fitting and forecasting periods are partitioned in various sub-periods, presented in Table 4.2. The HCR mortality methods were built into the R (R Core Team, 2017) statistical software, by creating our own routines.

Table 4.2 Selected fitting and forecasting periods.

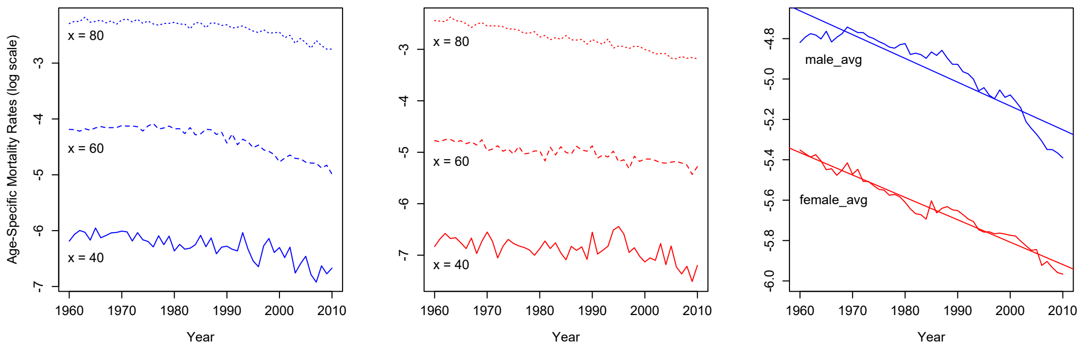
Fitting Length	**Fitting Period**	**Forecasting Horizon**	**Forecasting Period**
n	$[t_1, t_n]$	H	$[t_n+1, t_n+H]$
$n = 41$	$[1960, 2000]$		$[2001, 2010]$
$n = 31$	$[1970, 2000]$	$H = 10$	$[2001, 2010]$
$n = 21$	$[1980, 2000]$		$[2001, 2010]$
$n = 11$	$[1990, 2000]$		$[2001, 2010]$
$n = 31$	$[1960, 1990]$	$H = 20$	$[1991, 2010]$
$n = 21$	$[1970, 1990]$		$[1991, 2010]$



(a) Ireland



(b) Finland



(c) Norway

Figure 4.2 Observed $\log m(t, xgc)$ of period 1960-2010 in Ireland, Finland and Norway at the age of 40, 60 and 80 for males in the left panel and females in the middle panel. Average male and female $\log m(t, x)$ values over the ages 20-84 are illustrated in the right panel, where straight lines show the corresponding trends in mortality decline.

4.4.1 Forecasting Results

The original LC model (Chapter 3, Section 3.2.1) is fitted for the selected periods of Table 4.2, separately on six single populations (the males of Ireland, the females of Ireland, the males of Finland, the females of Finland, the males of Norway and the females of Norway), while the LC_{joK}, the LC_{acf}, the HCR_{IF} and the HCR_{MF} models are jointly applied to six populations.

The forecasting performances of the LC, the LC_{joK}, the LC_{acf} and the HCR models are comparatively evaluated using the mean absolute percentage forecast error (MAPFE) measure. MAPFE has been used in many studies to measure forecast errors. For some of them, see D'Amato et al. (2012), Lin and Tsai (2015) and Tsai and Wu (2018). MAPFE average (avg) values over the selected forecasting horizon H, for age x, gender g and country c are defined as

$$MAPFE_{avg} = \frac{1}{H \times k} \sum_{h=1}^{H} \sum_{x=x_1}^{x_k} \left| \frac{\exp\left[\log \widehat{m}_{t_n+h,xgc}^{Cred}\right] - m(t_n+h,xgc)}{m(t_n+h,xgc)} \right| \times 100 \,.$$

Accuracy results over the 10-year forecasting period $[2001, 2010]$ are presented for both genders of Ireland, Finland and Norway in Tables 4.3 (a), (b), (c) and (d), for fitting periods $[1960, 2000]$, $[1970, 2000]$, $[1980, 2000]$ and $[1990, 2000]$, respectively. The corresponding results over the 20-year forecasting period are given in Tables 4.4 (a) and (b), for fitting periods $[1960, 1990]$ and $[1970, 1990]$, respectively. MAPFE results indicate that HCR models gave us the better forecasts in average, compared with the single and the multi-population Lee-Carter models for both genders of the three countries (six populations in total). Boldface numbers indicate the lowest error values over the corresponding fitting periods for each gender, while average values over the six populations are given in the last row of each fitting subtable. The ranking order (based on the lowest values) is given in brackets and shows that for each one of the selected fitting and forecasting periods, the HCR_{MF} and the HCR_{IF} models outperform in average.

From the Lee-Carter type models, the LC_{acf} gave us the lowest error values for Finland males of fitting period 1960-2000. Also, the single population LC model produced the lowest errors for Norway male data of years 1960-2000. For consistency, the forecasting performance was also evaluated with other well-known measures, such as the mean absolute forecast error (MAFE), the root sum of squared forecast error (RSSFE) and the root mean of squared forecast error (RMSFE), leading us to the same ranking results, with the HCR_{MF} and the HCR_{IF} models being on top.

Table 4.3 Averaged MAPFE values (%) over the 10 year forecasting period [2001, 2010] for the Lee-Carter type and the proposed hierarchical models fitted on years (a) [1960, 2000], (b) [1970, 2000], (c) [1980, 2000] and (d) [1990, 2000], for both genders of Ireland, Finland and Norway.

$\mathbf{MAPFE}^{[2001,2010]}_{[t_1,2000]}$		**Lee-Carter Models**			**Hierarchical Models**	
Country	Gender	LC	LC_{joK}	LC_{acf}	HCR_{IF}	HCR_{MF}
				(a) Fitting period: [1960,2000]		
Ireland	Male	28.99	31.92	29.05	30.45	**27.86**
	Female	27.14	23.95	27.42	21.47	**20.44**
Finland	Male	12.65	16.07	**12.14**	14.52	13.18
	Female	19.29	16.33	18.13	16.04	**15.94**
Norway	Male	**22.85**	26.96	23.62	26.24	23.43
	Female	16.18	15.71	15.80	**14.82**	14.87
Average		21.18(4)	21.82(5)	21.03(3)	20.59(2)	19.29(1)
				(b) Fitting period: [1970,2000]		
Ireland	Male	26.48	27.18	25.75	23.74	**22.40**
	Female	25.38	21.93	25.55	18.87	**18.75**
Finland	Male	12.42	14.29	12.24	**12.08**	12.38
	Female	19.31	17.32	18.60	**16.40**	17.19
Norway	Male	19.92	21.91	19.95	19.78	**18.81**
	Female	16.59	16.47	16.08	**15.28**	15.89
Average		20.02(5)	19.85(4)	19.69(3)	17.69(2)	17.57(1)
				(c) Fitting period: [1980,2000]		
Ireland	Male	24.91	25.44	23.57	**21.72**	22.49
	Female	24.38	22.43	24.39	**19.86**	20.77
Finland	Male	13.17	15.16	13.87	13.12	**11.38**
	Female	21.42	19.83	21.40	18.58	**17.12**
Norway	Male	18.75	17.11	18.54	15.77	**14.13**
	Female	18.15	17.64	16.88	16.45	**15.28**
Average		20.13(5)	19.60(3)	19.78(4)	17.58(2)	16.86(1)
				(d) Fitting period: [1990,2000]		
Ireland	Male	28.39	30.26	26.13	25.14	**23.13**
	Female	27.52	28.06	29.05	**23.06**	24.79
Finland	Male	12.20	12.42	10.78	**9.84**	9.97
	Female	20.32	18.81	21.86	**15.35**	15.47
Norway	Male	17.19	15.80	16.92	**13.86**	16.17
	Female	17.78	17.46	17.32	**14.04**	15.17
Average		20.57(5)	20.47(4)	20.34(3)	16.88(1)	17.45(2)

Table 4.4 Averaged MAPFE values (%) over the 20 year forecasting period [1991, 2010] for the Lee-Carter type and the proposed hierarchical models fitted on years (a) [1960, 1990] and (b) [1970, 1990], for both genders of Ireland, Finland and Norway.

$\mathbf{MAPFE}^{[1991,2010]}_{[t_1,1990]}$		**Lee-Carter Models**			**Hierarchical Models**	
Country	Gender	LC	LC_{joK}	LC_{acf}	HCR_{IF}	HCR_{MF}
(a) Fitting period: [1960,1990]						
Ireland	Male	28.53	34.06	28.16	32.39	**26.88**
	Female	25.14	25.67	23.60	21.44	**18.53**
Finland	Male	23.27	19.42	23.04	16.64	**13.87**
	Female	18.28	16.06	16.73	16.83	**16.68**
Norway	Male	29.42	32.32	30.10	31.46	**27.77**
	Female	16.44	15.18	16.12	**14.04**	14.53
Average		23.51(4)	23.78(5)	22.96(3)	22.13(2)	19.71(1)
(b) Fitting period: [1970,1990]						
Ireland	Male	25.79	28.24	23.13	23.07	**19.61**
	Female	23.24	23.62	23.06	17.82	**16.07**
Finland	Male	19.42	16.07	19.28	**12.25**	13.59
	Female	18.78	16.58	17.42	**16.20**	17.00
Norway	Male	24.17	27.26	**23.43**	24.47	25.07
	Female	17.88	16.43	17.71	**14.83**	16.43
Average		21.55(5)	21.37(4)	20.67(3)	18.11(2)	17.96(1)

From the results of Tables 4.3 and 4.4, we can also observe that deviations in the average MAPFE values between the best performed HCR and Lee-Carter models range between 8% - 17%. This means that if we use the HCR_{MF} instead of the LC_{acf} to model mortality for years 1960-1990, the 20-year average forecasting performance for both genders of the three countries gets improved by 14%, while for years 1960-2000 the 10-year forecasting performance gets improved by 8% and if we use the HCR_{IF} instead of the LC_{acf} for years 1990-2000 the 10-year forecasting performance gets improved by 17%.

4.5 Concluding Remarks

In this chapter, we proposed a hierarchical credibility regression method to model the mortality data of multiple counties, genders and ages, structured in a hierarchical form. We considered different extrapolation strategies to derive future mortality rates, and then, we compared the forecasting performances between the hierarchical model and two Lee-Carter extensions for multiple populations. Based on the accuracy results, the proposed models gave us a better forecasting performance in comparison with the

Lee-Carter models. This indicates that hierarchical credibility regression modelling can be effectively applied to mortality datasets of multiple populations, with possible similarities in their demographic or socio-economic structure.

