A Hierarchical Credibility Approach to Modelling Mortality Rates for Multiple Populations

by

Adelaide Di Wu

B.Sc., Simon Fraser University, 2016

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

Approval

Name: **Adelaide Di Wu**

Degree: **Master of Science (Actuarial Science)**

Title: ***A Hierarchical Credibility Approach to Modelling Mortality Rates for Multiple Populations***

Examining Committee: **Chair:** Dr. Yi Lu
Associate Professor

Dr. Cary Chi-Liang Tsai
Senior Supervisor
Associate Professor

Ms. Barbara Sanders
Supervisor
Associate Professor

Dr. LiangLiang Wang
Internal Examiner
Assistant Professor

Date Defended: May 8^{th}, 2018

Abstract

A hierarchical credibility model is a generalization of the Bühlmann credibility model and the Bühlmann-Straub credibility model with a tree structure of four or more levels. This project aims to incorporate the hierarchical credibility theory, which is used in property and casualty insurance, to model the dependency of multi-population mortality rates. The forecasting performances of the three/four/five-level hierarchical credibility models are compared with those of the classical Lee-Carter model and its three extensions for multiple populations (joint-k, cointegrated and augmented common factor Lee-Carter models). Numerical illustrations based on mortality data for both genders of the US, the UK and Japan with a series of fitting year spans and three forecasting periods show that the hierarchical credibility approach contributes to more accurate forecasts measured by the AMAPE (average of mean absolute percentage errors). The proposed model is convenient to implement and can be further applied to projecting a mortality index for pricing mortality-indexed securities.

Keywords: Hierarchical Credibility Theory; Bühlmann Credibility Theory; Lee-Carter Model; Multi-population Mortality Model

Dedication

I would like to dedicate this project to my beloved parents for their selfless love, unconditional support, and ongoing encouragement.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my senior supervisor Dr. Cary Tsai who encouraged me to pursue a master's degree of Actuarial Science at Simon Fraser University and provided comprehensive guidance and constant support throughout my master's study. I am fortunate to receive professional guidance from a knowledgeable supervisor who not only has expertise in academic literature and rich industrial experience, but also are unlimitedly patient and kind to me. I benefited enormously from his innovative ideas and intuitive analysis. Without his enlightening and numerous support, I would never have been able to finish this project.

In addition, I am very grateful to my committee members, Ms. Barbara Sanders and Dr. Liangliang Wang, for their thorough reviews and insightful comments on this project. Furthermore, I would like to thank Dr. Yi Lu for taking the precious time out of her busy schedule to serve as the chair of the examining committee, and many thanks are further extended to the whole department, all the faculty members and staffs who provided all kinds of help and support to me.

Last but not least, I would like express my sincere appreciation to my parents and friends for providing me with unfailing support and continuous encouragement throughout the years of my study and through the process of writing this degree project. This accomplishment would not have been possible without them.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

As medical services, health care and living conditions have improved dramatically, the life expectancy of human have increased over the recent decades, which greatly affects life insurance, annuities, pension plans, and social security systems. Policyholders pay premiums to insurance companies in exchange for financial protections. With accurate forecasts of mortality rates, the companies can sell their products with competitive premiums, set up adequate reserves, and maintain financial solvency. However, inaccurate mortality forecasts may lead insurance companies to financial insolvency due to underpricing or loss of market shares because of overpricing. Therefore, constructing an effective and accurate mortality model is essential to pricing and reserving of life insurance, annuities, and mortality-linked securities.

In a global world, medicine, public health and living environment of developed countries progress at a similar pace. Therefore, the mortality rates for these well-developed countries are inclined to be correlated. Especially, within a country, the mortality experience of the female population is highly correlated to that of the male population since they are exposed to the same medical and environmental situations. A mortality index is a weighted average of the realized mortality rates over an age span and both genders of some selected countries. Recently, mortality-indexed/-linked securities and derivatives, such as longevity bonds, longevity index, and longevity swaps, are arising. Life insurance companies and pension providers are interested in those financial securities and derivatives to hedge mortality and longevity risks. Consequently, building a multi-population mortality model that takes into account the dependency of multi-population mortality rates and provides a high degree of accuracy in projecting mortality rates and index is important in these days.

Credibility theory is widely applied in property and casualty insurance, where the credibility estimate for the next year is a weighted average of the sample mean of the past claim data of a policyholder and the true mean of claims. The claim data can be severities or frequencies of claims, and the claim data for all policyholders in a group can be treated as a tree structure of three levels. Taking group auto insurance as an example, level one is the individual claim data in the past years for each policyholder, and level two represents the group risk to which the policyholder belongs, which is shown within the dashed box of Figure 1.1. The sample mean is calculated based on the past experiences of a given policyholder, and the true mean can be estimated from the past experiences of all policyholders in the whole group. Tsai and Lin (2017a, b) incorporated the Bühlmann credibility model to forecast mortality rates for a single population, which provides more accurate forecasts than the Lee-Carter model.

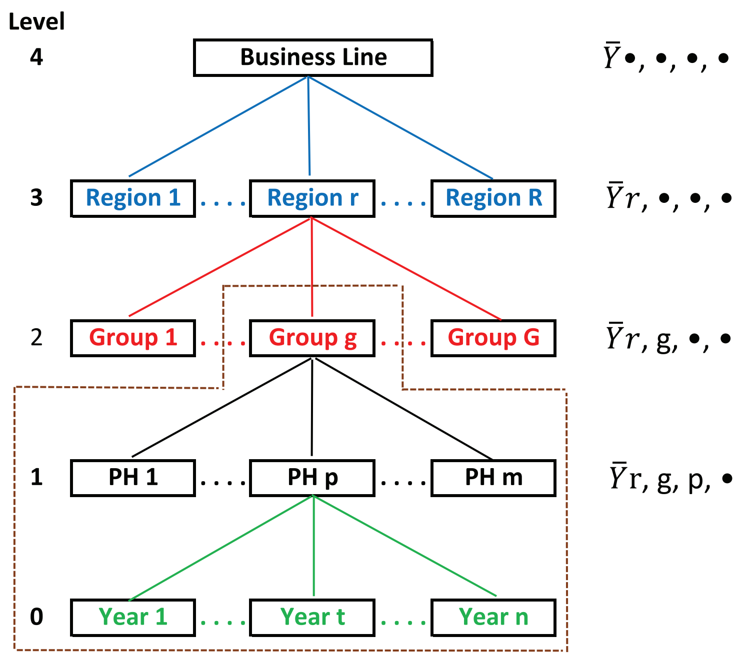


Figure 1.1: Five-level tree structure

Hierarchical credibility is similar to the classical credibility theory. Actually, it is a generalization of the credibility theory with a tree structure of more levels. Using the same example for illustration, hierarchical credibility can be used to model claim severities/frequencies in a tree structure of five levels. As the tree structure shown in Figure 1.1, level two stands for groups with a number of policyholders for each group, and level three represents regions with a number of groups for each region, operating a specific business line. The non-parametric Bühlmann estimate follows a procedure where the sample means are calculated in a hierarchical order from bottom to top: firstly, the sample mean of the past claim severities/frequencies in level zero for policyholder p of group g in region r is calculated to get the policyholder sample mean $\overline{Y}_{r,g,p,\bullet}$ for all policyholders in level one; secondly, the average of the policyholder sample means $\overline{Y}_{r,g,p,\bullet}$ over all policyholders of group g in region r is computed to get the group sample mean $\overline{Y}_{r,g,\bullet,\bullet}$ (an estimate of the true mean of claim severities/frequencies for group g in region r) for all groups in level two; thirdly, the average of the group sample means $\overline{Y}_{r,g,\bullet,\bullet}$ over all groups in region r is calculated to obtain the region sample mean $\overline{Y}_{r,\bullet,\bullet,\bullet}$ (an estimate of the true mean of claim severities/frequencies for region r) for all regions in level three; and finally, the average of the region sample means $\overline{Y}_{r,\bullet,\bullet,\bullet}$ over all regions is computed to achieve the business line sample mean $\overline{Y}_{\bullet,\bullet,\bullet,\bullet}$ (an estimate of the true mean of claim severities/frequencies for the business line) in level four. Four weights determined by Bühlmann and Gisler (2005), summing to one, are assigned to the policyholder sample mean $\overline{Y}_{r,g,p,\bullet}$, the group sample mean $\overline{Y}_{r,g,\bullet,\bullet}$, the region sample mean $\overline{Y}_{r,\bullet,\bullet,\bullet}$, and the business line sample mean $\overline{Y}_{\bullet,\bullet,\bullet,\bullet}$, respectively. The non-parametric hierarchical credibility estimate for region r, group g, policyholder p and the next year is then obtained by a weighted average of the four sample means. That is, the credibility estimate is determined by the four sample means and corresponding weights, and thus each of the past claim data for all policyholders, groups and regions has a different degree of contribution to the value of the hierarchical credibility estimate.

We notice that the multi-country mortality data also have a hierarchical structure. Therefore, the aim of this project is to generalize the Bühlmann credibility mortality model from a single population to multiple populations by incorporating the hierarchical credibility approach into multi-country mortality data. To apply the four-level (five-level) hierarchical credibility to multi-population mortality data, the levels from bottom to top are year, age, and gender (year, age, gender and country). Same as the Bühlmann credibility model approach to modelling mortality rates (see Tsai and Lin, 2017a, b), there are two strategies for forecasting mortality rates for two and more years, the expanding window (EW) and the moving window (MW). We compare the forecasting performances with an error measure among the three-level, four-level and five-level hierarchical credibility mortality models and the classical Lee-Carter model and its three variations for multiple populations. Numerical illustrations show that the proposed hierarchical credibility approaches produce more accurate forecasts.

The remainder of this project proceeds as follows: Chapter 2 provides the literature review on existing single and multi-population mortality models, including the Lee-Carter model and its extensions, the CBD model and other models; it also reviews the Bühlmann credibility model and the hierarchical credibility model. In Chapter 3, we introduce the four-level and five-level hierarchical credibility approaches to modelling multi-population mortality rates, and apply them to projecting mortality rates for six populations (US male, US female, UK male, UK female, Japan male and Japan female). Chapter 4 compares the forecasting performances of the proposed hierarchical approaches with the Bühlmann credibility mortality model and the Lee-Carter model and its variations for multiple populations. The comparisons are shown in seven figures and three tables. Chapter 5 concludes this project.

Chapter 2

Literature Review

2.1 Mortality

Over the last few decades, lots of mortality models were developed to provide accurate predictions of mortality rates. The Lee-Carter model proposed by Lee and Carter (1992) is a significant milestone. As an effective and most popular mortality model in actuarial literature, it provides long term forecasting by modelling the logarithm of central death rate as a function of a period-specific factor and two age-specific parameters. Another famous mortality model is the Cairns-Blake-Dowd (CBD) model proposed by Cairns et al. (2006). The CBD model focuses on forecasting post-age-60 mortality rates by a two-factor stochastic model where one factor has the same influence on all ages and the other factor has more influence on elder ages than younger ages. Tsai and Yang (2015) introduced a linear relational approach to modelling mortality rates, which is easy to implement and understand, and also provides features and potential applications that are not available in the Lee-Carter and CBD models. Tsai and Lin (2017a, b) incorporated the Bühlmann credibility theory, which is commonly applied in property and casualty insurance, into modelling of mortality rates. Comparing its forecasting performances with the Lee-Carter and CBD models, the model proposed by Tsai and Lin (2017a, b) has better forecasting performances based on the measure of MAPE.

Numerous extensions of the Lee-Carter and CBD models were developed. Renshaw and Haberman (2006) generalized the Lee-Carter model to a non-linear model which includes age-specific cohort effects and age-specific period effects; Li et al. (2009) proposed an extension of the Lee-Carter model that provides more conservative interval forecasts of the central death rate by considering individual differences in each age-period cell; Plat (2009) gave a model that combines some nice features of the Lee-Carter, CBD and Renshaw-Haberman models while eliminating the disadvantages of those models; Mitchell et al. (2013) introduced a model based on the idea of bilinear modelling of age and time from the original Lee-Carter model but it suggested to model the change in the logarithm of central death rate instead of the level of central death rate. Lin et al. (2015) proposed AR-GARCH models to forecast mortality rates for a given age and employed a copula method to capture the inter-age mortality dependence.

Besides the single population mortality models, there are intensive developments in multi-population mortality models. Various extensions of the Lee-Cater model focus on modelling the dependency of multi-population mortality rates. Carter and Lee (1992) proposed the joint-k model to reflect the relation between two populations by using a common time-varying index for both populations. Li and Hardy (2011) suggested that the time-varying index for population j ($j \geq 2$) is linearly related to the time-varying index for the base population, which is called the cointegrated Lee-Carter model. Li and Lee (2005) gave the augmented common factor Lee-Carter model to model and forecast mortality rates for multiple populations in a coherent way, which not only considers the commonalities in the historical experience but also includes the individual differences in the trends.

Mortality rates in different countries might be correlated with each other. Besides the extensions of the Lee-Carter model, there are vast literature focusing on multi-population mortality modelling. Cairns et al. (2011) proposed a Bayesian stochastic mortality model to deal with the dynamics of mortality rates in a pair of populations, which is designed for a large population coupled with a small sub-population. It models the difference in the stochastic factors between two populations using a mean-reverting autoregressive process. Therefore, the mortality forecasts do not diverge over the long run. This model fully allows parameter uncertainty and has flexibility to deal with missing data. Zhou et. al (2014) gave an intuitive extension which models stochastic factors using a vector error correction model. Yang and Wang (2013) suggested a vector error correction model to deal with multi-country longevity risk. Kleinow (2015) introduced a common age effect model to govern the multi-population mortality rates. Copula models have been incorporated into multi-population mortality rates modelling in the last few years. Wang et. al (2015) captured the mortality dependence between multi-country mortality rates with a time-varying copula model. The typical multi-population mortality models, taking the models above as examples, assume the mortality rates across different countries converge in the long time. Chen et al. (2015) thought this assumption is too strong to model short-term mortality rates and proposed a two-stage procedure, an ARMA-GARCH process followed by a one-factor copula model, to model the mortality dependence for multiple populations.

2.2 Credibility

Credibility theory is widely applied in property and casualty insurance such as auto insurance, workers' compensation and fire insurance. Bühlmann (1967) proposed a distribution-free credibility formula to determine the credibility premium based on the past experiences of the risks. The premium is calculated by the weighted average of the collective mean and the individual mean, where the weight is determined by the least expected square deviation of the expected value and its linear estimation. Bühlmann credibility theory has built the foundation of credibility model. Furthermore, Bühlmann and Straub (1970) extended the Bühlmann model by allowing unequal number of exposure units for each risk, which enlarges the applicable situations. For such a model, drivers with different number of years of experiences can be grouped together to determine their premiums for the next year.

The Bühlmann and Bühlmann-Straub credibility models assume claim counts or sizes are independent over their associated risks, which simplifies the premium calculations. However, it is obvious that the assumption is not practical. Over the past few decades, numerous extensions of Bühlmann and Bühlmann-Straub credibility models were proposed to account for the dependency of risks. Jewell (1975) modified the credibility formula by using the collateral data, and built a hierarchical model to include the correlation among the risks. Dannenburg (1995) generalised Jewell's hierarchical model and introduced the crossed classification credibility (CCC) model which governs all risk factors symmetrically, with an assumption that the numbers of risk factors are the same for all contracts. Goulet (2001) proposed a generalized crossed classification credibility model (GCCC) by allowing a various number of risk factors for each contract. To allow the risks to be generally dependent, Wen and Wu (2011) extended the Bühlmann and Bühlmann-Straub models to a regression credibility model by re-building the credibility estimators under a general dependence structure. Yeo and Valdez (2006) and Wen et al. (2009) focused on modelling the dependence caused by common effects. Poon and Lu (2015) studied the Bühlmann-Straub credibility model by considering two kinds of dependences, the dependence among risk factors and the conditional spatial cross-sectional dependence.