To summarize, the averaged MAPFE results show that the best performing method in average is the HCR_{MF}, which indicates that mortality forecasts should be generally be based on more recent observations. In addition, the forecasting performance of the hierarchical model was even better when it was applied applied only for Finland and Norway datasets of both genders (4 populations), resulting to even smaller MAPFE values for the selected periods.

Chapter 5

A Crossed Classification Credibility Mortality Model for Multiple Populations

5.1 Introduction

In the previous chapter, we proposed a hierarchical credibility regression method to model the mortality between multiple counties, genders and ages. Under this method, age-specific mortality data for each gender are assumed to be nested into multiple countries. An extension of the hierarchical credibility is the crossed classification credibility model, in which mortality factors can also be modelled symmetrically.

Dannenburg et al. (1996) proposed the two-way crossed classification model, where every contract in an insurance portfolio is assumed to be affected by the same number of risk factors. Goulet (2001) generalized this model, allowing for a variable number of risk factors per contract and Fung and Xu (2008) estimated the structural parameters of the crossed classification credibility model using linear mixed models.

In this chapter, we present a multi-population mortality model, based on crossed classification credibility techniques. Differently from the standard Lee-Carter methodology, where the Lee-Carter time index is assumed to follow an appropriate time series process, under the proposed approach, period dynamics of mortality are modelled under a two-way crossed classification credibility framework. This approach estimates the impact of gender and country characteristics on mortality, allowing for possible interaction effects between them.

This chapter is organized as follows. Section 5.2 describes a different formulation of the original Lee-Carter model and then proposes a crossed classification credibility model for multiple populations, with a detailed description for the parameter estimation and the forecasting procedure. Section 5.3 presents an empirical illustration of the proposed method on males and females from three developed countries, the United Kingdom (UK), the USA and Japan. Section 5.4 concludes this chapter.

5.2 A Credible Extension of the Lee-Carter Method for Multiple Populations

In this section, we first review the original Lee-Carter (1992) model for a single population and then we propose a credible extension for multiple populations. Let us denote $D_{t,xgc}$ as the observed number of deaths and $E_{t,xgc}$ as the average population (exposure to risk) in consecutive calendar years $t = t_1, \ldots, t_n$, ages $x = x_1, \ldots, x_k$, genders $g = 1, \ldots, G$ $(G = 2,$ i.e., male $= 1,$ female $= 2)$ and countries $c = 1, \ldots, C$. Then, mortality rates $m_{t,xgc}$ are obtained by the ratio $D_{t,xgc}/E_{t,xgc}$. The original Lee-Carter (1992) model is given as follows

$$\log m_{t,xgc} = \alpha_{xgc} + \beta_{xgc}\ \kappa_{t,gc} + \varepsilon_{t,xgc}, \tag{5.1}$$

where α_{xgc} is an age parameter that reflects the average mortality at age x for gender g and country c, $\kappa_{t,gc}$ is a period parameter which indicates the general level of mortality in year t for gender g and country c and β_{xgc} is an age parameter that indicates the corresponding deviation from the average mortality, as the general level of mortality changes. Errors $\varepsilon_{t,xgc}$ are assumed to be normally distributed, with zero mean and constant variance, $\varepsilon_{t,xgc} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, reflecting period, age, gender and country effects not captured by the model. Lee and Carter (1992) estimated the model parameters using a close approximation to the singular value decomposition method under the constraints

$$\textstyle\sum_{t=t_1}^{t_n} \kappa_{t,gc} = 0 \text{ and } \sum_{t=t_1}^{t_n} \varepsilon_{t,xgc} = 0, \tag{5.2}$$

which give

$$\widehat{\alpha}_{xgc} = \tfrac{1}{t_n - t_1 + 1} \textstyle\sum_{t=t_1}^{t_n} \log m_{t,xgc} \tag{5.3}$$

and constraints

$$\textstyle\sum_{x=x_1}^{x_k} \beta_{xgc} = 1 \text{ and } \sum_{x=x_1}^{x_k} \varepsilon_{t,xgc} = 0, \tag{5.4}$$

which yield

$$\widehat{\kappa}_{t,gc} = \textstyle\sum_{x=x_1}^{x_k} \left[\log m_{t,xgc} - \widehat{\alpha}_{xgc}\right]. \tag{5.5}$$

Then, by regressing $(\log m_{t,xgc} - \widehat{\alpha}_{xgc})$ on $\widehat{\kappa}_{t,gc}$ for each age x (without an intercept term), we obtain

$$\widehat{\beta}_{xgc} = \frac{\sum_{t=t_1}^{t_n} \left[\log m_{t,xgc} - \widehat{\alpha}_{xgc}\right] \widehat{\kappa}_{t,gc}}{\sum_{t=t_1}^{t_n} \widehat{\kappa}_{t,gc}^2}. \tag{5.6}$$

After estimating the model parameters, Lee and Carter (1992) suggested a random walk with a drift parameter θ_{gc} to model period estimates for each gender g of country c, i.e., $\widehat{\kappa}_{t,gc} = \widehat{\kappa}_{t-1,gc} + \theta_{gc} + \varepsilon_{t,gc}$, where the time trend errors $\varepsilon_{t,gc}$ are assumed to be independent and identically distributed, and independent of the model errors $\varepsilon_{t,xgc}$. The drift parameter is then estimated by $\widehat{\theta}_{gc} = \frac{1}{t_n - t_1} \sum_{t=t_1+1}^{t_n} (\widehat{\kappa}_{t,gc} - \widehat{\kappa}_{t-1,gc}) = \frac{1}{n-1}(\widehat{\kappa}_{t_n,gc} - \widehat{\kappa}_{t_1,gc})$. Hence, $\widehat{\kappa}_{t_n,gc}$ are first projected $h = 1, 2, \ldots$ years ahead using $\widehat{\kappa}_{t_n+h,gc} = \widehat{\kappa}_{t_n,gc} + \widehat{\theta}_{gc} \cdot h$ and then future mortality estimates in year $t_n + h$, for age x, gender g and country c are obtained by

$$\log \widehat{m}_{t_n+h,xgc} = \widehat{\alpha}_{xgc} + \widehat{\beta}_{xgc} \cdot \widehat{\kappa}_{t_n+h,gc} = \log \widehat{m}_{t_n,xgc} + (\widehat{\beta}_{xgc} \cdot \widehat{\theta}_{gc}) \cdot h\,. \tag{5.7}$$

5.2.1 Modelling the Period Improvements of Mortality

Instead of modelling the mortality rate levels according to the original Lee-Carter methodology, many studies have shown that is more advantageous to target on the mortality improvement rates. In order to capture the dependence structure between ages (more accurately), Mitchell et al. (2013) reformulated the Lee-Carter model by targeting to the log mortality improvement. This formulation was later adopted by Schinzinger et al. (2016) and Tsai and Lin (2017b) to model the log improvement rates of the original Lee-Carter model under a classical credibility framework. Following the same approach, we consider the random variable of the log improvement rates

$$M_{t,xgc} = \log \frac{m_{t,xgc}}{m_{t-1,xgc}} = \log m_{t,xgc} - \log m_{t-1,xgc}, \tag{5.8}$$

for $t = t_2, \ldots, t_n,\ x = x_1, \ldots, x_k,\ g = 1, \ldots, G$ and $c = 1, \ldots, C$. Then by substituting (5.1) into (5.8), we can get

$$M_{t,xgc} = \beta_{xgc} \cdot \Delta\kappa_{t,gc} + \Delta\varepsilon_{t,xgc}, \tag{5.9}$$

where $\Delta\kappa_{t,gc} = \kappa_{t,gc} - \kappa_{t-1,gc}$ expresses the general level of mortality improvement between consecutive years, while β_{xgc} reflects the sensitivity to this mortality improvement. The errors $\Delta\varepsilon_{t,xgc} = \varepsilon_{t,xgc} - \varepsilon_{t-1,xgc}$ are assumed to be independent, identically distributed and independent from $\Delta\kappa_{t,gc}$. Given each age x, gender g and country c, we assume that $\Delta\varepsilon_{t,xgc}$ is a white noise process, $\Delta\varepsilon_{t,xgc} \sim \mathcal{N}(0, 2 \cdot \sigma_\varepsilon^2)$ such that the mortality improvement rates $M_{t,xgc}$ are independent for $t = t_2, \ldots, t_n$. To ensure identifiability of model (5.9), again constraints in (5.4) are considered. Also, summing $M_{t,xgc}$ over x using (5.4) gives

$$M_{t,\cdot gc} = \sum_{x=x_1}^{x_k} M_{t,xgc} = \sum_{x=x_1}^{x_k} \beta_{xgc}\, \Delta\kappa_{t,gc} + \sum_{x=x_1}^{x_k} \Delta\varepsilon_{t,xgc} = \Delta\kappa_{t,gc}, \tag{5.10}$$

which connects the aggregate (for all ages) mortality improvement rates for consecutive years with the general level of mortality improvement between these years, without involving the age structure x, for gender g and country c, implied by β_{xgc}. This advantage has also been highlighted by Schinzinger et al. (2016). Expression (5.10) can be equivalently obtained by taking the first differences of $\widehat{\kappa}_{t,gc}$ with respect to t in (5.5).

Under formulation (5.9), one-year ahead mortality estimates for age x, gender g and country c can be obtained by

$$\log \widehat{m}_{t_n+1,xgc} = \log \widehat{m}_{t_n,xgc} + \widehat{\beta}_{xgc} \cdot \widehat{\Delta\kappa}_{t_n+1,gc}, \tag{5.11}$$

or recursively, for $h = 2, \ldots H$ years ahead by

$$log \widehat{m}_{t_n+h,xgc} = \log \widehat{m}_{t_n,xgc} + \widehat{\beta}_{xgc} \sum_{s=1}^{h} \widehat{\Delta\kappa}_{t_n+s,gc}, \tag{5.12}$$

where $\widehat{\beta}_{xgc}$ is estimated by (5.6) and $\widehat{\Delta\kappa}_{t_n+h,gc}$ is projected using a time series model. Selecting an appropriate time series model for the $\widehat{\Delta\kappa}_{t,gc}$ process is a crucial stage in mortality projection. Slightly different time series models can lead to evident deviations in future mortality estimates. This sensitivity, as well as possible limitations in data availability, can make forecasting even harder. However, the random walk with a drift model is widely adopted, serving as a compromise for the majority of mortality modelling studies.