Hierarchical structure is frequently used in calculating premiums. The expected aggregate premium of a line of insurance is calculated and then distributed to lower levels, such as regions and individuals. Bühlmann and Gisler (2005) incorporated the idea of hierarchical structure to the credibility theory to achieve the hierarchical credibility model. The observations are assumed to be independent conditioning on their next higher level of risk parameters. It can be seen that the hierarchical credibility is a generalization of the Bühlmann model and the Bühlmann-Straub model with a higher order of tree structure. This project applies the hierarchical credibility approach to modelling multi-population morality rates, which accounts for the dependence among populations and makes decent contributions to the multi-population mortality modelling.

Chapter 3

Hierarchical Credibility Mortality Model

This Chapter applies the hierarchical credibility idea from Bühlmann and Gisler (2005) to propose a hierarchical credibility mortality model for multi-country populations, which is a generalization of the non-parametric Bühlmann credibility mortality model for a single population proposed by Tsai and Lin (2017b).

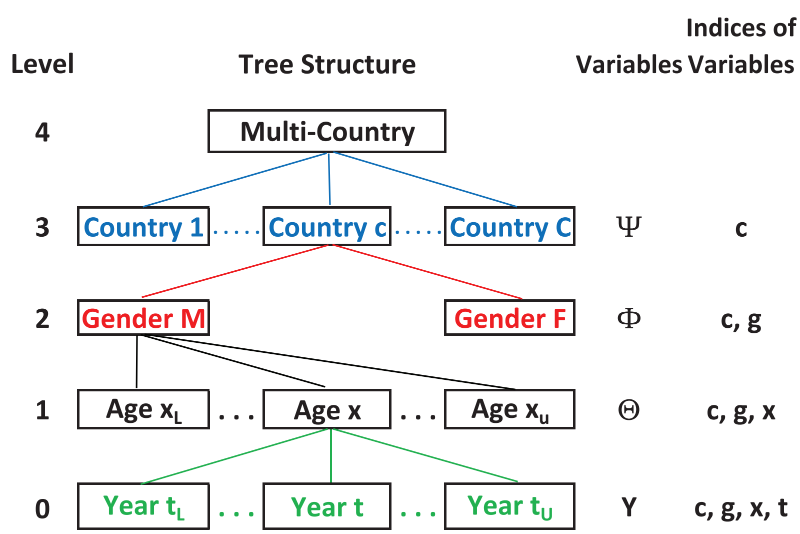
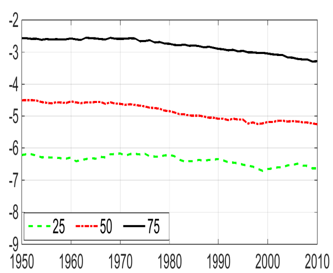
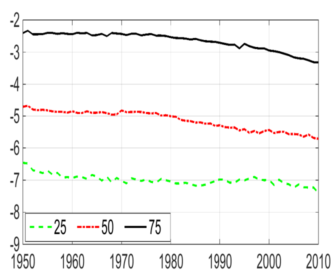


Figure 3.1: A five-level hierarchical tree structure for multi-country populations

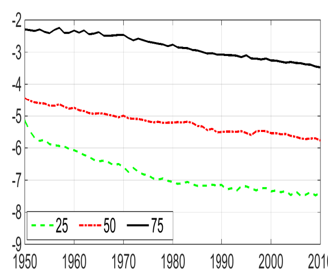
Figure 3.1 structures mortality data for multiple countries in a five-level (levels $0-4$) hierarchical tree with the top level (level 4) being the multi-country, which consists of mortality data from C countries. Each country is broken down into G genders ($G = 2$). Within each gender, there are consecutive ages $x_L, \cdots, x_U$. Finally, each age has yearly data from year t_L to year t_U, which is the bottom level (level 0). Denote Ψ the risk factor related to countries, Φ the risk factor related to countries and genders and Θ the risk factor related to countries, genders and ages. Denote $m_{c,\,g,\,x,\,t}$ the central death rate for country c, gender g, age x and year t. The Lee-Carter model and its three variations for multiple populations use $\ln(m_{c,\,g,\,x,\,t})$ to model mortality rates. Figure 3.2 shows that the historical mortality data $\ln(m_{c,\,g,\,x,\,t})$s from the Human Mortality Database for the US, the UK and Japan display a downward trend over year $t = 1950, \cdots, 2010$ for $x = 25, 50, 75$. As with the Bühlmann credibility mortality model proposed by Tsai and Lin (2017a, b), we apply the hierarchical credibility approach to modelling $Y_{c,\,g,\,x,\,t} = \ln(m_{c,\,g,\,x,\,t}) - \ln(m_{c,\,g,\,x,\,t-1})$, the yearly decrement of the logarithm of central death rate for country c, gender g and age x over $[t-1,\, t]$, in order to eliminate the downward trend (see Figure 3.3) and more importantly make the yearly decrement $Y_{c,\,g,\,x,\,t}$ for all t be independent and identically distributed given c, g and x. Since we will use indices $x = 1, \cdots, X$ and $t = 1, \cdots, T$ for simplifying notations, given mortality data $\ln(m_{c,\,g,\,x,\,t})$s in an age-year rectangle $[x_L,\, x_U] \times [t_L,\, t_U]$ for a population of country c and gender g, the age span $[x_L,\, x_U]$ and the year span $[t_L,\, t_U]$ for $\ln(m_{c,\,g,\,x,\,t})$ correspond to $[1,\, X]$ and $[1,\, T]$ for $Y_{c,\,g,\,x,\,t}$, respectively; that is, $Y_{c,\,g,\,x,\,t} \triangleq \ln(m_{c,\,g,\,x_L+x-1,\,t_L+t}) - \ln(m_{c,\,g,\,x_L+x-1,\,t_L+t-1})$ for $x = 1, \cdots, X$ and $t = 1, \cdots, T$, where $X = x_U - x_L + 1$ and $T = t_U - t_L$. Our goal is to apply a hierarchical credibility approach to the yearly decrements $Y_{c,\,g,\,x,\,t}$s to obtain the hierarchical credibility estimates $\hat{Y}_{c,\,g,\,x,\,T+t}$ for $t = 1, \cdots, \tau$. Then the hierarchical credibility estimate $\ln(\hat{m}_{c,\,g,\,x_L+x-1,\,t_U+\tau}) = \ln(m_{c,\,g,\,x_L+x-1,\,t_U}) + \sum_{t=1}^{\tau} \hat{Y}_{c,\,g,\,x,\,T+t}$.



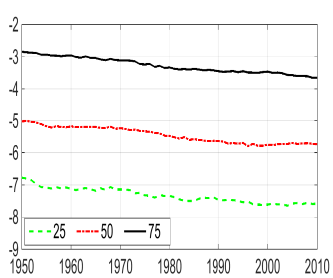
(a) US Males



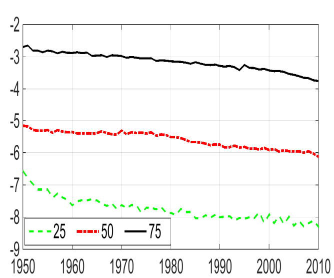
(b) UK Males



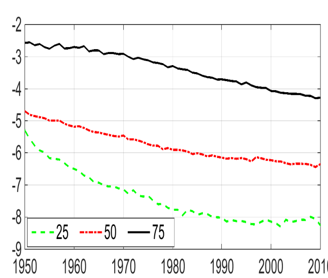
(c) Japan Males



(d) US Females

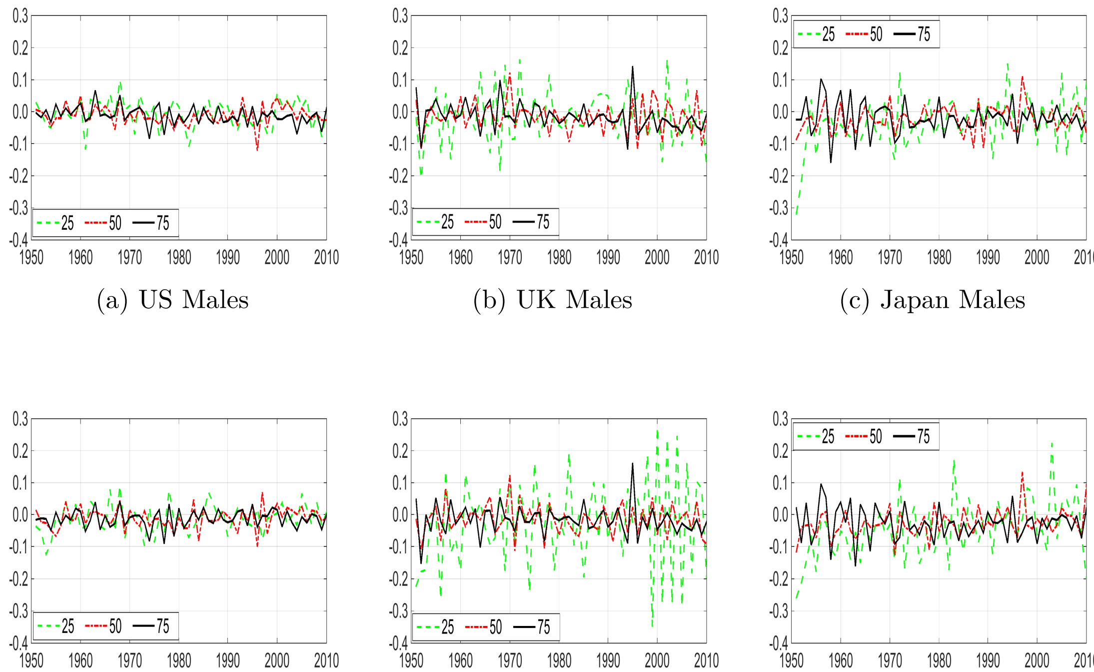


(e) UK Females



(f) Japan Females

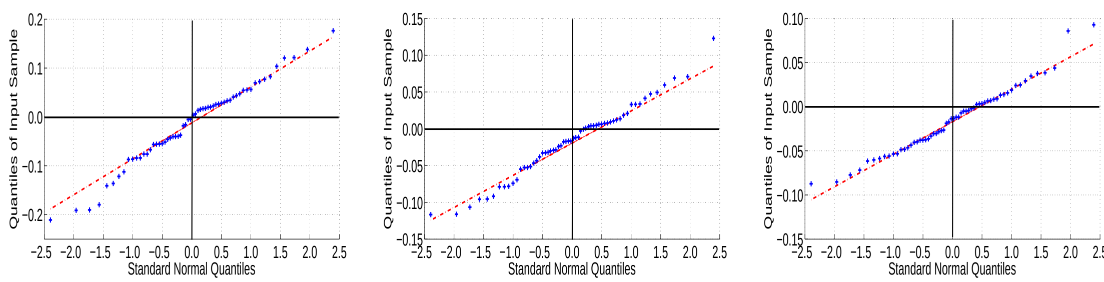
Figure 3.2: $\ln(m_{c,\,g,\,x,\,t})$ against t for age $x = 25$, 50 and 75



(a) US Males (b) UK Males (c) Japan Males

(d) US Females (e) UK Females (f) Japan Females

Figure 3.3: $Y_{c,g,x,t} = \ln(m_{c,g,x,t}) - \ln(m_{c,g,x,t-1})$ against t for age $x = 25$, 50 and 75



(a) age $x = 25$ (b) age $x = 50$ (c) age $x = 75$

Figure 3.4: Q-Q plots of $Y_{c,g,x,t}$ for UK males

3.1 Assumptions and Notations

This section first gives the assumptions and then the notations for conditional means, mean, conditional variances and variances for a hierarchical tree based on Chapter 6 of Bühlmann and Gisler (2005). The assumptions for levels 0–3 of a hierarchical tree are given as follows:

- Level 3: Ψ_c, $c = 1, \cdots, C$, at level 3 are independent and identically distributed;
- Level 2: given Ψ_c at level 3 for fixed country c, $\Phi_{c,g}$, $g = 1, \cdots, G$, at level 2 are independent and identically distributed;

- Level 1: given $\Phi_{c,g}$ at level 2 for fixed country c and gender g, $\Theta_{c,g,x}$, $x = 1, \cdots, X$, at level 1 are independent and identically distributed; and
- Level 0: given $\Theta_{c,g,x}$ at level 1 for fixed country c, gender g and age x, $Y_{c,g,x,t}$, $t = 1, \cdots, T$, at level 0 are independent and identically distributed.

The Q-Q plots of historical yearly decrements for UK males aged 30, 50 and 70 are plotted in Figure 3.4 as an example. The Q-Q plots follow the line $y = x$ which indicates the yearly decrements for all t are independently and identically distributed given country, gender and age.

First, we denote the mean for level 4 and the conditional means for levels 1–3 below:

- Level 4: $\mu_4 \triangleq E[Y_{c,g,x,t}]$, the unconditional expectation of $Y_{c,g,x,t}$ for all years t, ages x, genders g, and countries c at levels 0–3 (the common mean is one of the key assumptions for the classical Bühlmann credibility model);
- Level 3: $\mu_3(\Psi_c) \triangleq E[Y_{c,g,x,t}|\Psi_c]$, the conditional expectation of $Y_{c,g,x,t}$ for all years t, ages x, and genders g at levels 0–2 under country c, given Ψ_c at level 3;
- Level 2: $\mu_2(\Phi_{c,g}) \triangleq E[Y_{c,g,x,t}|\Phi_{c,g}]$, the conditional expectation of $Y_{c,g,x,t}$ at for all years t and ages x at levels 0–1 under gender g and country c, given $\Phi_{c,g}$ at level 2; and
- Level 1: $\mu_1(\Theta_{c,g,x}) \triangleq E(Y_{c,g,x,t}|\Theta_{c,g,x})$, the conditional expectation of $Y_{c,g,x,t}$ for all years t at level 0 under age x, gender g and country c, given $\Theta_{c,g,x}$ at level 1.

By the law of total expectation, we can show that

- $\mu_2(\Phi_{c,g}) = E[Y_{c,g,x,t}|\Phi_{c,g}] = E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}] = E[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}]$,
- $\mu_3(\Psi_c) = E[Y_{c,g,x,t}|\Psi_c] = E\{E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\} = E[\mu_2(\Phi_{c,g})|\Psi_c]$, and
- $\mu_4 = E[Y_{c,g,x,t}] = E\{E\{E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\}\} = E[\mu_3(\Psi_c)]$.