5.2.2 The Crossed Classification Credibility Framework

Differently from the standard Lee-Carter methodology, in which $\Delta\kappa_{t,gc}$s are assumed to follow an appropriate time series process, herein we propose a two-way crossed classification credibility approach to model the crossed effects between genders and countries. Under this two-way model, mortality improvement is affected by two qualitative risk factors and the possible interactions between them. Let us consider two categories of risk factors, associated with gender and country characteristics, denoted by the random variables $\mathcal{K}_g^{(1)}$, $g = 1, \ldots, G$ and $\mathcal{K}_c^{(2)}$, $c = 1, \ldots, C$, respectively. Under this specification, the gender and country dynamics of mortality improvement can be decomposed (additively) into independent variance components by assuming that the elements of the gender-related vector $(\mathcal{K}_1^{(1)}, \ldots, \mathcal{K}_G^{(1)})$ and the country-related vector $(\mathcal{K}_1^{(2)}, \ldots, \mathcal{K}_C^{(2)})$ are independent and identically distributed, with all of their elements being mutually independent across vectors. The conditional means and variance are then defined as

$$\mathrm{E}(\Delta\kappa_{t,gc} | \mathcal{K}_g^{(1)}, \mathcal{K}_c^{(2)}) \;=\; \mu_{12}(\mathcal{K}_g^{(1)}, \mathcal{K}_c^{(2)}),$$

$$\begin{aligned}
\mathrm{E}(\Delta\kappa_{t,gc}|\mathcal{K}_g^{(1)}) &= \mu_1(\mathcal{K}_g^{(1)}), \\
\mathrm{E}(\Delta\kappa_{t,gc}|\mathcal{K}_c^{(2)}) &= \mu_2(\mathcal{K}_c^{(2)}), \\
\mathrm{E}[\mathrm{Var}(\Delta\kappa_{t,gc}|\mathcal{K}_g,\mathcal{K}_c)] &= s^2.
\end{aligned}$$

Under the above assumptions, $\Delta\kappa_{t,gc}$ in (5.10) can be formulated into a two-way crossed classification model

$$\Delta\kappa_{t,gc} = \mu + k_g^{(1)} + k_c^{(2)} + k_{gc}^{(12)} + \varepsilon_{t,gc}, \tag{5.13}$$

for $t = t_2, \ldots, t_n$ ($n-1$ years), gender $g = 1$ (males), 2 (females) and country $c = 1, \ldots, C$ assuming that

$$\mathrm{E}(k_g^{(1)}) = \mathrm{E}(k_c^{(2)}) = \mathrm{E}(k_{gc}^{(12)}) = \mathrm{E}(\varepsilon_{t,gc}) = 0 \tag{5.14}$$

and

$$\mathrm{Var}(k_g^{(1)}) = \sigma_g^2,\ \mathrm{Var}(k_c^{(2)}) = \sigma_c^2,\ \mathrm{Var}(k_{gc}^{(12)}) = \sigma_{gc}^2,\ \mathrm{Var}(\varepsilon_{t,gc}) = s^2. \tag{5.15}$$

Table 5.1 Gender, country and interaction terms in the two-way crossed classification model.

Gender \ Country	$k_1^{(2)}$	$\cdots$	$k_C^{(2)}$
$k_1^{(1)}$	$k_{11}^{(12)}$	$\cdots$	$k_{1C}^{(12)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k_G^{(1)}$	$k_{G1}^{(12)}$	$\cdots$	$k_{GC}^{(12)}$

The gender and country factors, along with their interaction terms of model (5.13) are given in Table 5.1. The one-year ahead credibility estimator of the general level of mortality improvement for gender g and country c in the two-way crossed classification model is given by the following theorem. For a detailed proof of this theorem, we refer to Dannenburg et al. (1996, pp. 109-113).

Theorem 5.1. *The credibility estimator of $\Delta\kappa_{t_n+1,gc}$ under the two-way crossed classification model is given by*

$$\widehat{\Delta\kappa}_{t_n+1,gc}^{Cred} = Z^{(12)}\overline{\Delta\kappa}_{\cdot,gc} + (1-Z^{(12)})(\widehat{\mu} + \widehat{k}_g^{(1)} + \widehat{k}_c^{(2)}). \tag{5.16}$$

Sketch of Proof: Note that the credibility estimator of $\Delta\kappa_{t_n+1,gc}$ is directly obtained by using the credibility estimators for each one of the terms appeared in model (5.13) as follows

$$\widehat{\Delta\kappa}^{\text{Cred}}_{t_n+1,gc} = \widehat{\mu} + \widehat{k}^{(1)}_g + \widehat{k}^{(2)}_c + \widehat{k}^{(12)}_{gc}, \tag{5.17}$$

where the credibility estimators of gender and country factors $k^{(1)}_g$, $k^{(2)}_c$ and $k^{(12)}_{gc}$ are given by

$$\widehat{k}^{(12)}_{gc} = Z^{(12)}_{gc}(\overline{\Delta\kappa}_{.,gc} - \widehat{\mu} - \widehat{k}^{(1)}_g - \widehat{k}^{(2)}_c), \tag{5.18}$$

$$\widehat{k}^{(1)}_g = Z^{(1)}(\overline{\Delta\kappa}_{.,g.} - \widehat{\mu}) - Z^{(1)}\frac{1}{C}\sum_{c=1}^{C}\widehat{k}^{(2)}_c, \tag{5.19}$$

$$\widehat{k}^{(2)}_c = Z^{(2)}(\overline{\Delta\kappa}_{.,.c} - \widehat{\mu}) - Z^{(2)}\frac{1}{G}\sum_{g=1}^{G}\widehat{k}^{(1)}_g. \tag{5.20}$$

Thus, expression (5.16) is derived by substituting (5.18) into (5.17). □

In terms of credibility theory, future estimates for the general level of mortality improvement given in (5.16) can be interpreted as the weighted average of the characteristics within a specific gender-country selection $\overline{\Delta\kappa}_{.,gc}$ (individual) and the overall mean $\widehat{\mu}$ (collective), plus two terms that correspond to gender $\widehat{k}^{(1)}_g$ and country $\widehat{k}^{(2)}_c$ factors, estimated by (5.19) and (5.20), respectively. The estimators for the terms, appeared in (5.16), (5.19), (5.20) and (5.18), are presented in the next section.

5.2.3 Estimation of Model Parameters

Following Dannenburg et al. (1996), the credibility factors are given by

$$\begin{aligned} Z^{(12)} &= \frac{\sigma^2_{gc}}{\sigma^2_{gc} + s^2/(n-1)}, \\ Z^{(1)} &= \frac{C\cdot\sigma^2_g}{C\cdot\sigma^2_g + \sigma^2_{gc} + s^2/(n-1)}, \\ Z^{(2)} &= \frac{G\cdot\sigma^2_c}{G\cdot\sigma^2_c + \sigma^2_{gc} + s^2/(n-1)} \end{aligned} \tag{5.21}$$

and the means are defined as

$$\overline{\Delta\kappa}_{.,gc} = \frac{1}{n-1}\sum_{t=t_2}^{t_n}\Delta\kappa_{t,gc}, \quad \overline{\Delta\kappa}_{.,g.} = \frac{1}{C}\sum_{c=1}^{C}\overline{\Delta\kappa}_{.,gc}, \quad \overline{\Delta\kappa}_{.,.c} = \frac{1}{G}\sum_{g=1}^{G}\overline{\Delta\kappa}_{.,gc}. \tag{5.22}$$

The structural parameters are estimated by

$$\widehat{\mu} = \overline{\Delta\kappa}_{.,..} = \frac{1}{C \cdot G\,(n-1)} \sum_{c=1}^{C} \sum_{g=1}^{G} \sum_{t=t_2}^{t_n} \widehat{\Delta\kappa}_{t,gc} = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \overline{\Delta\kappa}_{.,gc}, \tag{5.23}$$

$$\widehat{s}^2 = \frac{1}{C \cdot G\,(n-2)} \sum_{c=1}^{C} \sum_{g=1}^{G} \sum_{t=t_2}^{t_n} (\widehat{\Delta\kappa}_{t,gc} - \overline{\Delta\kappa}_{.,gc})^2, \tag{5.24}$$

while the estimators of σ_g^2, σ_c^2 and σ_{gc}^2 can be obtained as solutions of the following linear system of equations

$$\begin{aligned}
\sum_{g=1}^{G} \frac{1}{G}\Big[\sum_{c=1}^{C} \frac{1}{C}(\overline{\Delta\kappa}_{.,gc} - \overline{\Delta\kappa}_{.,g.})^2 - \widehat{s}^2 \frac{C-1}{C(n-1)}\Big] &= (\sigma_c^2 + \sigma_{gc}^2)(1 - \sum_{g=1}^{G}\sum_{c=1}^{C} \frac{1}{G}(\frac{1}{C})^2), \\
\sum_{c=1}^{C} \frac{1}{C}\Big[\sum_{g=1}^{G} \frac{1}{G}(\overline{\Delta\kappa}_{.,gc} - \overline{\Delta\kappa}_{.,.c})^2 - \widehat{s}^2 \frac{G-1}{G(n-1)}\Big] &= (\sigma_g^2 + \sigma_{gc}^2)(1 - \sum_{c=1}^{C}\sum_{g=1}^{G} \frac{1}{C}(\frac{1}{G})^2), \\
\sum_{g=1}^{G}\sum_{c=1}^{C} \frac{1}{G \cdot C}(\overline{\Delta\kappa}_{.,gc} - \overline{\Delta\kappa}_{.,..})^2 - \widehat{s}^2 \frac{G \cdot C - 1}{G \cdot C(n-1)} &= \sigma_g^2(1 - \sum_{g=1}^{G}(\frac{1}{G})^2) + \\
&+ \sigma_c^2(1 - \sum_{c=1}^{C}(\frac{1}{C})^2) + \sigma_{gc}^2(1 - \sum_{g=1}^{G}\sum_{c=1}^{C}(\frac{1}{G \cdot C})^2).
\end{aligned}$$

Remark 5.2. Theoretically, the above estimation can yield non positive estimates for any of $\widehat{\sigma}_g^2$, $\widehat{\sigma}_c^2$, $\widehat{\sigma}_{gc}^2$. So if $\widehat{\sigma}_{gc}^2 \le 0$, then $Z_{gc}^{(12)}$ is set to be zero valued and from (5.16), $\widehat{\Delta\kappa}_{t_n+1,gc}^{\mathrm{Cred}}$ equals to the overall mean $\widehat{\mu}$. Moreover, if $\widehat{\sigma}_{gc}^2$ is positive, but one of $\widehat{\sigma}_g^2$ or $\widehat{\sigma}_c^2$ is non positive, then $Z^{(1)}$ or $Z^{(2)}$ are zero valued and $\widehat{\Delta\kappa}_{t_n+1,gc}^{\mathrm{Cred}}$ yields $Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})(\widehat{\mu} + \widehat{k}_c^{(2)})$ or $Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})(\widehat{\mu} + \widehat{k}_g^{(1)})$, respectively. In particular, if $\widehat{\sigma}_{gc}^2$ is positive, but both $\widehat{\sigma}_g^2$ and $\widehat{\sigma}_c^2$ are non positive, then estimator (5.16) reduces to the Bühlmann credibility formula, i.e., $\widehat{\Delta\kappa}_{t_n+1,gc}^{\mathrm{Cred}} = Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})\widehat{\mu}$.