Next, we denote the following conditional variances for levels 1–3:

- Level 3: $\sigma_3^2(\Psi_c) \triangleq Var[\mu_2(\Phi_{c,g})|\Psi_c] = E\{[\mu_2(\Phi_{c,g}) - \mu_3(\Psi_c)]^2|\Psi_c\}$, the conditional variance of $\mu_2(\Phi_{c,g})$ at level 2 given Ψ_c at level 3;
- Level 2: $\sigma_2^2(\Phi_{c,g}) \triangleq Var[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}] = E\{[\mu_1(\Theta_{c,g,x}) - \mu_2(\Phi_{c,g})]^2|\Phi_{c,g}\}$, the conditional variance of $\mu_1(\Theta_{c,g,x})$ at level 1 given $\Phi_{c,g}$ at level 2; and
- Level 1: $\dfrac{\sigma_1^2(\Theta_{c,g,x})}{w_{c,g,x,t}} \triangleq Var[Y_{c,g,x,t}|\Theta_{c,g,x}] = E\{[Y_{c,g,x,t} - \mu_1(\Theta_{c,g,x})]^2|\Theta_{c,g,x}\}$, the conditional variance of $Y_{c,g,x,t}$ at level 0 given $\Theta_{c,g,x}$ at level 1, where $w_{c,g,x,t}$ is a known exposure unit and not necessarily equal for all c, g, x and t (the traditional Bühlmann-Straub credibility model allows unequal exposure units for more applications, but the Bühlmann one requires equal exposure units).

Last, the variance for level 3 and the expected conditional variances for levels 0–2 are denoted as follows:

- Level 3: $\sigma_3^2 \triangleq Var[\mu_3(\Psi_c)]$, the variance of $\mu_3(\Psi_c)$ at level 3;
- Level 2: $\sigma_2^2 \triangleq E[\sigma_3^2(\Psi_c)] = E\{Var[\mu_2(\Phi_{c,\,g})|\Psi_c]\}$, the expectation of the conditional variance of $\mu_2(\Phi_{c,\,g})$ at level 2 given Ψ_c at level 3;
- Level 1: $\sigma_1^2 \triangleq E[\sigma_2^2(\Phi_{c,\,g})] = E\{Var[\mu_1(\Theta_{c,\,g,\,x})|\Phi_{c,\,g}]\}$, the expectation of the conditional variance of $\mu_1(\Theta_{c,\,g,\,x})$ at level 1 given $\Phi_{c,\,g}$ at level 2; and
- Level 0: $\sigma_0^2 \triangleq E[\sigma_1^2(\Theta_{c,\,g,\,x})] = E\{w_{c,\,g,\,x,\,t} \cdot Var[Y_{c,\,g,\,x,\,t}|\Theta_{c,\,g,\,x}]\}$, the expectation of $w_{c,\,g,\,x,\,t}$ times the conditional variance of $Y_{c,\,g,\,x,\,t}$ at level 0 given $\Theta_{c,\,g,\,x}$ at level 1.

It can be shown by the law of total expectation that $\sigma_0^2 = E\{w_{c,\,g,\,x,\,t} \cdot [Y_{c,\,g,\,x,\,t} - \mu_1(\Theta_{c,\,g,\,x})]^2\}$, $\sigma_1^2 = E\{[\mu_1(\Theta_{c,\,g,\,x}) - \mu_2(\Phi_{c,\,g})]^2\}$, $\sigma_2^2 = E\{[\mu_2(\Phi_{c,\,g}) - \mu_3(\Psi_c)]^2\}$, and $\sigma_3^2 = E\{[\mu_3(\Psi_c) - \mu_4]^2\}$.

Note that by the law of total variance, we have

$$\begin{aligned} Var[\mu_2(\Phi_{c,\,g})] &= E\{Var[\mu_2(\Phi_{c,\,g})|\Psi_c]\} + Var\{E[\mu_2(\Phi_{c,\,g})|\Psi_c]\} \\ &= \sigma_2^2 + Var[\mu_3(\Psi_c)] = \sigma_2^2 + \sigma_3^2, \end{aligned}$$

and

$$\begin{aligned} Var[\mu_1(\Theta_{c,\,g,\,x})] &= E\{Var[\mu_1(\Theta_{c,\,g,\,x})|\Phi_{c,\,g}]\} + Var\{E[\mu_1(\Theta_{c,\,g,\,x})|\Phi_{c,\,g}]\} \\ &= \sigma_1^2 + Var[\mu_2(\Phi_{c,\,g})] = \sigma_1^2 + \sigma_2^2 + \sigma_3^2. \end{aligned}$$

3.2 Model Prediction

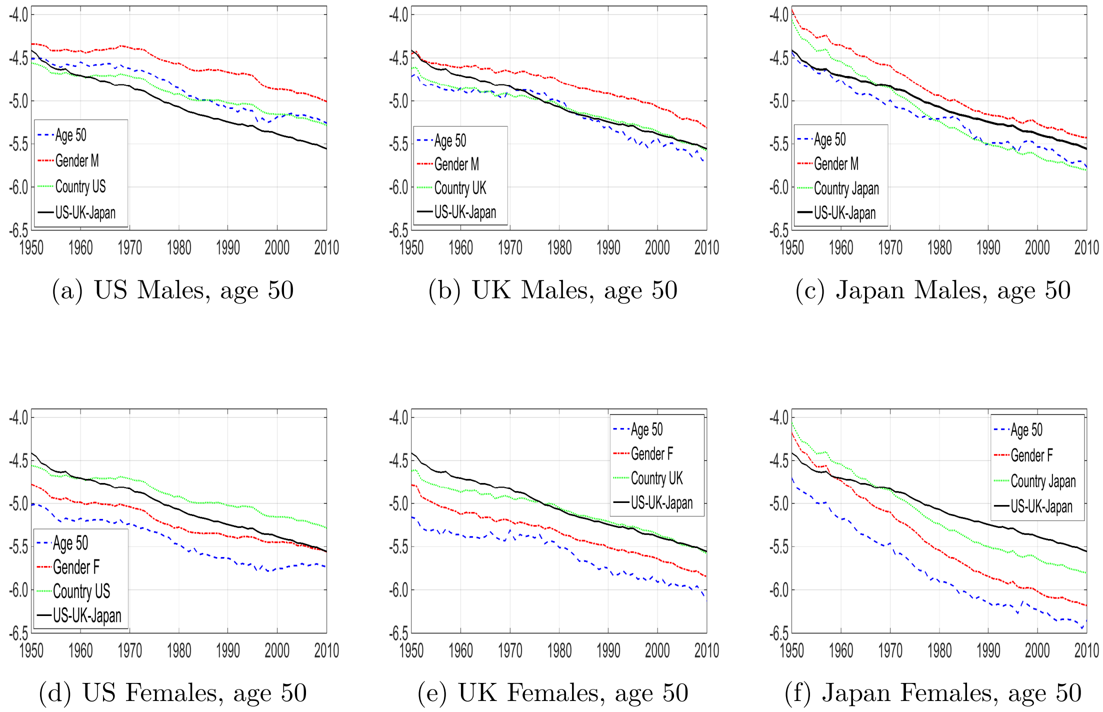
The estimation of structural parameters and determination of credibility factors given in Bühlmann and Gisler (2005) are quite complicated, and thus are placed in Appendix B.

When $w_{c,\,g,\,x,\,t} = 1$ (equal exposure units) for $c = 1, \cdots, C$, $g = 1, \cdots, G$, $x = 1, \cdots, X$, and $t = 1, \cdots, T$, the hierarchical credibility estimate of the decrement over $[T,\, T+1]$ in the logarithm of central death rate for country c, gender g and age x under this special case (refer to (A.6)) is

$$\begin{aligned} \hat{Y}_{c,\,g,\,x,\,T+1} &= \hat{\alpha}^{(1)} \cdot \overline{Y}_{c,\,g,\,x,\,\bullet} + [(1 - \hat{\alpha}^{(1)}) \cdot \hat{\alpha}^{(2)}] \cdot \overline{Y}_{c,\,g,\,\bullet,\,\bullet} \\ &\quad + [(1 - \hat{\alpha}^{(1)}) \cdot (1 - \hat{\alpha}^{(2)}) \cdot \hat{\alpha}^{(3)}] \cdot \overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet} \\ &\quad + [(1 - \hat{\alpha}^{(1)}) \cdot (1 - \hat{\alpha}^{(2)}) \cdot (1 - \hat{\alpha}^{(3)})] \cdot \overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}, \end{aligned} \tag{3.1}$$

where the expressions for $\hat{\alpha}^{(1)}$, $\hat{\alpha}^{(3)}$ and $\hat{\alpha}^{(3)}$ in (3.1) are given in Table B.3. Note that $\hat{Y}_{c,\,g,\,x,\,T+1}$ is the credibility-factor-weighted average of

- $\overline{Y}_{c,\,g,\,x,\,\bullet} = \frac{1}{T}\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} = \frac{1}{T}\sum_{t=1}^{T}[\ln(m_{c,\,g,\,x_L+x-1,\,t_L+t}) - \ln(m_{c,\,g,\,x_L+x-1,\,t_L+t-1})]$ (the average annual decrement of $\{\ln(m_{c,\,g,\,x_L+x-1,\,t_L+t}) : t = 0, \cdots, T\}$, the **age** time series over $[0,\, T]$ for age x under gender g and country c);



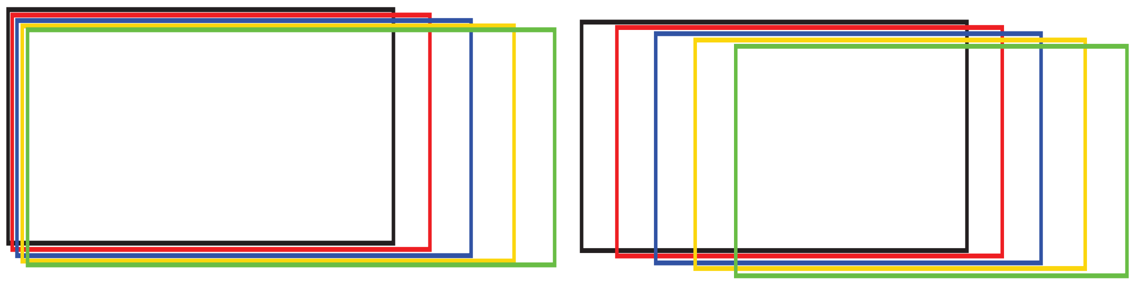
(a) US Males, age 50 (b) UK Males, age 50 (c) Japan Males, age 50

(d) US Females, age 50 (e) UK Females, age 50 (f) Japan Females, age 50

Figure 3.5: Time series for four different levels

- $\overline{Y}_{c,\,g,\,\bullet,\,\bullet} = \frac{1}{T}\sum_{t=1}^{T}\overline{Y}_{c,\,g,\,\bullet,\,t} = \frac{1}{T}\sum_{t=1}^{T}[\ln(m_{c,\,g,\,\bullet,\,t_L+t}) - \ln(m_{c,\,g,\,\bullet,\,t_L+t-1})]$ (the average annual decrement of $\{\ln(m_{c,\,g,\,\bullet,\,t_L+t}) = \frac{1}{X}\sum_{x=1}^{X}\ln(m_{c,\,g,\,x_L+x-1,\,t_L+t}) : t = 0, \cdots, T\}$, the **gender** time series over $[0,\,T]$ for gender g under country c);
- $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet} = \frac{1}{T}\sum_{t=1}^{T}\overline{Y}_{c,\,\bullet,\,\bullet,\,t} = \frac{1}{T}\sum_{t=1}^{T}[\ln(m_{c,\,\bullet,\,\bullet,\,t_L+t}) - \ln(m_{c,\,\bullet,\,\bullet,\,t_L+t-1})]$ (the average annual decrement of $\{\ln(m_{c,\,\bullet,\,\bullet,\,t_L+t}) = \frac{1}{G}\sum_{g=1}^{G}\ln(m_{c,\,g,\,\bullet,\,t_L+t}) : t = 0, \cdots, T\}$, the **country** time series over $[0,\,T]$ for country c); and
- $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet} = \frac{1}{T}\sum_{t=1}^{T}\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,t} = \frac{1}{T}\sum_{t=1}^{T}[\ln(m_{\bullet,\,\bullet,\,\bullet,\,t_L+t}) - \ln(m_{\bullet,\,\bullet,\,\bullet,\,t_L+t-1})]$ (the average annual decrement of $\{\ln(m_{\bullet,\,\bullet,\,\bullet,\,t_L+t}) = \frac{1}{C}\sum_{c=1}^{C}\ln(m_{c,\,\bullet,\,\bullet,\,t_L+t}) : t = 0, \cdots, T\}$, the **multi-country** time series over $[0,\,T]$ for all C countries).

Figure 3.5 displays time series for four different levels: $\{\ln(m_{c,\,g,\,50,\,1950+t}) : t = 0, \cdots, 60\}$, the **age** time series for age 50 under gender g and country c; $\{\ln(m_{c,\,g,\,\bullet,\,1950+t}) : t = 0, \cdots, 60\}$, the **gender** time series for gender g under country c; $\{\ln(m_{c,\,\bullet,\,\bullet,\,1950+t}) : t = 0, \cdots, 60\}$, the **country** time series for country c; and $\{\ln(m_{\bullet,\,\bullet,\,\bullet,\,1950+t}) : t = 0, \cdots, 60\}$, the **multi-country** (US, UK and Japan) time series. Taking the first-order difference of each time series to get the yearly decrement data and then calculating the average of the resulting yearly decrements for each level produces $\overline{Y}_{c,\,g,\,x,\,\bullet}$ (called the age sample mean), $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}$ (called the gender sample mean), $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ (called the country sample mean), and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$ (called the multi-country sample mean).



(a) Expanding Window (EW) (b) Moving Window (MW)

Figure 3.6: Expanding Window (EW) and Moving Window (MW)

The expression in (3.1) gives the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+1}$ for country c, gender g and age x in year $T+1$. To obtain the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+\tau}$ for year $T+\tau$ ($\tau \geq 2$), which is denoted by

$$\begin{aligned}\hat{Y}_{c,\,g,\,x,\,T+\tau} \quad = \quad & \hat{\alpha}_{\tau}^{(1)} \cdot \overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau} + [(1-\hat{\alpha}_{\tau}^{(1)}) \cdot \hat{\alpha}_{\tau}^{(2)}] \cdot \overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau} \\ & +[(1-\hat{\alpha}_{\tau}^{(1)}) \cdot (1-\hat{\alpha}_{\tau}^{(2)}) \cdot \hat{\alpha}_{\tau}^{(3)}] \cdot \overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}^{T+\tau} \\ & +[(1-\hat{\alpha}_{\tau}^{(1)}) \cdot (1-\hat{\alpha}_{\tau}^{(2)}) \cdot (1-\hat{\alpha}_{\tau}^{(3)})] \cdot \overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}, \qquad (3.2)\end{aligned}$$

we adopt the same two strategies, the expanding window (EW) strategy and the moving window (MW) strategy, as those in the Bühlmann credibility mortality model of Tsai and Lin (2017b).

Strategy EW: Expanding window by one year

The EW strategy expands the original fitting year span by $\tau-1$ years to $[1,\, T+\tau-1]$ from $[1,\, T]$ by adding $\{\hat{Y}_{c,\,g,\,x,\,T+1}, \cdots, \hat{Y}_{c,\,g,\,x,\,T+\tau-1}\}$ to the end of $\{Y_{c,\,g,\,x,\,1}, \cdots, Y_{c,\,g,\,x,\,T}\}$ for all c, g and x; see Figure 3.6 (a).