The results of Remark 5.2 are summarized as follows

$$\widehat{\Delta\kappa}_{t_n+1,gc}^{\mathrm{Cred}} = \begin{cases}
\widehat{\mu}, & \text{if } \widehat{\sigma}_{gc}^2 \le 0, \\
Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})(\widehat{\mu} + \widehat{k}_c^{(2)}), & \text{if } \widehat{\sigma}_{gc}^2 > 0,\ \widehat{\sigma}_c^2 > 0 \text{ and } \widehat{\sigma}_g^2 \le 0, \\
Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})(\widehat{\mu} + \widehat{k}_g^{(1)}), & \text{if } \widehat{\sigma}_{gc}^2 > 0,\ \widehat{\sigma}_g^2 > 0 \text{ and } \widehat{\sigma}_c^2 \le 0, \\
Z^{(12)}\overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)})\widehat{\mu}, & \text{if } \widehat{\sigma}_{gc}^2 > 0,\ \widehat{\sigma}_g^2 \le 0 \text{ and } \widehat{\sigma}_c^2 \le 0.
\end{cases}$$

5.2.4 Mortality Forecasting

Substituting (5.16) into (5.11) gives the one-year ahead credibility forecast of the mortality improvement rates, for age x, gender g and country c as follows

$$\widehat{M}^{\text{Cred}}_{t_n+1,xgc} = \widehat{\beta}_{xgc} \cdot \widehat{\Delta\kappa}^{\text{Cred}}_{t_n+1,gc} = \widehat{\beta}_{xgc} \left[Z^{(12)} \cdot \overline{\Delta\kappa}_{.,gc} + (1 - Z^{(12)}) \left(\widehat{\mu} + \widehat{k}^{(1)}_g + \widehat{k}^{(2)}_c \right) \right] \quad (5.25)$$

Yet recall that β_{xgc} is estimated by (5.6). Lemma 5.3 shows that the credibility estimator of the one-year ahead mortality improvement rates $\widehat{M}^{\text{Cred}}_{t_n+1,xgc}$ preserves the crossed classification credibility form, similarly with estimator (5.16) for the general level of mortality improvement.

Lemma 5.3. *The credibility estimator of the one-year ahead mortality improvement rates* $\widehat{M}^{Cred}_{t_n+1,.gc}$ *can be obtained by*

$$\widehat{M}^{Cred}_{t_n+1,xgc} = Z^{(12)} \cdot \overline{M}_{.,xgc} + (1 - Z^{(12)})(\widehat{\beta}_{xgc} \cdot \overline{M}_{.,...} + \widehat{\beta}_{xgc} \cdot \widehat{\mathcal{M}}^{(1)}_g + \widehat{\beta}_{xgc} \cdot \widehat{\mathcal{M}}^{(2)}_c), \quad (5.26)$$

Proof: For the proof of Lemma (5.3), note that for any arbitrary x, g and c, the credibility factor $Z^{(12)}$ is a scalar, thus expression (5.25) can be rearranged as

$$\widehat{M}^{\text{Cred}}_{t_n+1,xgc} = Z^{(12)} \left(\widehat{\beta}_{xgc} \cdot \overline{\Delta\kappa}_{.,gc} \right) + \left(1 - Z^{(12)} \right) \left(\widehat{\beta}_{xgc} \cdot \widehat{\mu} + \widehat{\beta}_{x,gc} \cdot \widehat{k}^{(1)}_g + \widehat{\beta}_{xgc} \cdot \widehat{k}^{(2)}_c \right). \quad (5.27)$$

Now, averaging (5.9) with respect to t yields

$$\overline{M}_{.,xgc} = \frac{1}{n-1} \sum_{t=t_2}^{t_n} M_{t,xgc} = \frac{1}{n-1} \sum_{t=t_2}^{t_n} \beta_{xgc} \cdot \Delta\kappa_{t,gc} = \beta_{xgc} \cdot \overline{\Delta\kappa}_{.,gc} \,, \quad (5.28)$$

while summing (5.28) over x gives $\overline{M}_{.,.gc} = \overline{\Delta\kappa}_{.,gc}$, which is connected with (5.28) by $\overline{M}_{.,xgc} = \widehat{\beta}_{xgc} \cdot \overline{M}_{.,.gc}$. Further averaging $\overline{M}_{.,.gc}$ with respect to g and c yields

$$\overline{M}_{.,.g.} = \frac{1}{G} \sum_{g=1}^{G} \overline{M}_{.,.gc} = \frac{1}{G} \sum_{g=1}^{G} \overline{\Delta\kappa}_{.,gc} = \overline{\Delta\kappa}_{.,.c} \,, \quad (5.29)$$

$$\overline{M}_{.,..c} = \frac{1}{C} \sum_{c=1}^{C} \overline{M}_{.,.gc} = \frac{1}{C} \sum_{c=1}^{C} \overline{\Delta\kappa}_{.,gc} = \overline{\Delta\kappa}_{.,g.} \quad (5.30)$$

and

$$\overline{M}_{.,...} = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \overline{M}_{.,.gc} = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \overline{\Delta\kappa}_{.,gc} = \overline{\Delta\kappa}_{.,..} = \widehat{\mu} \,. \quad (5.31)$$

Similarly, $\widehat{k}_g^{(1)}$ in (5.19) and $\widehat{k}_c^{(2)}$ in (5.20) may be rewritten as

$$\widehat{\mathcal{M}}_g^{(1)} = Z^{(1)}(\overline{M}_{.,.,c} - \overline{M}_{.,...}) - Z^{(1)}\frac{1}{C}\sum_{c=1}^{C}\widehat{\mathcal{M}}_c^{(2)} \tag{5.32}$$

and

$$\widehat{\mathcal{M}}_c^{(2)} = Z^{(2)}(\overline{M}_{.,.g.} - \overline{M}_{.,...}) - Z^{(2)}\frac{1}{G}\sum_{g=1}^{G}\widehat{\mathcal{M}}_g^{(1)}, \tag{5.33}$$

respectively. Then, substituting (5.28)–(5.33) into (5.27) yields (5.26). □

In the two-way crossed classification credibility estimator (5.26), the individual mean is represented by $\overline{M}_{.,xgc}$ $(= \widehat{\beta}_{xgc} \cdot \overline{M}_{.,.gc})$, while the collective mean by $\widehat{\beta}_{xgc} \cdot \overline{M}_{.,...}$, plus two additional factors, $\widehat{\beta}_{xgc} \cdot \widehat{\mathcal{M}}_g^{(1)}$ and $\widehat{\beta}_{xgc} \cdot \widehat{\mathcal{M}}_c^{(2)}$, which express the influence of gender and country effects.

Remark 5.4. Summing (5.26) over x gives

$$\widehat{M}^{\text{Cred}}_{t_n+1,.gc} = Z^{(12)} \cdot \overline{M}_{.,.gc} + (1 - Z^{(12)})\ (\overline{M}_{.,...} + \widehat{\mathcal{M}}_g^{(1)} + \widehat{\mathcal{M}}_c^{(2)}). \tag{5.34}$$

Estimator (5.34) allows to study the overall improvement of future mortality $\widehat{M}^{\text{Cred}}_{t_n+1,.gc}$ from the aggregate mortality dynamics $\overline{\Delta\kappa}_{.,gc}$, without involving the $\widehat{\beta}_{xgc}$ effects of age x, for gender g and country c.

To summarize, the credibility estimator of one-year ahead mortality rates $\log \widehat{m}^{\text{Cred}}_{t_n+1,xgc}$ can be directly obtained by substituting (5.25) in (5.8)

$$\log \widehat{m}^{\text{Cred}}_{t_n+1,xgc} = \log m_{t_n,xgc} + \widehat{M}^{\text{Cred}}_{t_n+1,xgc}\,, \tag{5.35}$$

or recursively, for h years ahead by

$$\log \widehat{m}^{\text{Cred}}_{t_n+h,xgc} = \log m_{t_n,xgc} + \sum_{s=1}^{h}\widehat{M}^{\text{Cred}}_{t_n+s,xgc}\,. \tag{5.36}$$

In order to obtain future mortality rates $\log \widehat{m}^{\text{Cred}}_{t_n+h,xgc}$, for $h = 2,3,\ldots,H$ years ahead, we first have to estimate mortality improvement rates $\widehat{M}^{\text{Cred}}_{t_n+h,xgc}$ in (5.36). Tsai and Lin (2017b) proposed two methods to determine future mortality rates under the Bülmann credibility approach. Similarly, we employ the moving fitting (MF) and the expanding fitting (EF) methods to estimate future mortality improvement rates under the crossed classification mortality framework as follows.

Method 1: Moving Fitting (MF)

Under this method, the one-year ahead credibility estimate of mortality improvement rates $\widehat{M}^{\text{Cred}}_{t_n+1,xgc}$, obtained by (5.25), is embedded to the existing fitting span, while the first observed rates $M_{t_2,xgc}$ are simultaneously excluded from it. Thus, the fitting span is moved by one year to $[t_3, t_n+1]$, keeping a constant fitting length, to form $[M_{t_3,xgc}, \widehat{M}^{\text{Cred}}_{t_n+1,xgc}]$, which by (5.9) are decomposed to $[\widehat{\Delta\kappa}_{t_3,gc}, \widehat{\Delta\kappa}_{t_n+1,gc}]$. After following the parameter estimation procedure of Section 5.2.3, $\widehat{\Delta\kappa}^{\text{Cred}}_{t_n+2,gc}$ is derived from (5.16) and then is incorporated into (5.25) to obtain $\widehat{M}^{\text{Cred}}_{t_n+2,xgc}$. By repeating the same steps, we can consecutively obtain $\widehat{M}^{\text{Cred}}_{t_n+3,xgc}, \ldots, \widehat{M}^{\text{Cred}}_{t_n+H,xgc}$, which by (5.36) yield $\widehat{m}^{\text{Cred}}_{t_n+3,xgc}, \ldots, \widehat{m}^{\text{Cred}}_{t_n+H,xgc}$, respectively. Under the MF method, a credibility expression for $\widehat{M}^{\text{Cred}}_{t_n+h,xgc}$ for $h = 2, 3, \ldots, H$, age x, gender g and country c equals

$$\widehat{\beta}_{xgc} \cdot Z^{(12)}(t_n+h) \cdot \overline{\Delta\kappa}_{\cdot,gc}(t_n+h) + \widehat{\beta}_{xgc} \left[1 - Z^{(12)}(t_n+h)\right] \left[\overline{\Delta\kappa}_{\cdot,\cdot\cdot}(t_n+h) + \widehat{k}^{(1)}_g + \widehat{k}^{(2)}_c\right], \quad (5.37)$$

where

$$\overline{\Delta\kappa}_{\cdot,gc}(t_n+h) = \frac{1}{n-1} \sum_{t=t_1+h}^{t_n+h-1} \widehat{\Delta\kappa}_{t,gc}, \quad (5.38)$$

$$\overline{\Delta\kappa}_{\cdot,\cdot\cdot}(t_n+h) = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \overline{\Delta\kappa}_{\cdot,gc}(t_n+h) \quad (5.39)$$

and

$$Z^{(12)}_{gc}(t_n+h) = \frac{\sigma^2_{gc}}{\sigma^2_{gc} + s^2/(n-1)}, \quad (5.40)$$

with $\overline{\Delta\kappa}_{\cdot,gc}(t_n+h)$, $\overline{\Delta\kappa}_{\cdot,\cdot\cdot}(t_n+h)$ and $Z^{(12)}(t_n+h)$ being the $\overline{\Delta\kappa}_{\cdot,gc}$, $\overline{\Delta\kappa}_{\cdot,\cdot\cdot}$ and $Z^{(12)}$ for year t_n+h, respectively. Note that $\widehat{\beta}_{xgc}$, as well as $\{s^2, \sigma^2_{gc}\}$ in $Z^{(12)}(t_n+h)$ and $\{\overline{\Delta\kappa}_{\cdot,g\cdot}, \overline{\Delta\kappa}_{\cdot,\cdot c}, \widehat{\mu}\}$ in $\widehat{k}^{(1)}_g$ and $\widehat{k}^{(2)}_c$ are estimated over $[t_1+h, t_n+h-1]$.

Method 2: Expanding Fitting (EF)

Here the one-year ahead credibility estimate of mortality improvement rates $\widehat{M}^{\text{Cred}}_{t_n+1,xgc}$ is embedded to the existing fitting span, but $M_{t_2,xgc}$ is not removed from it. Thus, the fitting span is expanded by one year to $[t_2, t_n+1]$ to form $[M_{t_2,xgc}, \widehat{M}^{\text{Cred}}_{t_n+1,xgc}]$, which by (5.9) are decomposed to $[\widehat{\Delta\kappa}_{t_2,gc}, \widehat{\Delta\kappa}_{t_n+1,gc}]$. By repeating the estimation procedure of Section 5.2.3, we derive $\widehat{\Delta\kappa}^{\text{Cred}}_{t_n+3,gc}$ and $\widehat{M}^{\text{Cred}}_{t_n+3,xgc}$, using (5.16) and (5.25), respectively. We follow the same steps to obtain $\widehat{M}^{\text{Cred}}_{t_n+2,xgc}, \ldots, \widehat{M}^{\text{Cred}}_{t_n+H,xgc}$, which by (5.36) yields $\widehat{m}^{\text{Cred}}_{t_n+2,xgc}, \ldots, \widehat{m}^{\text{Cred}}_{t_n+H,xgc}$. Again, the credibility formula of $\widehat{M}^{\text{Cred}}_{t_n+h,xgc}$, for $h = 2, 3, \ldots, H$ is given by (5.37), but now

$$\overline{\Delta\kappa}_{\cdot,gc}(t_n+h) = \frac{1}{n+h-2} \sum_{t=t_2}^{t_n+h-1} \widehat{\Delta\kappa}_{t,gc}, \quad (5.41)$$

$$\overline{\Delta\kappa}_{\cdot,\cdot\cdot}(t_n+h) = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \overline{\Delta\kappa}_{\cdot,gc}(t_n+h) \quad (5.42)$$

and

$$Z^{(12)}(t_n+h) = \frac{\sigma_{gc}^2}{\sigma_{gc}^2 + s^2/(n+h-2)}. \quad (5.43)$$

Under the EF method, the rest of model components should be estimated over $[t_2, t_n + h - 1]$. MF and EF formulas follow directly from the estimation procedure of Section 5.2.3 to generalize (5.25) for an h-years ahead forecasting horizon. We note that under the MF method, future estimates are based on the mortality experience, gained from the rates of the most recent calendar years, while under the EF method, the whole historical mortality experience is taken into account.

5.3 Empirical illustration

In this Section, we evaluate the forecasting performance of the proposed model, comparatively with the the joint-k (Carter and Lee, 1992) and the augmented common factor (Li and Lee, 2005) multi-populations models, presented in Chapter 4 (Section 4.2). For the numerical illustration, we use data for both genders of three developed countries, the UK, the USA and Japan. The crossed classification credibility mortality model was built into the R statistical software (R Core Team, 2017), by creating our own routines.

5.3.1 Numerical Results

In this Section, we fit the LC, the LC_{joK}, the LC_{acf} models and the proposed credibility model to mortality data for both genders of UK, USA and Japan. Mortality data were obtained from the Human Mortality Database (HMD, 2017), covering calendar years from 1960 to 2015. Figure 5.1 indicates the downward trend of $\log m_{t,xgc}$ rates over the period 1960–2015 for ages $x = 40,\ 60,\ 80$ of both genders in UK, USA and Japan, respectively. This downward trend is eliminated by considering the mortality improvement rates $M_{t,xgc} = \log m_{t,xgc} - \log m_{t-1,xgc}$, illustrated in Figure 5.2.

To proceed, we consider two risk factors, where gender G is the first factor and country C is the second factor. The first factor consists of two categories, with $g = 1$ for males and $g = 2$ for females, while the second factor includes three categories, $c = 1$ for UK, $c = 2$ for USA and $c = 3$ for Japan.

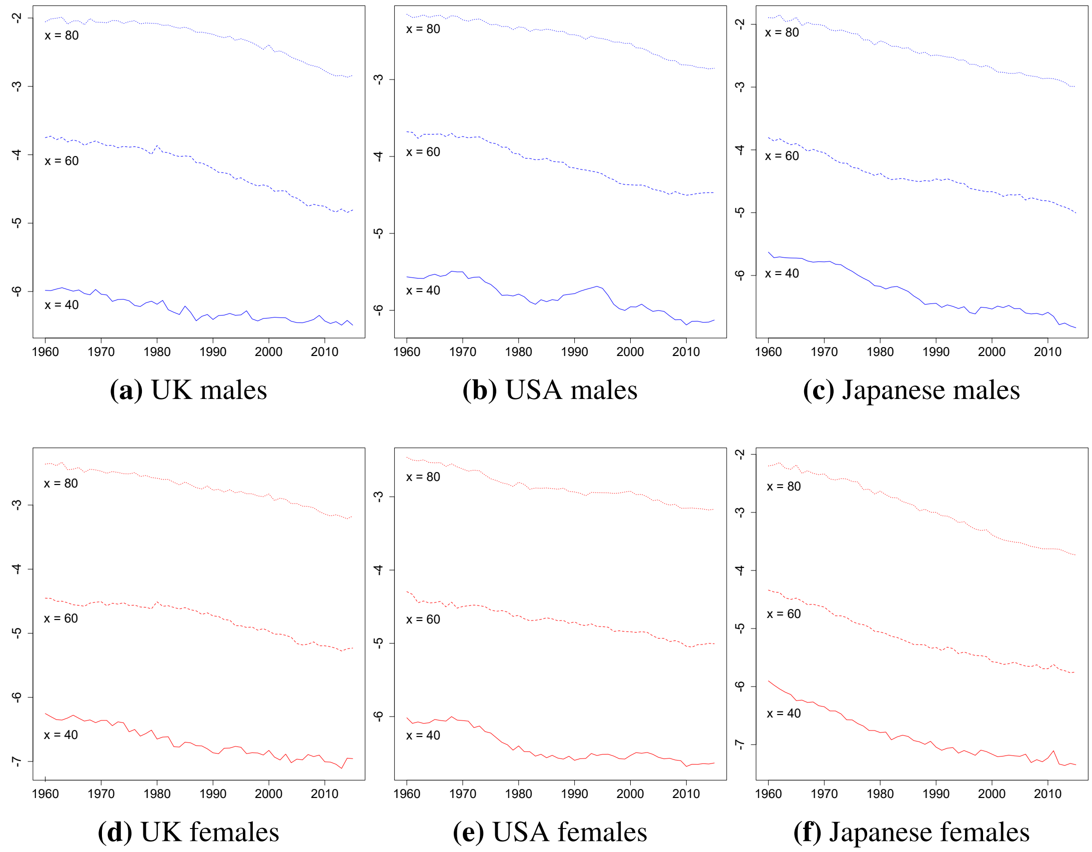
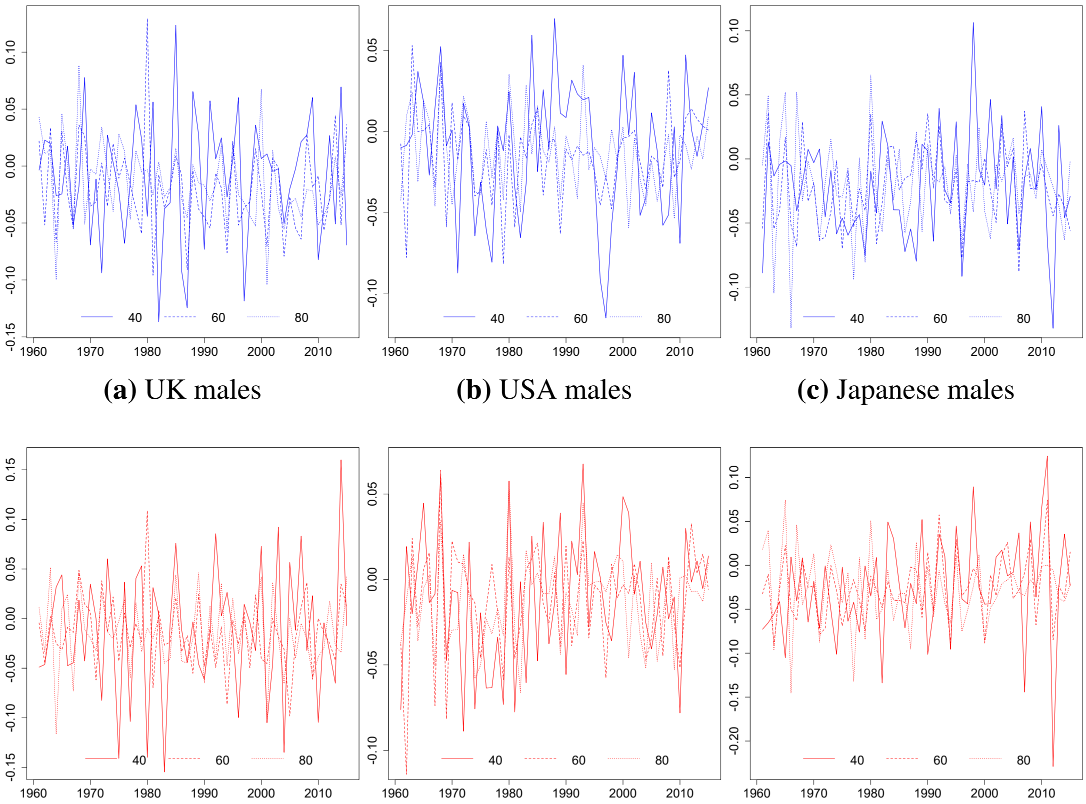


Figure 5.1 $\log m_{t,xgc}$ rates of period 1960–2015 for: (a) UK males, (b) USA males, (c) Japanese males, (d) UK females, (e) USA females and (f) Japanese females at the age of 40 (solid lines), 60 (dashed lines) and 80 (dotted lines).