First, the average annual decrement over the year span $[1,\, T+\tau-1]$ for country c, gender g and age x, $\overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau}$, $\tau \geq 2$, is calculated by

$$\overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau} = \frac{1}{T+\tau-1}\left[\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t}\right]. \qquad (3.3)$$

The quantities $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau}$, $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}$, and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}$ are obtained, using the same formula as $\tau = 1$, by

$$\overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau} = \frac{1}{X}\sum_{x=1}^{X} \overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau}, \qquad (3.4)$$

$$\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}^{T+\tau} = \frac{1}{G}\sum_{g=1}^{G} \overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau}, \qquad (3.5)$$

and

$$\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet} = \frac{1}{C}\sum_{c=1}^{C}\overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet}. \tag{3.6}$$

Next, $\hat{\alpha}^{(1)}_{\tau}$, the credibility factor assigned to $\overline{Y}^{T+2}_{c,\,g,\,x,\,\bullet}$, is calculated as

$$\hat{\alpha}^{(1)}_{\tau} = \frac{(T+\tau-1)\cdot\hat{\sigma}^2_1}{(T+\tau-1)\cdot\hat{\sigma}^2_1+\hat{\sigma}^2_0}. \tag{3.7}$$

The other two credibility factors $\hat{\alpha}^{(2)}_{\tau}$ and $\hat{\alpha}^{(3)}_{\tau}$ are given by

$$\hat{\alpha}^{(2)}_{\tau} = \frac{X\cdot\hat{\alpha}^{(1)}_{\tau}\cdot\hat{\sigma}^2_2}{X\cdot\hat{\alpha}^{(1)}_{\tau}\cdot\hat{\sigma}^2_2+\hat{\sigma}^2_1} = \frac{X\,(T+\tau-1)\cdot\hat{\sigma}^2_2}{X\,(T+\tau-1)\cdot\hat{\sigma}^2_2+(T+\tau-1)\cdot\hat{\sigma}^2_1+\hat{\sigma}^2_0}, \tag{3.8}$$

and

$$\begin{aligned}\hat{\alpha}^{(3)}_{\tau} &= \frac{G\cdot\hat{\alpha}^{(2)}_{\tau}\cdot\hat{\sigma}^2_3}{G\cdot\hat{\alpha}^{(2)}_{\tau}\cdot\hat{\sigma}^2_3+\hat{\sigma}^2_2}\\ &= \frac{G\,X\,(T+\tau-1)\cdot\hat{\sigma}^2_3}{G\,X\,(T+\tau-1)\cdot\hat{\sigma}^2_3+X\,(T+\tau-1)\cdot\hat{\sigma}^2_2+(T+\tau-1)\cdot\hat{\sigma}^2_1+\hat{\sigma}^2_0}. \end{aligned}\tag{3.9}$$

Also note that the values of $\hat{\sigma}^2_0$, $\hat{\sigma}^2_1$, $\hat{\sigma}^2_2$, $\hat{\sigma}^2_3$ are unchanged as τ increases.

Finally, the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+\tau}$ for country c, gender g and age x in year $T+\tau$ is obtained by (3.2).

Strategy MW: Moving window by one year

The MW strategy moves the original fitting year span by $\tau-1$ years to $[\tau,\,T+\tau-1]$ from $[1,\,T]$ by appending the hierarchical credibility estimates $\{\hat{Y}_{c,\,g,\,x,\,T+1},\ldots,\hat{Y}_{c,\,g,\,x,\,T+\tau-1}\}$ to and removing $\{\hat{Y}_{c,\,g,\,x,\,1},\cdots,\hat{Y}_{c,\,g,\,x,\,\tau-1}\}$ from $\{Y_{c,\,g,\,x,\,1},\cdots,Y_{c,\,g,\,x,\,T}\}$ for all c, g and x where $\hat{Y}_{c,\,g,\,x,\,t}=Y_{c,\,g,\,x,\,t}$ for $t\le T$; see Figure 3.6 (b).

First, we obtain $\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}$, the average annual decrement over the year span $[\tau,\,T+\tau-1]$ for country c, gender g and age x, by

$$\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} = \frac{1}{T}\sum_{t=\tau}^{T+\tau-1}\hat{Y}_{c,\,g,\,x,\,t}, \tag{3.10}$$

and $\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}$, $\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}$, and $\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$ are calculated using (3.4), (3.5) and (3.6), respectively.

Next, the credibility factor assigned to $\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}$ is achieved by $\hat{\alpha}^{(1)}_{\tau}=\dfrac{T\cdot\hat{\sigma}^2_1}{T\cdot\hat{\sigma}^2_1+\hat{\sigma}^2_0}$. As $\hat{\sigma}^2_0$, $\hat{\sigma}^2_1$, $\hat{\sigma}^2_2$, $\hat{\sigma}^2_3$ are unchanged for all τ, we have $\hat{\alpha}^{(1)}_{\tau}=\hat{\alpha}^{(1)}$, $\hat{\alpha}^{(2)}_{\tau}=\hat{\alpha}^{(2)}$ and $\hat{\alpha}^{(3)}_{\tau}=\hat{\alpha}^{(3)}$. Therefore, $\hat{\alpha}^{(1)}_{\tau}$, $\hat{\alpha}^{(2)}_{\tau}$ and $\hat{\alpha}^{(3)}_{\tau}$ are constant in τ under the MW strategy. Finally, we can calculate $\hat{Y}_{c,\,g,\,x,\,T+\tau}$, the decrement over $[T+\tau-1,\,T+\tau]$ in the logarithm of central death rate for country c, gender g and age x, using (3.2).

3.3 Properties

In this section, we propose some properties for the EW and MW strategies with two propositions under the five-level and four-level hierarchical credibility mortality models. The proofs of Propositions 1 and 2 under these two models are similar, so we only give proofs of Proposition 1 for the five-level hierarchical credibility mortality model in Appendix C.

Proposition 1. (a) *Under the EW and MW strategies, the average of the hierarchical credibility estimates* $\hat{Y}_{c,\,g,\,x,\,T+\tau}$*s over ages* $1,\cdots,X$*, genders* $1,\,\cdots,\,G$*, and countries* $1,\cdots,C$ *for year* $T+\tau$ *equals the average of* $\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}$ *(given in* (3.10)*) over the same age, gender and country spans for year* $T+\tau$*. Specifically,*

$$\frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{c,\,g,\,x,\,T+\tau}=\frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}=\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet},\tau=1,2,\cdots.$$

(b) *Under the EW strategy, the overall average of the hierarchical credibility estimates for year* $T+\tau$*,* $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}$*,* $\tau=1,\,2,\,\cdots$*, are constant in* τ*, i.e.,* $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}$*,* $\tau=2,3,\cdots$*.*
(c) *Under the EW strategy, the hierarchical credibility estimate* $\hat{Y}_{c,\,g,\,c,\,T+\tau}$ *is constant for* $\tau=1,\,2,\,\cdots$*; that is,* $\hat{Y}_{c,\,g,\,x,\,T+\tau}=\hat{Y}_{c,\,g,\,x,\,T+1},\tau=2,3,\cdots$*.*

From Proposition 1 (c), under the EW strategy we have

$$\begin{aligned}\ln(\hat{m}_{c,\,g,\,x_L+x-1,\,t_U+\tau}) &= \ln(m_{c,\,g,\,x_L+x-1,\,t_U})+\sum_{t=1}^{\tau}\hat{Y}_{c,\,g,\,x,\,T+t}\\ &= \ln(m_{c,\,g,\,x_L+x-1,\,t_U})+(\hat{Y}_{c,\,g,\,x,\,T+1})\cdot\tau,\end{aligned}$$

which shows that $\ln(\hat{m}_{c,\,g,\,x_L+x-1,\,t_U+\tau})$ is a linear function of τ with slope $\hat{Y}_{c,\,g,\,x,\,T+1}$ and intercept $\ln(m_{c,\,g,\,x_L+x-1,\,t_U})$.

Proposition 2 below gives the properties under a four-level hierarchical credibility mortality model.

Proposition 2. (a) *Under the EW and MW strategies, the average of the hierarchical credibility estimates* $\hat{Y}_{g,\,x,\,T+\tau}$*s over ages* $1,\cdots,X$ *and genders* $1,\,\cdots,\,G$ *for year* $T+\tau$ *equals the average of* $\overline{Y}^{T+\tau}_{g,\,x,\,\bullet}$ *over the same age and gender spans for year* $T+\tau$*. Specifically,*

$$\frac{1}{G\cdot X}\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{g,\,x,\,T+\tau}=\frac{1}{G\cdot X}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}^{T+\tau}_{g,\,x,\,\bullet}=\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet},\qquad\tau=1,2,\cdots.$$

(b) *Under the EW strategy, the overall average of the hierarchical credibility estimates for year* $T+\tau$*,* $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet}$*,* $\tau=1,\,2,\,\cdots$*, are constant in* τ*, i.e.,* $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet}$*,* $\tau=2,3,\cdots$*.*

(c) *Under the EW strategy, the hierarchical credibility estimate* $\hat{Y}_{g,\,c,\,T+\tau}$ *is constant for* $\tau=1,\,2,\,\cdots$*; that is,* $\hat{Y}_{g,\,x,\,T+\tau}=\hat{Y}_{g,\,x,\,T+1}$*,* $\tau=2,3,\cdots$*.*

Chapter 4

Numerical Illustrations

This chapter applies the model and parameter estimations introduced in Chapter 3 to forecasting mortality rates for both genders of three developed countries for illustrations. The mortality data are obtained from the Human Mortality Database (HMD). We fit three-level, four-level, and five-level hierarchical credibility models with a wide age span and a series of fitting year spans, and make out-of-sample forecasts for future consecutive years. The same data set is also fitted to the classical Lee-Carter model, the joint-k, the cointegrated, and the augmented common factor Lee-Carter models (please refer to Appendix A) with six populations and two populations of each of three countries, respectively. The forecasting performance is measured by the average of mean absolute percentage error (AMAPE), which shows that all of the three-level, four-level and five-level hierarchical credibility mortality models overall outperform the classical and three multi-population Lee-Carter models.

The structure of this chapter is as follows. Section 4.1 gives the assumptions for numerical examples. In Section 4.2, the definition and formula of AMAPE are presented. Lastly, seven figures and three tables are provided in Section 4.3 to numerically compare the forecasting performances of the underlying mortality models; some informative observations from the figures and tables are also summarized in this section.

4.1 Model Assumptions

4.1.1 Mortality Data

The mortality data applied in this project come from the Human Mortality Database (HMD, www.mortality.org), which is a public database providing detailed mortality and population data for thirty-nine countries or areas around the world. This database contains data such as birth counts, death counts, and life tables for male or female. The six populations used in this project are both genders of the US, the UK, and Japan.

For each of the three-level, four-level, and five-level hierarchical credibility models, the EW and MW strategies are adopted to forecast mortality rates for 10, 20 and 30 years. We denote EW-l and MW-l for the EW and MW strategies, respectively, under the l-level hierarchical credibility model where $l = 3, 4, 5$. The three-level hierarchical credibility model with a tree structure of year, age and population (male or female of a country) is applied to each of the six populations. Under the four-level (five-level) hierarchical credibility model, the tree structure from the bottom to the top is specified as year, age, gender and country (year, age, gender, country and multi-country), and is applied to two genders of each of three countries (all six populations). To compare the forecasting performance of the three-level hierarchical credibility models, the mortality data for each of the six populations are respectively fitted to the classical Lee-Carter model, which is denoted as LC1-Ind; and to compare the forecasting performance of the four-level (five-level) hierarchical credibility model, the mortality data for both genders of each of three countries (six populations) are respectively fitted to the joint-k, the cointegrated, and the augmented common factor Lee-Carter models with the male of a country (the US male) as the base population for the cointegrated Lee-Carter model, which are denoted by LC2-JoK, LC2-CoI and LC2-ACF (LC6-JoK, LC6-CoI and LC6-ACF).

4.1.2 Age Span, Fitting Year Span, and Forecasting Year Span

Let $[T_1, T_2]$ be the study period where mortality rates are available. Assume that we stand at the end of year t_U and would like to fit the models with mortality data in the rectangle $[x_L, x_U] \times [t_L, t_U]$, project mortality rates for the rectangle $[x_L, x_U] \times [t_U + 1, T_2]$, and evaluate the forecast performances of the underlying mortality models. Below are detailed assumptions.

- For the age span $[x_L, x_U]$, we choose 20-84, i.e. $x_L = 20$ and $x_U = 84$ and the length of the age span $m = 65$.
- For the study period $[T_1, T_2]$, we use a 63-year period $1951 - 2013$, i.e., $T_1 = 1951$ and $T_2 = 2013$ which is the most recent year where the mortality rates are available for both genders of the US, the UK, and Japan.
- For the fitting year spans $[t_L, t_U]$, a series of periods, $[1951, t_U], \ldots, [t_U - 4, t_U]$, are selected where t_U takes three values, 1983, 1993 and 2003, and the shortest period $[t_U - 4, t_U]$ is five years.
- For the forecasting year span $[t_U + 1, T_2]$, we choose [2004, 2013] (10 years wide), [1994, 2013] (20 years wide) and [1984, 2013] (30 years wide).

The following table displays the three forecasting year spans conducted in this project.

Table 4.1: Summary of the fitting and forecasting year spans

Fitting year spans	$[t_L,\, t_U]$	[1951, 2003] [1952, 2003] $\vdots$ [1999, 2003]	[1951, 1993] [1952, 1993] $\vdots$ [1989, 1993]	[1951, 1983] [1952, 1983] $\vdots$ [1979, 1983]
Ending year of fitting year spans	t_U	2003	1993	1983
Number of fitting year spans	J	49	39	29
Forecasting year spans	$[t_U+1, T_2]$	$[2004, 2013]$	$[1994, 2013]$	$[1984, 2013]$
Width of forecasting year spans	$T_2 - t_U$	10	20	30

4.2 Forecasting Measurements

We compare the forecasting performances of the hierarchical credibility models with the classical, the joint-k, the cointegrated, and the augmented common factor Lee-Carter models by the measure of mean absolute percentage error (MAPE), which is a common measurement as used in Tsai and Lin (2017a, b). Specifically, the $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]}$ for gender g and country c over the forecasting age-year window $[x_L,\, x_U] \times [t_U+1,\, T_2]$ with the fitting year span $[t_L,\, t_U]$ is defined by

$$MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]} = \frac{1}{T_2 - t_U} \cdot \frac{1}{m} \sum_{\tau=1}^{T_2-t_U} \sum_{x=x_L}^{x_U} \left| \frac{\hat{q}_{c,\, g,\, x,\, t_U+\tau} - q_{c,\, g,\, x,\, t_U+\tau}}{q_{c,\, g,\, x,\, t_U+\tau}} \right|, \tag{4.1}$$

where $\hat{q}_{c,\, g,\, x,\, t_U+\tau}$ is the forecast mortality rate, $q_{c,\, g,\, x,\, t_U+\tau}$ is the observed mortality rate, and $\hat{q}_{c,\, g,\, x,\, t_U+\tau} = 1 - e^{-\hat{m}_{c,\, g,\, x,\, t_U+\tau}}$ is based on the assumption of constant force of mortality over $[x,\, x+1] \times [t_U+\tau,\, t_U+\tau+1]$. The value of $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]}$ largely depends on the fitting year span $[t_L,\, t_U]$. To evaluate the overall forecasting performance of a mortality model, we average the $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]}$ over the fitting year spans $[t_L,\, t_U]$ for $t_L = T_1,\, T_1+1, \cdots, t_U-4$ to get the $AMAPE_{c,\, g,\, [t_U+1,\, T_2]}$. Specifically,

$$AMAPE_{c,\, g,\, [t_U+1,\, T_2]} = \frac{1}{t_U - 4 - T_1 + 1} \sum_{t_L=T_1}^{t_U-4} MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]}.$$

A smaller $AMAPE_{c,\, g,\, [t_U+1,\, T_2]}$ produced from a mortality model indicates an overall more accurate forecast for the period $[t_U+1,\, T_2]$. The underlying mortality models in this project will be ranked based on the $AMAPE_{c,\, g,\, [t_U+1,\, T_2]}$.

4.3 Numerical Results

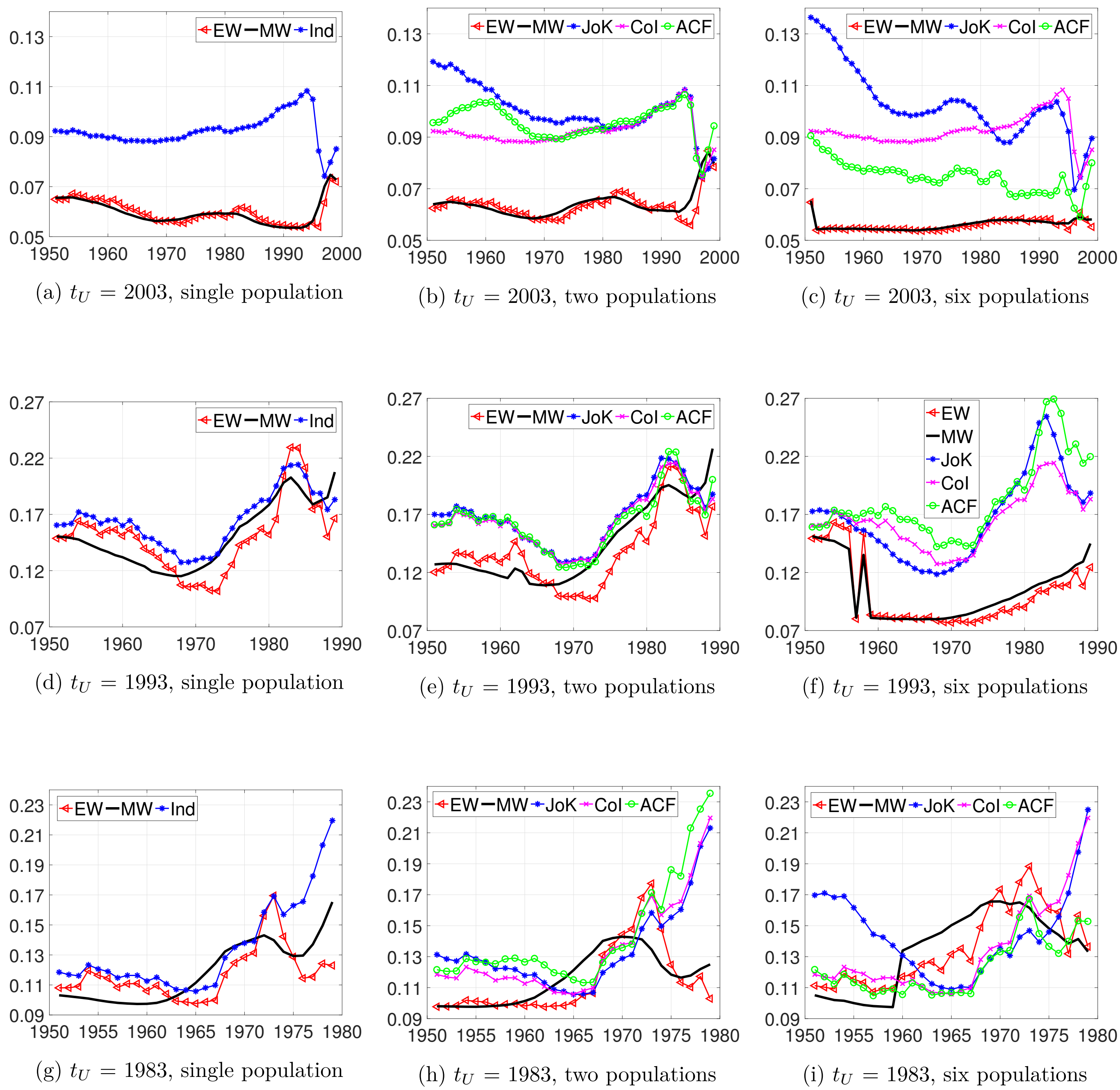
Based on mortality data from the Human Mortality Database, we produce seven figures and construct three tables for three forecasting year spans [2004, 2013], [1994, 2013] and

[1984, 2013]. Figures 4.1–4.6 display the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L, where $t_U = 2003, 1993, 1983$ for three forecasting year spans, for each of six populations (US male, US female, UK male, UK female, Japan male and Japan female), respectively, and Figure 4.7 exhibits the average of the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ over the six populations against t_L. Within each figure, the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ plots for three forecasting year spans are shown in three rows, and those for a single population, two populations, and six populations to which the models are respectively fitted are given in three columns. Specifically, the first column displays the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L (the start year of the fitting year span) for the three-level hierarchical credibility model and the classical Lee-Carter model for a single population; the second column exhibits the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L for the four-level hierarchical credibility model and the three Lee-Carter models for two populations; and the third column presents the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L for the five-level hierarchical credibility model and the three Lee-Carter models for six populations. Note that the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ under the three Lee-Carter models for two populations and six populations are different. For the two-population graphs, the joint-k, the cointegrated, and the augmented common factor Lee-Carter models are fitted into both genders of a country; for example, the three Lee-Carter models are applied to both genders of the US, and the corresponding $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ for the US male and female are given in (b), (e) and (h) of Figures 4.1 and 4.2, respectively. For the six-population graphs, the joint-k, the cointegrated, and the augmented common factor Lee-Carter models are fitted to all six populations, and the corresponding $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ for the US male and female are given in (c), (f) and (i) of Figures 4.1 and 4.2, respectively. Observations from the figures are summarized below.

- The $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values for the hierarchical credibility models and Lee-Carter models are generally decreasing in t_U, which means the wider the forecasting period, the higher the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ value.

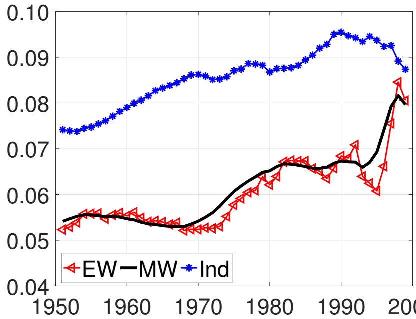
- The $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values for all models and two strategies are neither monotonically decreasing nor increasing with t_L (the start year of the fitting year span), i.e., the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values depend on the length and location of the fitting year span. The pattern of $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curves largely depends on the data set. For example, the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curve for the MW strategy in Figure 4.2 (h) increases for the first twenty t_L values, then decreases for the next five t_L values, and finally increases in the last few t_L values. However, Figure 4.3 (h) shows the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curve for the MW strategy is decreasing in t_L except for a few t_L values at both ends of the domain. Since we do not know which fitting year span will result in the lowest $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$, we calculate the $AMAPE_{c,\,g,\,[t_U+1,\,T_2]}$ (the average of the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ over all values of t_L) and use it to rank the underlying mortality models for their forecasting performances.

- The $MAPE_{c,\,g,\,[t_U+1,\,T_2]}^{[t_L,\,t_U]}$ values for the EW and MW strategies of the hierarchical credibility models are overall lower than those for the Lee-Carter models except for a few cases, for example Figure 4.1 (i) for the US male, Figure 4.2 (i) for the US female, and Figure 4.5 (h) for the Japan male; the corresponding $AMAPE_{c,\,g,\,[t_U+1,\,T_2]}$ values are shown in Table 4.4. Moreover, the $MAPE_{c,\,g,\,[t_U+1,\,T_2]}^{[t_L,\,t_U]}$ curves for the MW strategy look smoother in t_L than the EW one.

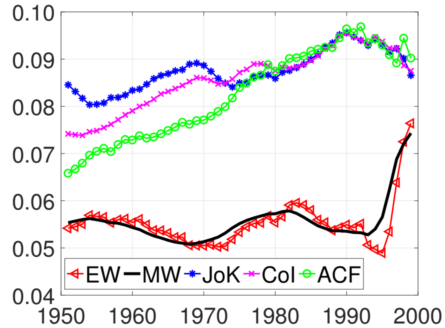


(a) $t_U = 2003$, single population
(b) $t_U = 2003$, two populations
(c) $t_U = 2003$, six populations
(d) $t_U = 1993$, single population
(e) $t_U = 1993$, two populations
(f) $t_U = 1993$, six populations
(g) $t_U = 1983$, single population
(h) $t_U = 1983$, two populations
(i) $t_U = 1983$, six populations

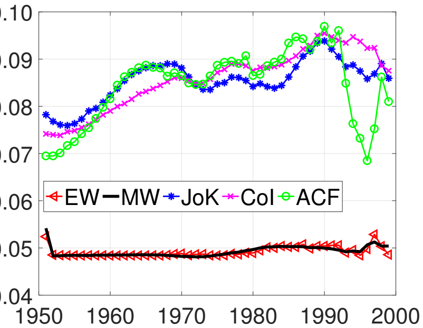
Figure 4.1: $MAPE_{c,\,g,\,[t_U+1,\,2013]}^{[t_L,\,t_U]}$ against t_L for US Male with age span 20–84



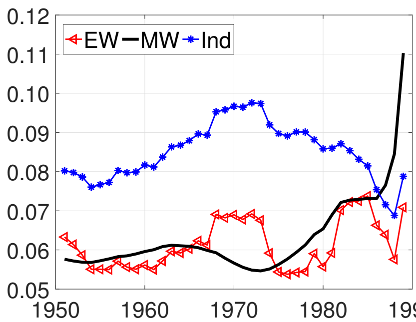
(a) $t_U = 2003$ single population



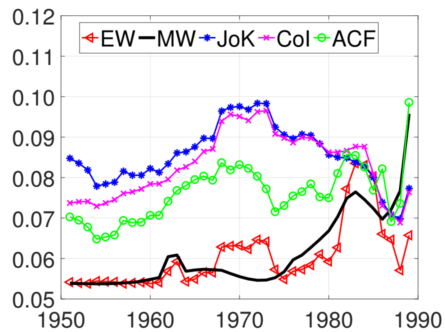
(b) $t_U = 2003$, two populations



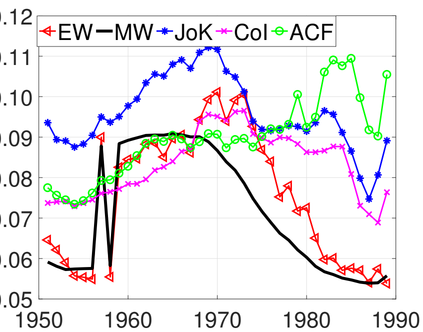
(c) $t_U = 2003$, six populations



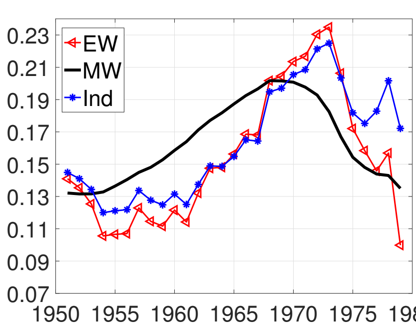
(d) $t_U = 1993$, single population



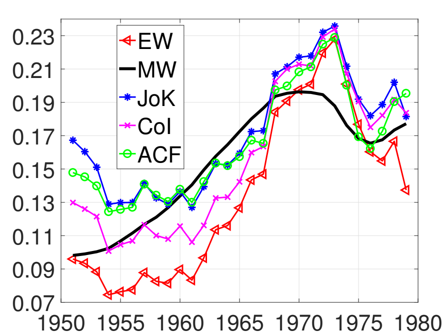
(e) $t_U = 1993$, two populations



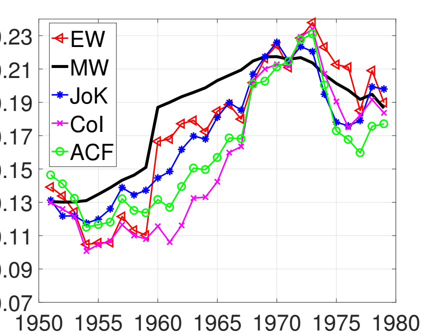
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

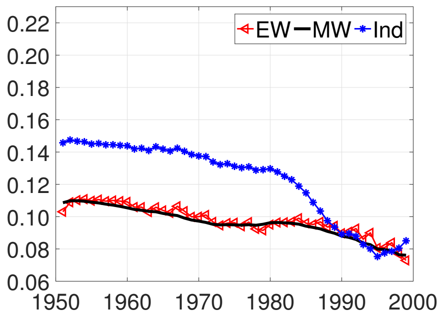


(h) $t_U = 1983$, two populations

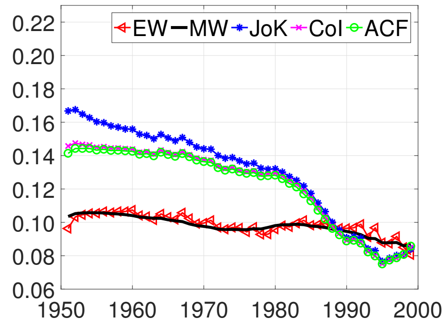


(i) $t_U = 1983$, six populations

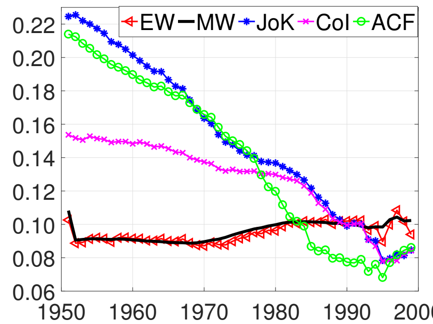
Figure 4.2: $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, 2013]}$ against t_L for US Female with age span 20–84