Model (5.13) for the general level of mortality improvement $\Delta\kappa_{t,gc}$ can be represented by Table 5.2. For the numerical illustration that follows, we set same number of fitted years n and ages k for any selected gender g and country c. Thus, we consider the mortality improvement rates $M_{t,xgc}$ of $n = 56$ $([t_1, t_n] = [1960, 2015])$ years and $k = 65([x_1, x_k] = [20, 84])$ ages, for males and females $(G = 2)$ of selected countries $(C = 3)$. Model (5.9) decomposes mortality improvement into β_{xgc} and $\Delta\kappa_{t,gc}$ parameters, given by (5.6) and (5.10), respectively. The one year ahead estimates $\Delta\kappa_{t_n+1,gc}$ for the general level of mortality improvement are estimated by (5.16) and the corresponding mortality improvement rates can then be derived by (5.25).



(a) UK males **(b)** USA males **(c)** Japanese males

(d) UK females **(e)** USA females **(f)** Japanese females

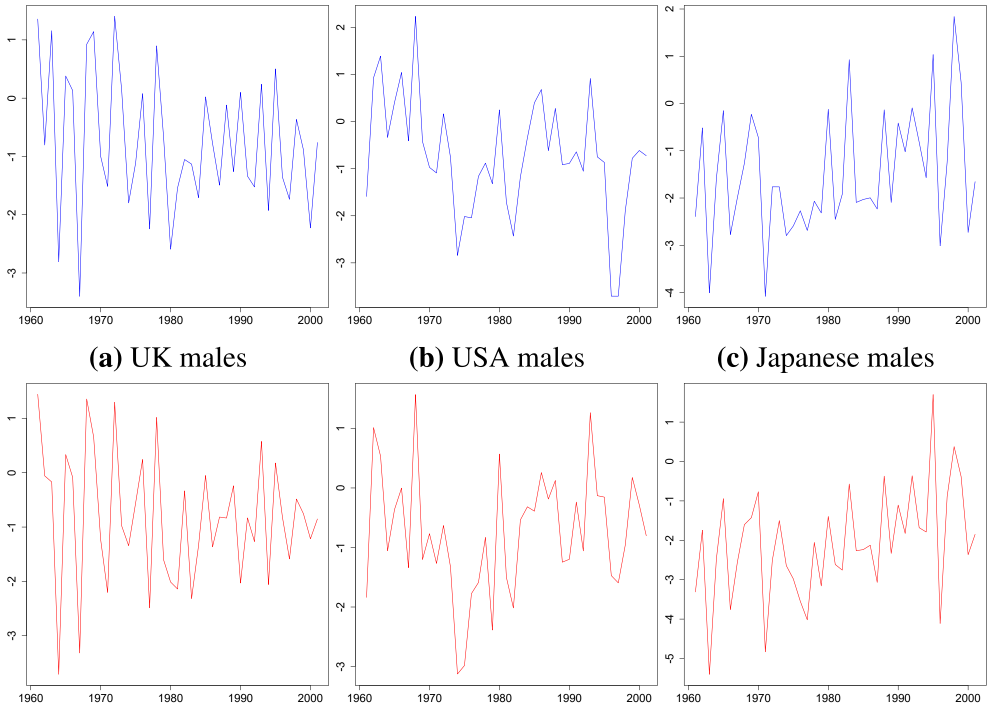
Figure 5.2 $M_{t,xgc}$ improvement rates of period 1961–2015 for: (a) UK males, (b) USA males, (c) Japanese males, (d) UK females, (e) USA females and (f) Japanese females at the age of 40 (solid lines), 60 (dashed lines) and 80 (dotted lines).

Table 5.2 Two-way tabular formulation of $\Delta\kappa_{t,gc}$, for $t = t_2, \ldots, t_n$.

Gender \ Country	UK $(c = 1)$	USA $(c = 2)$	Japan $(c = 3)$
Males $(g = 1)$	$\Delta\kappa_{t_2,11}$ $\vdots$ $\Delta\kappa_{t_n,11}$	$\Delta\kappa_{t_2,12}$ $\vdots$ $\Delta\kappa_{t_n,12}$	$\Delta\kappa_{t_2,13}$ $\vdots$ $\Delta\kappa_{t_n,13}$
Females $(g = 2)$	$\Delta\kappa_{t_2,21}$ $\vdots$ $\Delta\kappa_{t_n,21}$	$\Delta\kappa_{t_2,22}$ $\vdots$ $\Delta\kappa_{t_n,22}$	$\Delta\kappa_{t_2,23}$ $\vdots$ $\Delta\kappa_{t_n,23}$

5.3.2 Interpretation of Credibility Estimates

To better understand how the proposed credibility approach contributes in multi-population mortality modelling, let us consider the observed mortality rates for years 1960 – 2000, ages [20,84] for both genders of UK, USA and Japan. Our aim is to estimate mortality rates for the next year. Figure 5.3 depicts the estimated values for the general level of mortality improvement over the whole period, including the credibility estimates for year 2001, while Table 5.3 presents the estimated parameters of model (5.16).



(a) UK males **(b)** USA males **(c)** Japanese males

(d) UK females **(e)** USA females **(f)** Japanese females

Figure 5.3 $\widehat{\Delta\kappa}_{t,gc}$ estimates for the general level of 1961–2001 mortality improvement for: (a) UK males, (b) USA males, (c) Japanese males, (d) UK females, (e) USA females and (f) Japanese females.

For example, the one year ahead estimates for the general level of mortality improvement of year 2001 are directly obtained by substituting the values of Table 5.3 into (5.16). Thus, the estimate for the UK males can be written as

$$-0.760690 = 0.202901 \cdot (-0.747607) + (1 - 0.202901) \cdot (-1.106842 + 0.046297 + 0.296524).$$

We can easily observe that the credibility factors within cells $Z^{(12)} = 20\%$ and within genders $Z^{(1)} = 40\%$ are both much smaller than $Z^{(2)} = 93\%$. This means that more credibility should be given to the mortality experience within countries. Also, the existence of $\widehat{k}_g^{(1)}$ and $\widehat{k}_c^{(2)}$ is significant for the derivation of future estimates. A credibility factor $Z^{(12)}$ equal to 20% implies that the above formula assigns more weight to the overall mean -1.1068, which is corrected by the sum of the estimated gender and country factors $(0.046297 + 0.296524)$. Under this correction, the one year ahead estimate (-0.760689) is quite close to the estimated mean value (-0.747607), finally differing only by 1.7%.

Table 5.3 Estimated parameters for the one year ahead credibility estimates for year 2001.

(a) Credibility factors

$Z^{(12)}$	$Z^{(1)}$	$Z^{(2)}$
0.202901	0.397808	0.929935

(b) Gender factors

$\widehat{k}_g^{(1)}$	$g = 1$ (Males)	0.046297
	$g = 2$ (Females)	-0.046308

(c) Country factors

$\widehat{k}_c^{(2)}$	$c = 1$ (UK)	$c = 2$ (USA)	$c = 3$ (Japan)
	0.296524	0.337690	-0.634199

(d) Estimated means

$\overline{\Delta\kappa}_{.,gc}$	$c = 1$ (UK)	$c = 2$ (USA)	$c = 3$ (Japan)
$g = 1$ (Males)	-0.747607	-0.731227	-1.492506
$g = 2$ (Females)	-0.828357	-0.756201	-2.085154

(e) Credibility estimates

$\widehat{\Delta\kappa}^{\text{Cred}}_{2001,gc}$	$c = 1$ (UK)	$c = 2$ (USA)	$c = 3$ (Japan)
$g = 1$ (Males)	-0.760690	-0.724553	-1.653710
$g = 2$ (Females)	-0.850890	-0.803436	-1.847775

(f) Structural parameters

$\widehat{\mu}$	$\widehat{s}^2$	σ_g^2	σ_c^2	σ_{gc}^2
-1.106842	1.560731	0.0107789	0.324847	0.009932

5.3.3 Evaluation of Forecasts

After estimating the $\widehat{\Delta\kappa}_{t,gc}$, the long-term mortality estimates can be derived by (5.36), choosing between one of the MF or EF extrapolation methods. Thus, the LC_{joK}, the LC_{acf}, the MF and the EF models were fitted to mortality data of six populations, including both genders ($G = 2$) of UK, USA and Japan ($C = 3$), while the single population LC model was fitted separately on males and females for each country. To ensure robustness of forecasts, all the methods were applied for various fitting and forecasting periods, presented in Table 5.4. The forecasting performance of each model was evaluated using the mean absolute percentage forecast error (MAPFE) measure.

The averaged (avg) MAPFE values over the selected forecasting horizon H, for age x, gender g and country c are defined by

$$MAPFE_{avg} = \frac{1}{H \times k} \sum_{h=1}^{H} \sum_{x=x_1}^{x_k} \left| \frac{\exp\left[\log \widehat{m}(t_n+h,xgc)\right] - m(t_n+h,xgc)}{m(t_n+h,xgc)} \right| \times 100 \, .$$

Table 5.4 Fitting and forecasting periods.