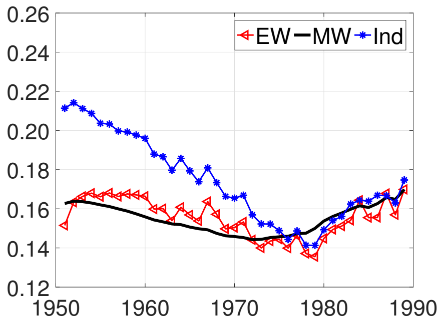
(a) $t_U = 2003$, single population



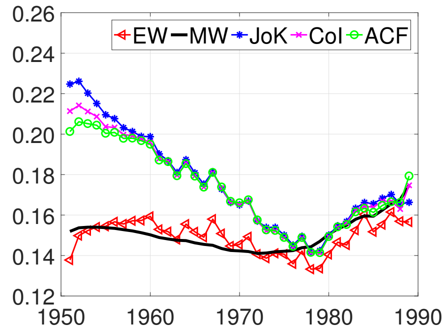
(b) $t_U = 2003$, two populations



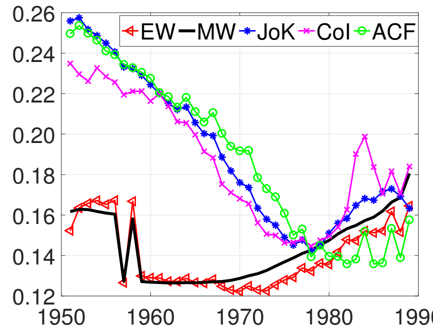
(c) $t_U = 2003$, six populations



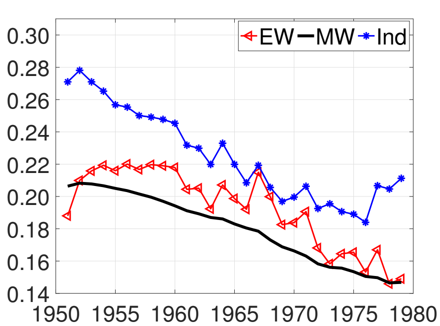
(d) $t_U = 1993$, single population



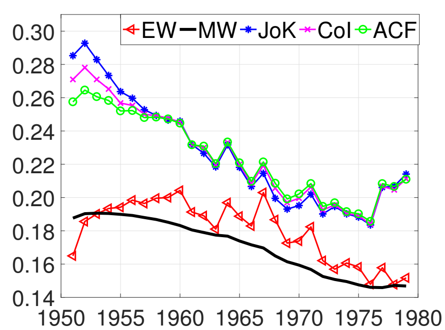
(e) $t_U = 1993$, two populations



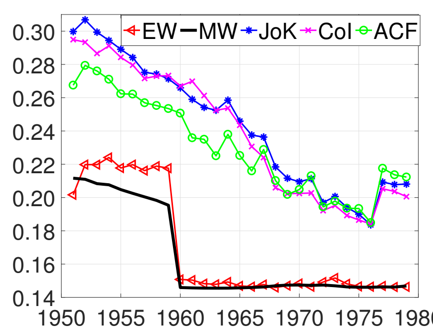
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

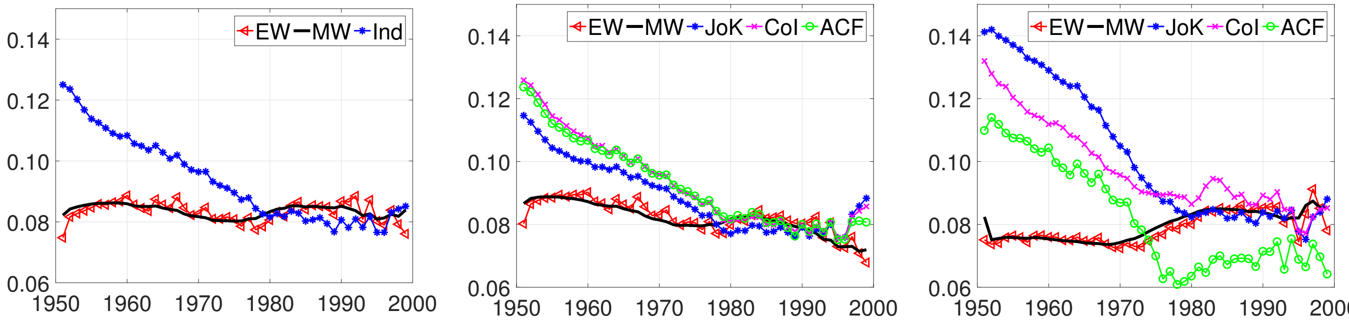


(h) $t_U = 1983$, two populations

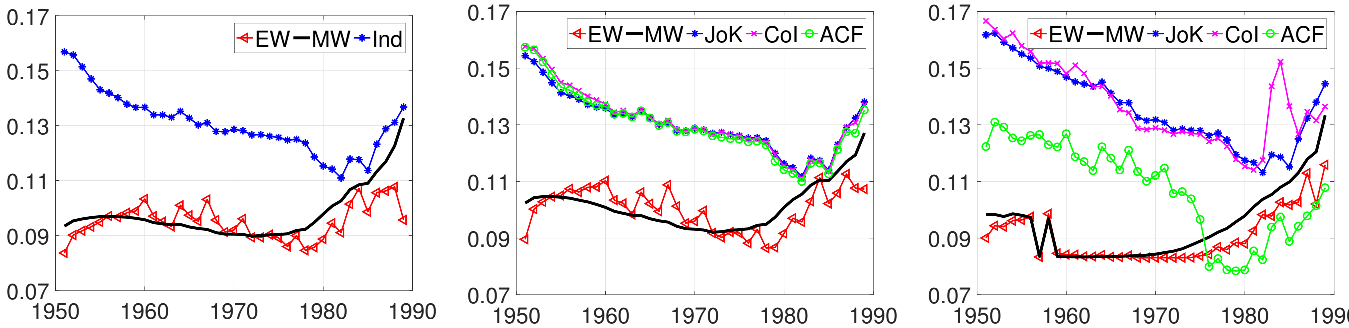


(i) $t_U = 1983$, six populations

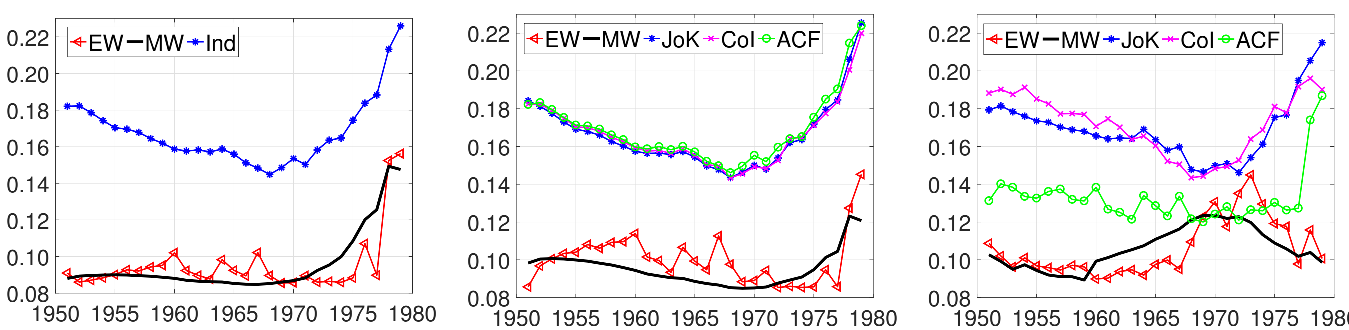
Figure 4.3: $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, 2013]}$ against t_L for UK Male with age span 20–84



(a) $t_U = 2003$, single population (b) $t_U = 2003$, two populations (c) $t_U = 2003$, six populations

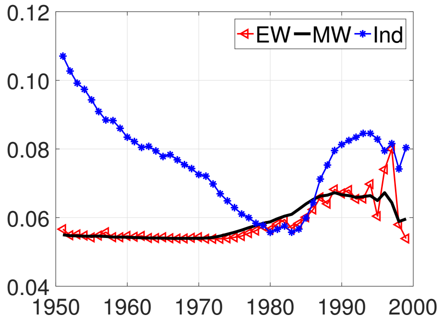


(d) $t_U = 1993$, single population (e) $t_U = 1993$, two populations (f) $t_U = 1993$, six populations

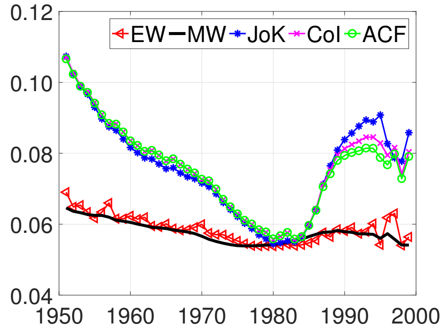


(g) $t_U = 1983$, single population (h) $t_U = 1983$, two populations (i) $t_U = 1983$, six populations

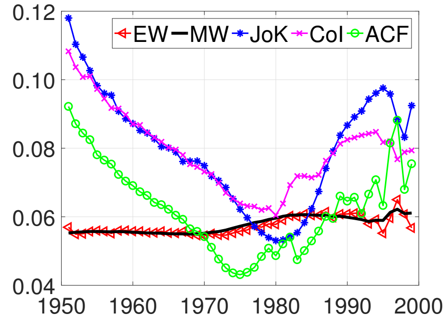
Figure 4.4: $MAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for UK Female with age span 20–84



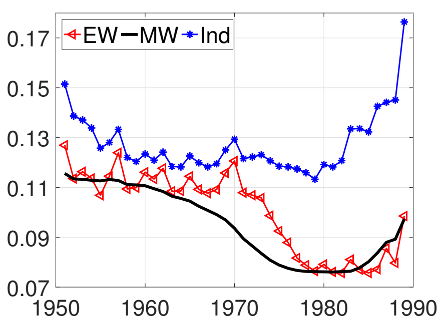
(a) $t_U = 2003$, single population



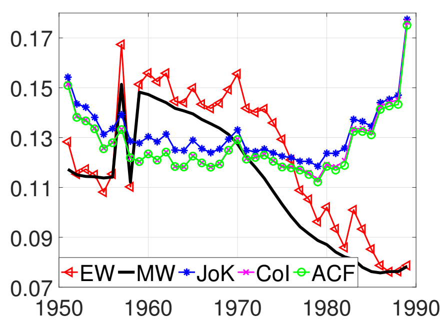
(b) $t_U = 2003$, two populations



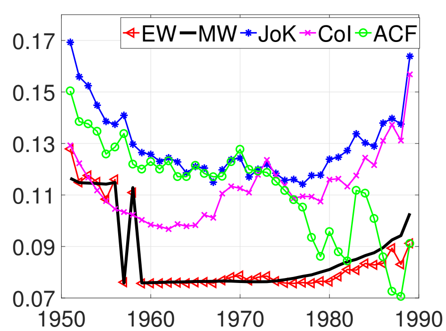
(c) $t_U = 2003$, six populations



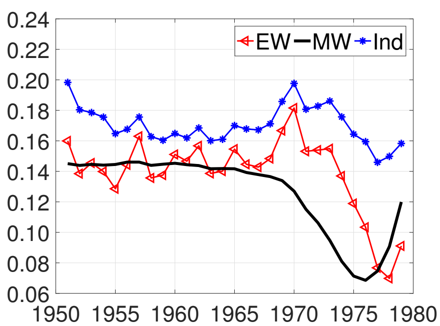
(d) $t_U = 1993$, single population



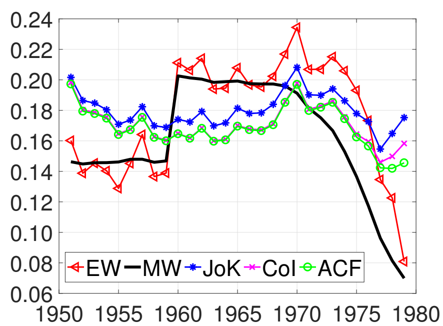
(e) $t_U = 1993$, two populations



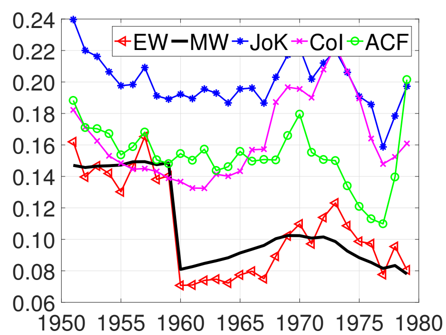
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

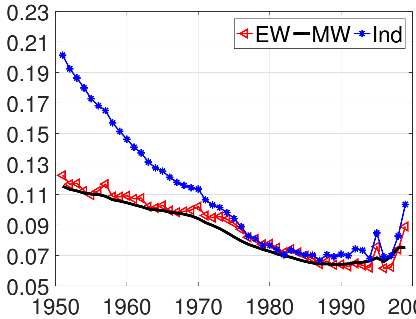


(h) $t_U = 1983$, two populations

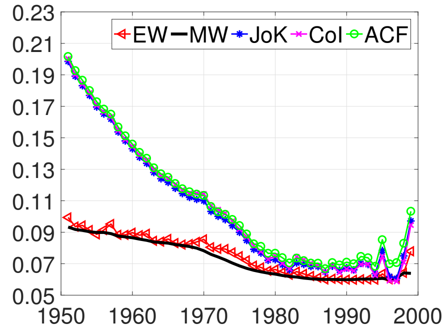


(i) $t_U = 1983$, six populations

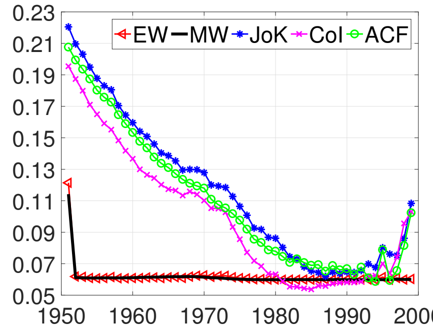
Figure 4.5: $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, 2013]}$ against t_L for Japan Male with age span 20–84



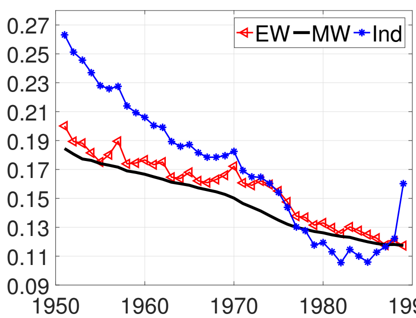
(a) $t_U = 2003$, single population



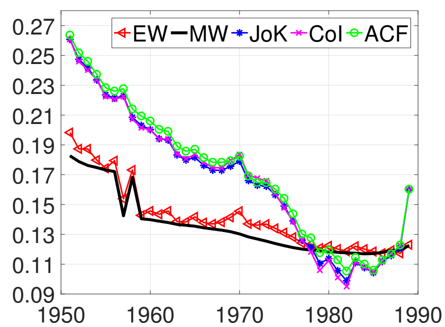
(b) $t_U = 2003$, two populations



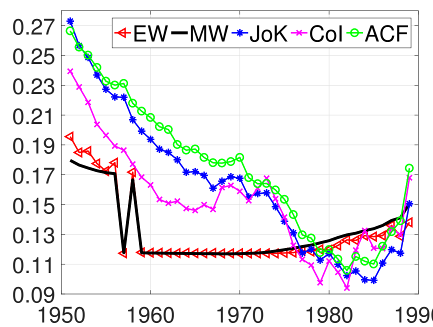
(c) $t_U = 2003$, six populations



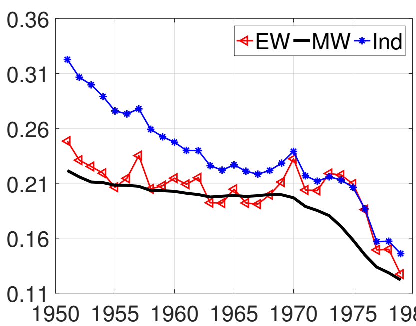
(d) $t_U = 1993$, single population



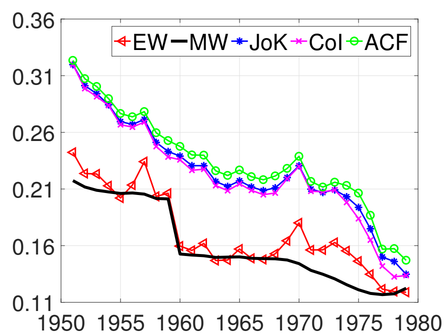
(e) $t_U = 1993$, two populations



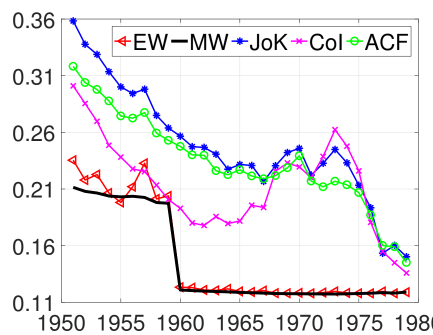
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

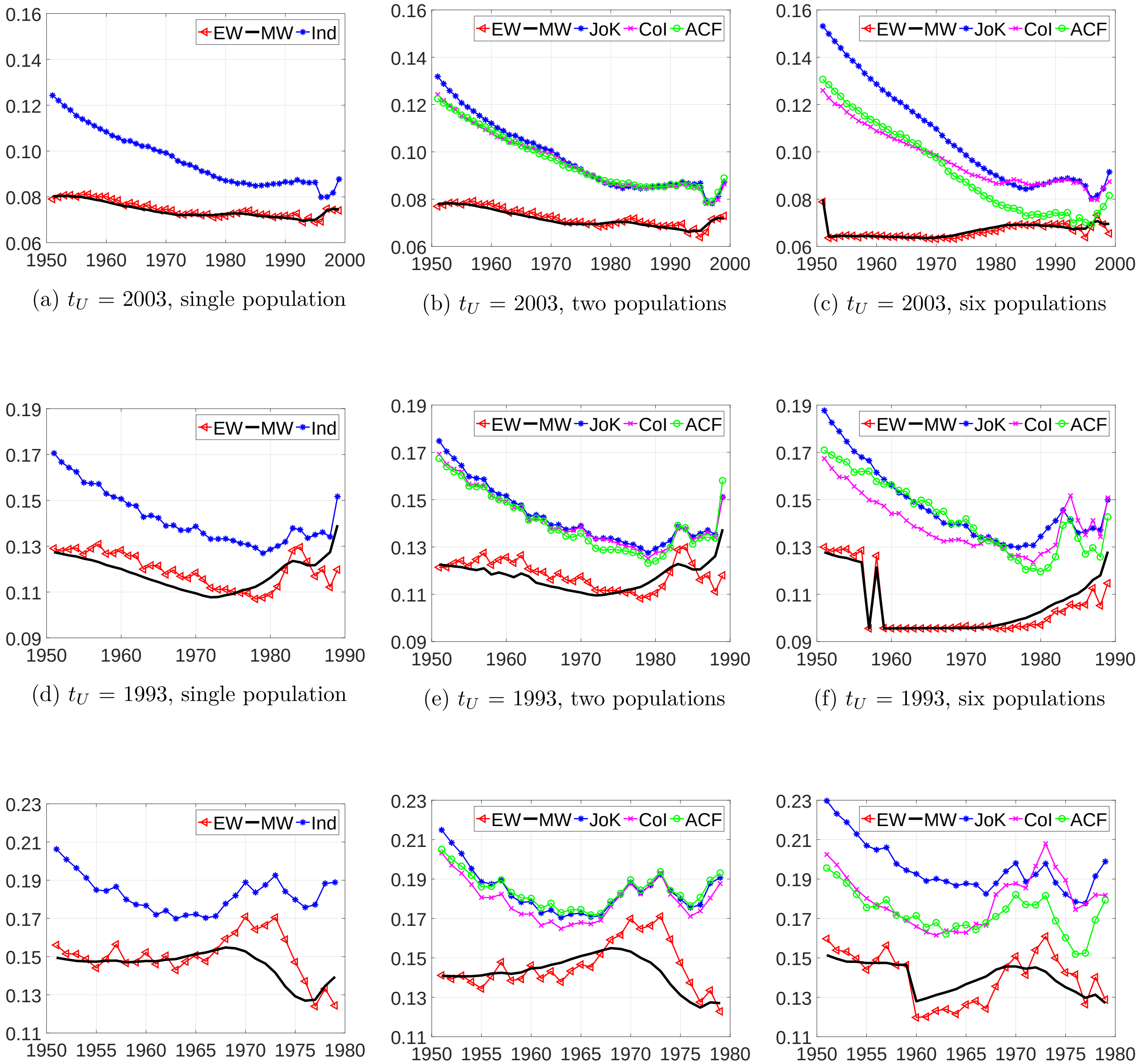


(h) $t_U = 1983$, two populations



(i) $t_U = 1983$, six populations

Figure 4.6: $MAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for Japan Female with age span 20–84



(a) $t_U = 2003$, single population

(b) $t_U = 2003$, two populations

(c) $t_U = 2003$, six populations

(d) $t_U = 1993$, single population

(e) $t_U = 1993$, two populations

(f) $t_U = 1993$, six populations

(g) $t_U = 1983$, single population

(h) $t_U = 1983$, two populations

(i) $t_U = 1983$, six populations

Figure 4.7: The average of $MAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, 2013]}$ over six populations against t_L with age span 20–84

Below are observations from Tables 4.2–4.4, which exhibit the $AMAPE_{c,\, g,\, [t_U+1,\, 2013]}$, the "Avg 2" (the average of the $AMAPE_{c,\, g,\, [t_U+1,\, 2013]}$ over both genders of a country) and the "Avg 6" (the average of the $AMAPE_{c,\, g,\, [t_U+1,\, 2013]}$ over all six populations) for $t_U = 2003$, 1993 and 1983, respectively.

- Observing the values in Tables 4.2–4.4, it is obvious that the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ values become larger as the length of the forecasting year span increases from 10 to 30 years, which is consistent with the observations in Figures 4.1–4.6 that the $MAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values increase in the width of the forecasting year span.
- Based on the average of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations ("Avg 6"), we observe that the hierarchical credibility model with more levels provides more accurate forecasting performance. However, this conclusion does not apply to "Avg 2" (the average of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over both genders of a country). The forecasting performance ranking based on "Avg 2" depends on country. For example, according to the "Avg 2" in Table 4.4 for the forecasting period $[1984, 2013]$, the five-level hierarchical credibility model has the worst forecast accuracy and the four-level one performs the best for the US, whereas the five-level hierarchical credibility model has the best forecast accuracy and the four-level one performs the worst for Japan.
- The $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ values and their averages "Avg 6" and "Avg 2" under both of the EW and MW strategies, given the same level of the hierarchical credibility model, are close to each other.
- Among the three Lee-Carter models applied to six populations, "Avg 6" (the average of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over all six populations) shows that the augmented common factor model LC6-ACF is the most accurate for the forecasting periods [2004, 2013] and [1984, 2013], the cointegrated model LC6-CoI performs the best for the forecasting period [1994, 2013], and the joint-k model LC6-JoK is the least accurate for all three forecasting periods. Moreover, the LC2-JoK model outperforms the LC6-JoK model for all three forecasting periods, the LC2-CoI model is better than the LC6-CoI model for $[2004, 2013]$ and $[1984, 2013]$, and the LC2-ACF model is worse than the LC6-ACF model for $[2004, 2013]$ and $[1984, 2013]$.
- From Tables 4.2–4.4, we observe that most of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ values under the hierarchical credibility models for all three forecasting periods and six populations are lower than those under the Lee-Carter models. As a result, the averages of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over both genders of a country and over six populations for all the three forecasting periods under the hierarchical credibility models are far lower than those under the Lee-Carter models. For example, for the 30-year forecasting period $[1984, 2013]$, the averages of the $AMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations for the joint-k, the cointegrated, and the augmented common factor Lee-Carter models applied to six populations and two populations and for the classical Lee-Carter model applied to a single population are 19.57% (LC6-Jok), 17.97% (LC6-CoI), 17.26% (LC6-ACF), 18.40% (LC2-Jok), 17.92% (LC2-CoI), 18.41% (LC2-ACF) and 18.25% (LC1-Ind), respectively, whereas those for the EW and MW strategies under the five/four/three-level hierarchical credibility models are 14.01% (EW-5), 14.02% (MW-5), 14.60% (EW-4), 14.28% (MW-4), 15.03% (EW-3) and 14.55% (MW-3), re-

spectively. Therefore, the numerical illustrations highly support the conclusion that the hierarchical credibility models outperform the Lee-Carter models.

AMAPE	**Country**	**US**			**UK**			**Japan**		
Model	**Avg 6**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**
	Hierarchical Credibility (HC) model									
EW-5	6.63	5.58	4.92	5.25	9.48	7.87	8.68	5.74	6.20	5.97
MW-5	6.66	5.60	4.93	5.26	9.57	7.95	8.76	5.77	6.16	5.97
EW-4	7.23	6.36	5.53	5.95	9.86	8.24	9.05	5.86	7.52	6.69
MW-4	7.16	6.41	5.56	5.99	9.79	8.17	8.98	5.76	7.27	6.52
EW-3	7.47	6.00	6.04	6.02	9.78	8.33	9.06	5.83	8.85	7.34
MW-3	7.41	5.96	6.14	6.05	9.61	8.36	8.98	5.85	8.54	7.19
	Lee-Carter model applied to all six populations (LC6)									
LC6-JoK	10.61	10.27	8.54	9.40	15.10	10.23	12.66	7.98	11.56	9.77
LC6-CoI	9.69	9.23	8.57	8.90	12.64	9.79	11.22	7.93	9.95	8.94
LC6-ACF	9.22	7.46	8.45	7.96	13.94	8.24	11.09	6.26	10.96	8.61
	Lee-Carter model applied to both genders of a country (LC2)									
LC2-JoK	9.77	10.06	8.75	9.41	12.95	8.89	10.92	7.61	10.39	9.00
LC2-CoI	9.59	9.23	8.57	8.90	12.28	9.32	10.80	7.63	10.50	9.06
LC2-ACF	9.60	9.58	8.25	8.92	12.16	9.26	10.71	7.57	10.77	9.17
	Lee-Carter model applied to a single population (LC1)									
LC1-Ind	9.64	9.23	8.57	8.90	12.28	9.36	10.82	7.63	10.77	9.20

Table 4.2: $AMAPE_{c,\,g,\,[2004,2013]}$s (%)

AMAPE	**Country**	**US**			**UK**			**Japan**		
Model	**Avg 6**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**
	Hierarchical Credibility (HC) model									
EW-5	10.41	10.12	7.56	8.84	14.00	9.05	11.53	8.52	13.18	10.85
MW-5	10.55	10.41	7.13	8.77	14.39	9.45	11.92	8.66	13.24	10.95
EW-4	11.85	13.77	6.05	9.91	14.94	10.03	12.48	12.19	14.10	13.15
MW-4	11.74	14.39	6.12	10.26	15.07	10.14	12.61	11.17	13.56	12.37
EW-3	11.98	14.97	6.14	10.55	15.57	9.50	12.53	10.05	15.63	12.84
MW-3	11.81	15.04	6.32	10.68	15.48	9.79	12.63	9.49	14.75	12.12
	Lee-Carter model applied to all six populations (LC6)									
LC6-JoK	14.71	16.64	9.59	13.11	19.11	13.63	16.37	12.93	16.37	14.65
LC6-CoI	13.98	16.48	8.32	12.40	18.71	13.81	16.26	11.31	15.27	13.29
LC6-ACF	14.25	18.19	8.95	13.57	18.92	10.75	14.83	11.30	17.39	14.34
	Lee-Carter model applied to both genders of a country (LC2)									
LC2-JoK	14.31	16.75	8.55	12.65	17.64	13.05	15.34	13.19	16.71	14.95
LC2-CoI	14.14	16.48	8.32	12.40	17.43	13.12	15.28	12.74	16.73	14.73
LC2-ACF	14.02	16.36	7.61	11.98	17.34	13.01	15.17	12.68	17.14	14.91
	Lee-Carter model applied to a single population (LC1)									
LC1-Ind	14.23	16.48	8.50	12.49	17.43	13.07	15.25	12.74	17.14	14.94

Table 4.3: $AMAPE_{c,\,g,\,[1994,\,2013]}$s (%)

AMAPE	Country	US			UK			Japan		
Model	Avg 6	M	F	Avg 2	M	F	Avg 2	M	F	Avg 2
	Hierarchical Credibility (HC) model									
EW-5	14.01	13.57	17.33	15.45	16.94	10.63	13.78	10.69	14.90	12.79
MW-5	14.02	13.48	18.26	15.87	16.44	10.60	13.52	10.85	14.50	12.67
EW-4	14.60	11.45	13.42	12.44	18.00	10.03	14.01	17.63	17.09	17.36
MW-4	14.28	11.63	15.38	13.50	17.03	9.53	13.28	16.12	15.98	16.05
EW-3	15.03	11.71	15.41	13.56	19.26	9.56	14.41	13.88	20.39	17.13
MW-3	14.55	11.83	16.37	14.10	17.98	9.58	13.78	12.62	18.94	15.78
	Lee-Carter model applied to all six populations (LC6)									
LC6-JoK	19.57	14.43	17.01	15.72	24.29	16.90	20.60	19.99	24.79	22.39
LC6-CoI	17.97	13.43	15.60	14.52	23.86	17.15	20.51	16.39	21.37	18.88
LC6-ACF	17.26	12.32	16.15	14.24	23.02	13.30	18.16	15.36	23.43	19.40
	Lee-Carter model applied to both genders of a country (LC2)									
LC2-JoK	18.40	13.42	17.11	15.27	22.63	16.59	19.61	17.99	22.61	20.30
LC2-CoI	17.92	13.43	15.60	14.52	22.53	16.62	19.58	17.05	22.26	19.65
LC2-ACF	18.41	14.32	16.49	15.41	22.43	16.89	19.66	16.90	23.46	20.18
	Lee-Carter model applied to a single population (LC1)									
LC1-Ind	18.25	13.43	16.26	14.85	22.53	16.78	19.66	17.05	23.44	20.24

Table 4.4: $AMAPE_{c,\,g,\,[1984,\,2013]}$s (%)

In summary, a hierarchical credibility model with more levels produces better prediction results, and the EW and MW strategies have similar forecasting performances. Regardless of the length of the fitting year span and forecasting year span, the hierarchical credibility models overall provide more accurate forecasts than the Lee-Carter models. Therefore, we conclude that the hierarchical credibility model is an effective approach to modelling multi-population mortality rates.

Chapter 5

Conclusions

This project applies the hierarchical credibility theory with tree structures of three, four and five levels to modelling multi-population mortality rates. The five-level tree structure from the bottom to the top is "year t", "age x", "gender g", "country c" and "multi-country", and the four-level tree structure is "year t", "age x", "gender g" and "country c". The hierarchical credibility mortality models are fitted with male and female mortality data of three developed countries (the US, the UK, and Japan) from the Human Mortality Database for an age span 25–84 and a series of fitting year spans. The classical Lee-Carter model and its three extensions for multiple populations (joint-k, cointegrated and augmented common factor) are also fitted with the same data set for comparisons.