Fitting Length	**Fitting Period**	**Forecasting Horizon**	**Forecasting Period**
n	$[t_1, t_n]$	H	$[t_n+1, t_n+H]$
$n = 51$	$[1960, 2010]$	$H = 5$	$[2011, 2015]$
$n = 41$	$[1970, 2010]$		$[2011, 2015]$
$n = 46$	$[1960, 2005]$	$H = 10$	$[2006, 2015]$
$n = 36$	$[1970, 2005]$		$[2006, 2015]$
$n = 41$	$[1960, 2000]$	$H = 15$	$[2001, 2015]$
$n = 31$	$[1970, 2000]$		$[2001, 2015]$

The evaluation results are presented in Table 5.5, for six different fitting and forecasting periods of both genders in UK, USA and Japan. MAPFE values indicate that the proposed method gave us the better forecasts in average (ranking order in brackets), compared with the single and the multi-population Lee-Carter models for both genders of the three countries (six populations in total). Regarding the Lee-Carter type models, the LC_{acf} gave us the lowest error values for the majority of the selected periods. Surprisingly, the single population Lee-Carter model produced lower average errors in comparison with its multi-population counterparts for the fitting period of years 1970-2000. For consistency, forecasting performance was also evaluated with other well-known measures such as the mean absolute forecast error (MAFE), the root sum of squared forecast error (RSSFE) and the root mean of squared forecast error (RMSFE), leading us to the same ranking results, with the MF and EF methods being again on top.

Table 5.5 Averaged MAPFE values (%) for the Lee-Carter type and the proposed credibility models fitted on years (a) $[1960, 2010]$, (b) $[1960, 2005]$, (c) $[1960, 2000]$, (d) $[1970, 2010]$, (e) $[1970, 2005]$ and (f) $[1970, 2000]$ for both genders of UK, USA and Japan, aged $20 - 84$.

(a) Fitting period: $[1960, 2010]$, forecasting period: $[2011, 2015]$						
MAPFE		**Lee-Carter Models (LC)**			**Credibility Models (CM)**	
Country	**Gender**	**LC**	**LC_{joK}**	**LC_{acf}**	**CM_{MF}**	**CM_{EF}**
UK	Male	11.29	17.47	10.59	6.49	6.51
	Female	7.19	10.12	7.47	5.83	5.84
USA	Male	6.62	7.22	6.39	3.65	3.64
	Female	7.99	8.13	7.78	3.86	3.85
Japan	Male	7.72	7.66	7.87	5.28	5.25
	Female	8.87	15.81	6.60	5.41	5.39
Average		8.28(4)	11.07(5)	7.78(3)	5.09(2)	5.08(1)
(b) Fitting period: $[1960, 2005]$, forecasting period: $[2006, 2015]$						
Country	**Gender**	**LC**	**LC_{joK}**	**LC_{acf}**	**MF**	**EF**
UK	Male	12.95	19.91	13.38	8.31	8.56
	Female	9.80	12.80	9.75	7.54	7.72
USA	Male	8.42	10.06	8.26	5.89	5.96
	Female	8.51	8.74	8.74	5.34	5.28
Japan	Male	8.43	8.19	8.90	5.57	5.58
	Female	12.14	15.91	10.22	5.65	5.78
Average		10.04(4)	12.60(5)	9.87(3)	6.38(1)	6.48(2)
(c) Fitting period: $[1960, 2000]$, forecasting period: $[2001, 2015]$						
Country	**Gender**	**LC**	**LC_{joK}**	**LC_{acf}**	**CM_{MF}**	**CM_{EF}**
UK	Male	17.27	21.71	18.24	14.19	14.71
	Female	12.39	13.82	12.57	9.05	9.45
USA	Male	9.50	12.13	8.59	8.57	8.57
	Female	8.55	9.55	8.89	8.62	8.29
Japan	Male	9.03	10.48	9.50	6.22	6.34
	Female	16.07	18.11	14.69	7.71	8.54
Average		12.14(4)	14.30(5)	12.08(3)	9.06(1)	9.32(2)
(d) Fitting period: $[1970, 2010]$, forecasting period: $[2011, 2015]$						
Country	**Gender**	**LC**	**LC_{joK}**	**LC_{acf}**	**MF**	**EF**
UK	Male	11.29	13.68	10.32	6.43	6.42
	Female	6.30	7.47	6.37	5.75	5.74
USA	Male	7.05	6.55	6.90	4.05	4.02
	Female	8.23	7.87	7.72	4.18	4.18
Japan	Male	5.93	5.93	5.99	5.59	5.47
	Female	5.81	12.48	5.81	5.66	5.61
Average		7.43(4)	9.00(5)	7.19(3)	5.27(2)	5.24(1)

Table 5.5 *Cont.*

(e) Fitting period: [1970, 2005], forecasting period: [2006, 2015]

Country	Gender	LC	LC_{joK}	LC_{acf}	MF	EF
UK	Male	12.60	15.76	12.06	7.63	7.79
	Female	8.90	10.16	8.65	7.08	7.16
USA	Male	8.49	8.99	8.55	5.65	5.58
	Female	9.31	9.29	9.35	5.38	5.46
Japan	Male	6.91	6.70	7.12	5.87	5.54
	Female	8.39	12.77	7.94	5.81	5.77
Average		9.10(4)	10.61(5)	8.95(3)	6.24(2)	6.22(1)

(f) Fitting period: [1970, 2000], forecasting period: [2001, 2015]

Country	Gender	LC	LC_{joK}	LC_{acf}	MF	EF
UK	Male	16.28	18.04	16.33	13.57	13.73
	Female	11.22	11.67	11.40	8.46	8.68
USA	Male	10.06	10.88	9.69	8.92	8.68
	Female	9.84	10.38	10.12	9.11	9.24
Japan	Male	8.45	9.36	8.94	5.70	6.31
	Female	12.54	15.03	12.02	6.56	7.58
Average		11.40(3)	12.56(5)	11.42(4)	8.72(1)	9.04(2)

5.4 Concluding Remarks

In this chapter, we presented a credibility formulation of the Lee-Carter method particularly designed for multi-population mortality modelling. Differently from the standard Lee-Carter methodology, where the time index is assumed to follow an appropriate time series process, under our approach, the period dynamics of mortality are modelled under a crossed classification credibility formulation. This approach allows to model the gender and country mortality effects, as well as the possible interactions that may exist between multiple genders and countries.

The future mortality rates were derived by incorporating different extrapolation methods and the forecasting performance between the proposed method and two Lee-Carter extensions for multiple populations was thoroughly evaluated. Numerical results on mortality data of different periods, for both genders of the UK, the USA and Japan show that the MF and the EF credibility methods have the best forecasting performance, based on MAPFE values, for all the fitting and forecasting periods.

Chapter 6

General Conclusions

This thesis combined aspects of life and non-life insurance to propose novel mortality modelling methods. Based on the actuarial credibility modelling techniques, we developed innovative methods, aiming to model and forecast mortality for a single or multiple populations. Before introducing these methods, in Chapter 2, we examined the fitting and forecasting performance of the most used stochastic mortality models on Greek data. Furthermore, an application of mortality modelling in pricing insurance-related products was also included.

However, it is highly important to point out that modelling efficiency and forecasting reliability may also depend on unexpected events or other factors. For this reason, in Chapter 3, we proposed a credibility regression mortality framework with random coefficients to model mortality data. The applicability of this modelling approach was comparatively illustrated on Greek mortality data with a pricing application on insurance-related products.

In Chapter 4, we extended the credibility regression mortality framework to a multi-level hierarchical structure, which models mortality for multiple populations of different ages, genders and countries in a hierarchical form. The efficiency of this method was illustrated on multi-population mortality data for both genders of three northern European countries, the Ireland, the Finland and the Norway.

In Chapter 5, we presented a mortality model for multiple populations, based on crossed classification credibility techniques. More specifically, period dynamics of mortality are modelled under a two-way crossed classification credibility framework. Under this specification, modelling procedure takes into account the impact of gender and country characteristics on mortality improvement and allows for possible interaction effects that may exist between them. The proposed method was thoroughly illustrated on mortality data for both genders of three well-developed countries, the UK, the USA and Japan.

To sum up, the proposed methods aimed to derive future mortality rates per age, gender and country, which are essential for the construction of a life table (period or cohort). Then, any quantity of demographic or actuarial interest (e.g. life expectancies, annuities) can be derived from this table.

Regarding some possible demographic extensions of our work, we note that even if this thesis deals with applications on all-cause mortality data, the use of our methods is not restricted to these data. For example, the hierarchical model of Chapter 4 can accommodate datasets for different causes of death, per age, gender or country. In addition, Lee (1993) suggested the use of the Lee-Carter method to model to forecast the time series of US fertility. Accordingly, the crossed classification model of Chapter 5 can be implemented on fertility modelling studies.

Finally, we have to mention that all of the proposed methods throughout this thesis were implemented using the "open source" R statistical software (R Core Team, 2017). Specifically, we developed our own R routines to build the credibility regression mortality model of Chapter 3, the hierarchical credibility regression mortality model of Chapter 4 and the crossed classification credibility mortality model of Chapter 5. The numerical illustrations of each model are fully applicable and provide encouragement that credibility modelling approaches could contribute to demographic projections, beyond the scopes of this thesis.

Bibliography

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Aro, H. and Pennanen, T. (2011). A user-friendly approach to stochastic mortality modelling. *European Actuarial Journal*, 1(S2):151–167.

Booth, H., Hyndman, R. J., Tickle, L. and de Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research*, 15(101):289–310.

Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336.

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(2008):3–43.

Bozikas, A. and Pitselis, G. (2018). An Empirical Study on Stochastic Mortality Modelling under the Age-Period-Cohort Framework: The Case of Greece with Applications to Insurance Pricing. *Risks*, 6(2), 44.

Bozikas, A. and Pitselis, G. (2019). Credible Regression Approaches to Forecast Mortality for Populations with Limited Data. *Risks*, 7(1), 27.

Brouhns, N., Denuit, M. and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

Bühlmann, H. (1967). Experience Rating and Credibility. *ASTIN Bulletin*, 4(3):199–207.

Bühlmann, H. and Gisler, A. (2005). *A course in credibility theory and its applications*. The Netherlands: Springer.

Butt, Z., Haberman, S. and Shang, H. L. (2014). The ilc Package in R: Generalised Lee-Carter Models using Iterative Fitting Algorithms, R package version 1.0. Available at http://cran.r-project.org/package=ilc.

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

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D. and Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A. and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.

Carter, L. R. and Lee, R. D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, 8(3):393–411.

Currie, I. D. (2006). Smoothing and forecasting mortality rates with P-splines. Talk given at the Institute of Actuaries, June 2006. Available at http://www.macs.hw.ac.uk/˜iain/research/talks.html.

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

D'Amato, V., Haberman, S., Piscopo, G. and Russolillo, M. (2012). Modelling dependent data for longevity projections. *Insurance: Mathematics and Economics*, 51(3):694–701.

D'Amato, V., Haberman, S., Piscopo, G., Russolillo, M. and Trapani, L. (2014). Detecting Common Longevity Trends by a Multiple Population Approach. *North American Actuarial Journal*, 18(1):139–149.

Dannenburg, D. R., Kaas, R. and Goovaerts, M. J. (1996). *Practical Actuarial Credibility Models*. Institute of Actuarial Science and Econometrics of the University of Amsterdam.

De Jong, P. and Tickle, L. (2006). Extending Lee–Carter Mortality Forecasting. *Mathematical Population Studies*, 13(1):1–18.

De Vylder, F. (1976). Geometrical credibility. *Scandinavian Actuarial Journal*, 1976(3):121–149.

De Vylder, F. (1978). Parameter estimation in credibility theory. *ASTIN Bulletin*, 10(1):99–112.

Debón, a., Martínez-Ruiz, F. and Montes, F. (2010). A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics and Economics*, 47(3):327–336.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366):427–431.

Fung, W. K. and Xu, X. (2008). Estimation of Structural Parameters in Crossed Classification Credibility Model Using Linear Mixed Models. *In: Brito P. (eds) COMPSTAT 2008. Physica-Verlag HD*, (2003):241–251.

Gaille, S. (2012). Forecasting mortality: when academia meets practice. *European Actuarial Journal*, 2(1):49–76.

Gong, Y., Li, Z., Milazzo, M., Moore, K. and Provencher, M. (2018). Credibility Methods for Individual Life Insurance. *Risks*, 6(4):144.

Goovaerts, M. J., Kaas, R., Van Heerwaarden, A. E. and Bauwelinckx, T. (1990). *Effective actuarial methods*. North-Holland.

Goulet, V. (2001). A generalized crossed classification credibility model. *Insurance: Mathematics and Economics*, 28(2):205–216.

Greene, W. H. (2012). *Econometric analysis 7th edition.* International ed. London: Pearson Education Limited.

Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55.

Hachemeister, C. (1975). Credibility for Regression Models with Application to Trend (Reprint). In Kahn, P., editor, *CREDIBILITY: Theory and Applications*, 307–348. Academic Press, Inc.

Hansen, H. (2013). The forecasting performance of mortality models. *AStA Advances in Statistical Analysis*, 97(1):11–31.

Hardy, M. R. and Panjer, H. H. (1998). A credibility approach to mortality risk. *Astin Bulletin*, 28(2):269–283.

Hatzopoulos, P. and Haberman, S. (2009). A parameterized approach to modeling and forecasting mortality. *Insurance: Mathematics and Economics*, 44(1):103–123.

Hatzopoulos, P. and Haberman, S. (2011). A dynamic parameterization modeling for the age-period-cohort mortality. *Insurance: Mathematics and Economics*, 49:155–174.

Hatzopoulos, P. and Haberman, S. (2013). Common Mortality Modeling and Coherent Forecasts: An Empirical Analysis of Worldwide Mortality Data. *Insurance: Mathematics and Economics*, 52(2):320–337.

Hatzopoulos, P. and Haberman, S. (2015). Modeling trends in cohort survival probabilities. *Insurance: Mathematics and Economics*, 64:162–179.

Hildreth, C. and Houck, J. P. (1968). Some Estimators for a Linear Model with Random Coefficients. *Journal of the American Statistical Association*, 63(322):584–595.

Hobcraft, J., Menken, J. and Preston, S. (1982). Age, period and cohort effects in demography: a review. *Population Index*, 48(1):4–43.

Hsiao, C. (1986). *Analysis of Panel Data. In Econometric Society Monographs. New York: Cambridge University Press.* Econometric Society Monographs. Cambridge University Press.

Huang, F. and Browne, B. (2017). Mortality forecasting using a modified Continuous Mortality Investigation Mortality Projections Model for China I: methodology and country-level results. *Annals of Actuarial Science*, 11(1):20–45.

Human Mortality Database (2017). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at https://www.mortality.org/.

Hunt, A. and Blake, D. (2015). On the Structure and Classification of Mortality Models. Pension Institute. Working Paper. Available at http://www.pensions-institute.org/workingpapers/wp1506.pdf.

Hunt, A. and Villegas, A. M. (2015). Robustness and convergence in the Lee-Carter model with cohort effects. *Insurance: Mathematics and Economics*, 64:186–202.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Hyndman, R. J., Booth, H., Tickle, L. and Maindonald, J. (2014). demography: Forecasting mortality and fertility data, R package version 1.18. Available at http://robjhyndman.com/software/demography/.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51(10):4942–4956.

Jewell, W. S. (1975). The use of collateral data in credibility theory: a hierarchical model. *IIASA Research Memorandum*, 1975(6):1–16.

Kalogirou, S., Tsimbos, C., Verropoulou, G. and Kotsifakis, G. (2012). Regional mortality differentials in Greece by selected causes of death: 2006-2008. *Journal of Maps*, 8(4):354–360.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147–152.

Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). *Loss models: from data to decisions. New York, John Wiley & Sons.*, volume 715. John Wiley & Sons.

Kostaki, A. and Zafeiris, K. N. (2019). Dealing with limitations of empirical mortality data in small populations. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1–6.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Ledolter, J., Klugman, S. and Lee, C.-S. (1991). Credibility Models with Time-varying Trend Components. *ASTIN Bulletin*, 21(1):73–91.

Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range and ultimate level. *International Journal of Forecasting*, 9:187–202.

Li, H. and Lu, Y. (2018). A Bayesian non-parametric model for small population mortality. *Scandinavian Actuarial Journal*, 2018(7):605–628.

Li, J. S. H. and Hardy, M. R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal*, 15(2):177–200.

Li, J. S. H., Zhou, R. and Hardy, M. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics*, 63:121–134.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography*, 42(3):575–594.

Li, N., Lee, R. and Tuljapurkar, S. (2004). Using the Lee–Carter Method to Forecast Mortality for Populations with Limited Data. *International Statistical Review*, 72(1):19–36.

Lin, T. and Tsai, C. C.-L. (2015). A Simple Linear Regression Approach to Modeling and Forecasting Mortality Rates. *Journal of Forecasting*, 34(5):543–559.

Lovász, E. (2011). Analysis of Finnish and Swedish mortality data with stochastic mortality models. *European Actuarial Journal*, 1(2):259–289.

Luan, X. (2015). *A Pseudo Non-Parametric Bühlmann Credibility Approach to Modeling Mortality Rates*. MSc thesis, Simon Fraser.

Maccheroni, C. and Nocito, S. (2017). Backtesting the Lee–Carter and the Cairns–Blake–Dowd Stochastic Mortality Models on Italian Death Rates. *Risks*, 5(3):34.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R. and Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics*, 52(2):275–285.

Norberg, R. (1980). Empirical bayes credibility. *Scandinavian Actuarial Journal*, 1980(4):177–194.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Pitacco, E., Denuit, M., Haberman, S. and Olivieri, A. (2009). *Modelling longevity dynamic for pensions and annuity business*. Oxford University Press.

Pitselis, G. (2004). Credibility models with cross-section effect and with both cross-section and time effects. *Blätter der DGVFM*, 26(4):643–663.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.

R Core Team (2017). R: A language and environment for Statistical Computing. Available at https://www.r-project.org/.

Rao, C. R. (1973). *Linear statistical inference and its applications. Second Edition.* Wiley New York.

Renshaw, A. E. and Haberman, S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.

Renshaw, A. E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics and Economics*, 42(2):797–816.

Salhi, Y. and Thérond, P. E. (2018). Age-specific adjustment of graduated mortality. *ASTIN Bulletin*, 48(2):543–569.

Salhi, Y., Thérond, P. E. and Tomas, J. (2016). A credibility approach of the Makeham mortality law. *European Actuarial Journal*, 6(1):61–96.

Schinzinger, E., Denuit, M. M. and Christiansen, M. C. (2016). A multivariate evolutionary credibility model for mortality improvement rates. *Insurance: Mathematics and Economics*, 69:70–81.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shang, H. L., Booth, H. and Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, 25:173–214.

Stoeldraijer, L., Van Duin, C., Van Wissen, L. and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research*, 29(13):323–354.

Sundt, B. (1979). A Hierarchical Credibility Regression Model. *Scandinavian Actuarial Journal*, 1979:107–114.

Sundt, B. (1980). A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, 1980(1):25–32.

Tsai, C. C.-L. and Lin, T. (2017a). A Bühlmann credibility approach to modeling mortality rates. *North American Actuarial Journal*, 21(2):204–227.

Tsai, C. C.-L. and Lin, T. (2017b). Incorporating the Bühlmann credibility into mortality models to improve forecasting performances. *Scandinavian Actuarial Journal*, 2017(5):419–440.

Tsai, C. C.-L. and Wu, A. D. (2018). A Hierarchical Credibility Approach to Modelling Mortality Rates for Multiple Populations. *Research Paper, Canadian Institute of Actuaries*.

Tsai, C. C.-L. and Yang, S. (2015). A Linear Regression Approach to Modeling Mortality Rates of Different Forms. *North American Actuarial Journal*, 19(1):1–23.

Tsai, C. C.-L. and Zhang, Y. (2019). A multi-dimensional Bühlmann credibility approach to modeling multi-population mortality rates. *Scandinavian Actuarial Journal*, 1–26.

Tsimbos, C., Kotsifakis, G., Verropoulou, G. and Kalogirou, S. (2011). Life expectancy in Greece 1991-2007: regional variations and spatial clustering. *Journal of Maps*, 7(1):280–290.

Turner, H. and Firth, D. (2015). Generalized nonlinear models in R: An overview of the gnm package. Available at http://cran.r-project.org/package=gnm.

Van Berkum, F., Antonio, K. and Vellekoop, M. (2016). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*, 2016(7):581–603.

Verropoulou, G. and Tsimbos, C. (2016). Mortality by Cause of Death Among Immigrants and Natives in a South European Country: The Case of Greece, 2011. *Journal of Immigrant and Minority Health*, 18(2):337–344.

Villegas, A. M., Millossovich, P. and Kaishev, V. K. (2017). StMoMo: An R Package for Stochastic Mortality Modelling.

Wan, C. and Bertschi, L. (2015). Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. *Insurance: Mathematics and Economics*, 63:66–75.

Wang, H. C., Yue, C. S. J. and Chong, C. T. (2018). Mortality models and longevity risk for small populations. *Insurance: Mathematics and Economics*, 78:351–359.

Wilson, C. (2001). On the Scale of Global Demographic Convergence 1950-2000. *Population and Development Review*, 27:155–171.

Xie, Y. (2013). animation : An R Package for Creating Animations and Demonstrating Statistical Methods. *Journal of Statistical Software*, 53(1):1–27.

Zafeiris, K. N. and Kostaki, A. (2017). Recent mortality trends in Greece. *Communications in Statistics - Theory and Methods*, 1–15.

Zhao, B. B. (2012). A modified Lee-Carter model for analysing short-base-period data. *Population Studies*, 66(1):39–52.