The formula for $\hat{Y}_{c,\,g,\,x,\,T+1}$, the hierarchical credibility estimate of the decrement over $[T,\,T+1]$ in the logarithm of central death rate for country c, gender g and age x, under a five-level hierarchical structure for the special case ($w_{c,\,g,\,x,\,t} = 1$ for all $c = 1, \cdots, C$, $g = 1, \cdots, G$, $x = 1, \cdots, X$, and $t = 1, \cdots, T$) is a credibility-factor-weighted average of $\overline{Y}_{c,\,g,\,x,\,\bullet}$, $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}$, $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$, the average annual decrements of the time series $\{\ln(m_{c,\,g,\,x_L+x-1,\,t_L+t})\}$, $\{\ln(m_{c,\,g,\,\bullet,\,t_L+t})\}$, $\{\ln(m_{c,\,\bullet,\,\bullet,\,t_L+t})\}$ and $\{\ln(m_{\bullet,\,\bullet,\,\bullet,\,t_L+t})\}$ for four different levels. For the hierarchical mortality model, we also adopt the expanding window (EW) and moving window (MW) strategies proposed by Tsai and Lin (2017a, b) to forecast mortality rates for two or more years. Under the expanding window strategy, the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+\tau} = \hat{Y}_{c,\,g,\,x,\,T+1}$ for $\tau = 2, 3, \cdots$. Thus, $\ln(\hat{m}_{c,\,g,\,x_L+x-1,\,t_U+\tau})$ is a linear function of τ with slope $\hat{Y}_{c,\,g,\,x,\,T+1}$ and intercept $\ln(m_{c,\,g,\,x_L+x-1,\,t_U})$.

The forecasting performance is measured by the MAPE resulting from a single fitting year span. Based on the figures displayed in Chapter 4, we conclude that models have larger MAPEs as the forecasting year span gets wider. Since we do not know which fitting year span will produce the lowest MAPE, and actually the MAPE varies largely in the fitting year span, we rank the models according to AMAPE, the average of MAPEs over all the fitting year spans. From Tables 4.2–4.4, we conclude that the hierarchical credibility model with more levels overall yields better prediction, which means that modelling mortality rates for multiple countries together with a five-level hierarchical tree structure can produce overall more accurate prediction than modelling those for a single population separately with a three-level hierarchical tree structure. Moreover, regardless of the forecasting year span of 10-, 20- or 30-year width, the hierarchical credibility models overall provide higher accurate forecasting results than the Lee-Carter models.

The proposed model contributes to the literature of multi-population mortality modelling by incorporating the hierarchical credibility theory, which is widely used in property and casualty insurance, to model multi-population mortality rates. The model is convenient to implement, and can be applied to a hierarchical tree of any arbitrary level to fit a data set. It is a generalization of the credibility mortality model proposed by Tsai and Lin (2017b), which can be considered as having a three-level tree structure with population, year and age. The mortality rates predicted from the hierarchical credibility mortality model for multiple populations can be further used to construct a mortality index for more accurately pricing mortality-indexed securities.

Appendix A

Lee-Carter model and its three extensions for multiple populations

A.1 Independent Lee-Carter model

Lee and Carter (1992) introduced the famous Lee-Carter model to forecast mortality rates. It models the natural logarithm of the central death rate using a bilinear function of two age-specific factors and one time-varying factor. The model for a specific population i is presented as follows:

$$\ln(m_{i,\,x,\,t}) = \alpha_{i,\,x} + \beta_{i,\,x} \times k_{i,\,t} + \varepsilon_{i,\,x,\,t}, \quad x = x_L, \cdots, x_U, \quad t = t_L, \cdots, t_U,$$

where $\alpha_{i,\,x}$ is the average age-specific mortality factor at age x for population i, $k_{i,\,t}$ is the index of the mortality level in year t for population i, $\beta_{i,\,x}$ is the age-specific reaction to $k_{i,\,t}$ at age x for population i, and the model errors for population i, $\varepsilon_{i,\,x,\,t}$, $t = t_L, \cdots, t_U$, capturing the age-specific effects not reflected in the model, are assumed independent and identically distributed.

The independent Lee-Carter model is subject to two constraints, $\sum_{t=t_L}^{t_U} k_{i,\,t} = 0$ and $\sum_{x=x_L}^{x_U} \beta_{i,\,x} = 1$. The first constraint leads the estimate of $\alpha_{i,\,x}$ to

$$\hat{\alpha}_{i,\,x} = \frac{\sum_{t=t_L}^{t_U} \ln(m_{i,\,x,\,t})}{t_U - t_L + 1}, \quad x = x_L, \cdots, x_U,$$

and the second constraint gives the estimate of $\hat{k}_{i,\,t}$ as

$$\hat{k}_{i,\,t} = \sum_{x=x_L}^{x_U} [\ln(m_{i,\,x,\,t}) - \hat{\alpha}_{i,\,x}], \quad t = t_L, \cdots, t_U.$$

Finally, to get $\hat{\beta}_{i,x}$, we regress $[\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}]$ on $\hat{k}_{i,t}$ without the constant term for each age x.

The time-varying index $\hat{k}_{i,t}$ is the key to project future mortality rates. Lee and Carter (1992) suggested to model the time-varying index $\hat{k}_{i,t}$ by a random walk with drift θ_i, that is, $\hat{k}_{i,t} = \hat{k}_{i,t-1} + \theta_i + \epsilon_{i,t}$, where the time trend errors $\epsilon_{i,t}$, $t = t_L, \cdots, t_U$, are assumed to follow an independent and identically distributed, and independent of the model errors $\varepsilon_{i,x,t}$. The assumption above implies that $(\hat{k}_{i,t} - \hat{k}_{i,t-1})$ are independent and identically distributed for all t, and the parameter θ_i can be estimated by

$$\hat{\theta}_i = \frac{1}{t_U - t_L} \sum_{t=t_L+1}^{t_U} (\hat{k}_{i,t} - \hat{k}_{i,t-1}) = \frac{\hat{k}_{i,t_U} - \hat{k}_{i,t_L}}{t_U - t_L}.$$

The natural logarithm of the central death rate for age x in year $t_U + \tau$ and population i is forecasted by

$$\begin{aligned} \ln(\hat{m}_{i,x,t_U+\tau}) &= \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times \hat{k}_{i,t_U+\tau} = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times (\hat{k}_{i,t_U} + \tau \times \hat{\theta}_i) \\ &= \ln(\hat{m}_{i,x,t_U}) + (\hat{\beta}_{i,x} \times \hat{\theta}_i) \times \tau, \quad \tau = 1, 2, \cdots, \end{aligned}$$

a linear function of τ with intercept $\ln(\hat{m}_{i,x,t_U})$ and slope $(\hat{\beta}_{i,x} \times \hat{\theta}_i)$, where $\ln(\hat{m}_{i,x,t_U}) = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times \hat{k}_{i,t_U}$. Then, the projected central death rate is given by

$$\hat{m}_{i,x,t_U+\tau} = \exp[\hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times (\hat{k}_{i,t_U} + \tau \times \hat{\theta}_i)],$$

and the predicted one-year death rate for age x in year $t_U + \tau$ and population i is given as

$$\hat{q}_{i,x,t_U+\tau} = 1 - \exp[-\exp(\hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times (\hat{k}_{i,t_U} + \tau \times \hat{\theta}_i))].$$

A.2 Joint-k Lee-Carter model

Carter and Lee (1992) proposed the joint-k model to govern the co-movements among the mortality rates for multiple populations. The joint-k model is constructed in the same way as the independent Lee-Carter model except that the time-varying index $k_{i,t} = k_t$, $i = 1, \cdots, r$. The logarithm of central death rates, $\ln(m_{i,x,t})$, for lives aged x in year t and population i can be expressed as

$$\ln(m_{i,x,t}) = \alpha_{i,x} + \beta_{i,x} \times k_t + \varepsilon_{i,x,t}, \quad i = 1, \cdots, r, \quad x = x_L, \cdots, x_U, \quad t = t_L, \cdots, t_U,$$

where $\alpha_{i,x}$ is the average age-specific mortality factor at age x for population i, $\beta_{i,x}$ is the age-specific reaction to k_t at age x for population i, k_t is the common index of the mortality level in year t, and the model errors $\varepsilon_{i,x,t}$, $t = t_L, \cdots, t_U$, capturing the age-specific effects not reflected in the model, are assumed independent and identically distributed.

There are two constraints $\sum_{t=t_L}^{t_U} k_t = 0$ and $\sum_{i=1}^{r}\sum_{x=x_L}^{x_U} \beta_{i,x} = 1$ for the joint-k Lee-Carter model. The first constraint $\sum_{t=t_L}^{t_U} k_t = 0$ gives the estimate of α_x,

$$\hat{\alpha}_{i,x} = \frac{\sum_{t=t_L}^{t_U} \ln(m_{i,x,t})}{t_U - t_L + 1}, \quad x = x_L, \cdots, x_U,$$

and the second constraint $\sum_{i=1}^{r}\sum_{x=x_L}^{x_U} \beta_{i,x} = 1$ implies the estimate of k_t,

$$\hat{k}_t = \sum_{i=1}^{r}\sum_{x=x_L}^{x_U} [\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}], \quad t = t_L, \cdots, t_U.$$

Finally, we regress $[\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}]$ on $\hat{k}_t$ without the constant term for each age x to obtain $\hat{\beta}_{i,x}$.

The common time-varying index $\hat{k}_t$ is assumed to follow a random walk with drift θ for mortality prediction: $\hat{k}_t = \hat{k}_{t-1} + \theta + \epsilon_t$, where the time trend errors ϵ_t, $t = t_L + 1, \cdots, t_U$, are assumed independent and identically distributed, and independent of the model errors $\varepsilon_{i,x,t}$. Then we can estimate the drift parameter θ with

$$\hat{\theta} = \frac{1}{t_U - t_L} \sum_{t=t_L+1}^{t_U} (\hat{k}_t - \hat{k}_{t-1}) = \frac{\hat{k}_{t_U} - \hat{k}_{t_L}}{t_U - t_L}.$$

The logarithm of the projected central death rate for age x in year $t_U + \tau$ and population i is given by

$$\ln(\hat{m}_{i,x,t_U+\tau}) = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times (\hat{k}_{t_U} + \tau \times \hat{\theta}) = \ln(\hat{m}_{i,x,t_U}) + (\hat{\beta}_{i,x} \times \hat{\theta}) \times \tau, \quad \tau = 1, 2, \cdots,$$

a linear function of τ with intercept $\ln(\hat{m}_{i,x,t_U})$ and slope $(\hat{\beta}_{i,x} \cdot \hat{\theta})$, where $\ln(\hat{m}_{i,x,t_U}) = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times \hat{k}_{t_U}$.

A.3 Cointegrated Lee-Carter model

Unlike the joint-k model, which assumes that all populations have the common time-varying index k, the cointegrated model proposed by Li and Hardy (2011) assumes the time-varying index for population i ($i \geqslant 2$) is linearly related to the time-varying index for population 1, the base population. Therefore, the time-varying index for population i ($i \geqslant 2$) needs to be re-estimated in this model.

Assume that the mortality rate for lives aged x in year t and population i follows the independent Lee-Carter model as follows:

$$\ln(m_{i,x,t}) = \alpha_{i,x} + \beta_{i,x} \times k_{i,t} + \varepsilon_{i,x,t}, \quad i = 1, \cdots, r, \quad x = x_L, \cdots, x_U, \quad t = t_L, \cdots, t_U.$$

There are two constraints $\sum_{t=t_L}^{t_U} k_{i,t} = 0$ and $\sum_{x=x_L}^{x_U} \beta_{i,x} = 1$, $i = 1, \cdots, r$, for the cointegrated Lee-Carter model. The estimate of $\alpha_{i,x}$ can be obtained by the constraint

$\sum_{t=t_L}^{t_U} k_{i,t} = 0$ as

$$\hat{\alpha}_{i,x} = \frac{\sum_{t=t_L}^{t_U} \ln(m_{i,x,t})}{t_U - t_L + 1}, \quad x = x_L, \cdots, x_U,$$

and $k_{i,t}$ can be estimated with the remaining constraint $\sum_{x=x_L}^{x_U} \beta_{i,x} = 1$ as

$$\hat{k}_{i,t} = \sum_{x=x_L}^{x_U} [\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}], \quad x = x_L, \cdots, x_U.$$

Again, we regress $[\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}]$ on $\hat{k}_{i,t}$ without the constant term for each age x to get $\hat{\beta}_{i,x}$.

The time trend $\hat{k}_{i,t}$ is assumed to follow a random walk with drift θ_i for mortality prediction: $\hat{k}_{i,t} = \hat{k}_{i,t-1} + \theta_i + \epsilon_{i,t}$, where the time trend errors $\epsilon_{i,t}$, $t = t_L + 1, \cdots, t_U$, are assumed independent and identically distributed, and independent of the model errors $\varepsilon_{i,x,t}$. Then the drift parameter θ_i for population i can be estimated by $\hat{\theta}_i = (\hat{k}_{i,t_U} - \hat{k}_{i,t_L})/(t_U - t_L)$, $i = 1, \cdots, r$.

The cointegrated Lee-Carter model assumes there is a linear relationship plus an error term $e_{i,t}$ between $\hat{k}_{1,t}$ and $\hat{k}_{i,t}$ for $i = 2, \cdots, r$. Specifically, $\hat{k}_{i,t} = a_i + b_i \times \hat{k}_{1,t} + e_{i,t}$, $i = 2, \cdots, r$. Then we re-estimate $k_{i,t}$ using the simple linear regression as $\hat{\hat{k}}_{i,t} = \hat{a}_i + \hat{b}_i \times \hat{k}_{1,t}$ for $i = 2, \cdots, r$, implying that the estimate of the drift of the time-varying index for population i, $\hat{\hat{\theta}}_i$, is given by

$$\hat{\hat{\theta}}_i = \begin{cases} \dfrac{1}{t_U - t_L} \displaystyle\sum_{t=t_L+1}^{t_U} (\hat{k}_{1,t} - \hat{k}_{1,t-1}) = \dfrac{\hat{k}_{1,t_U} - \hat{k}_{1,t_L}}{t_U - t_L} = \hat{\theta}_1, \quad i = 1, \\ \dfrac{\hat{\hat{k}}_{i,t_U} - \hat{\hat{k}}_{i,t_L}}{t_U - t_L} = \hat{b}_i \times \dfrac{\hat{k}_{1,t_U} - \hat{k}_{1,t_L}}{t_U - t_L} = \hat{b}_i \times \hat{\theta}_1, \quad i = 2, \cdots, r. \end{cases}$$

The logarithm of the forecasted central death rates for lives aged x in year $t_U + \tau$ and population i is given as

$$\ln(\hat{m}_{i,x,t_U+\tau}) = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times (\hat{k}_{i,t_U} + \tau \times \hat{\hat{\theta}}_i) = \ln(\hat{m}_{i,x,t_U}) + (\hat{\beta}_{i,x} \times \hat{\hat{\theta}}_i) \times \tau, \quad \tau = 1, 2, \cdots,$$

a linear function of τ with intercept $\ln(\hat{m}_{i,x,t_U})$ and slope $(\hat{\beta}_{i,x} \cdot \hat{\hat{\theta}}_i)$, where $\ln(\hat{m}_{i,x,t_U}) = \hat{\alpha}_{i,x} + \hat{\beta}_{i,x} \times \hat{k}_{i,t_U}$.

A.4 Augmented common factor Lee-Carter model

To deal with the divergence in forecasting multi-population mortality rates over the long-term, Li and Lee (2005) proposed the augmented common factor model which not only considers the commonalities in the historical experience but also includes the individual differences in the trends.

First, the independent Lee-Carter model is modified to a common factor model by setting a common age-specific term β_x ($\beta_{i,x} = \beta_x$) and a uniform time-varying index k_t ($k_{i,t} = k_t$) for all populations as follows:

$$\ln(m_{i,x,t}) = \alpha_{i,x} + \beta_x \times k_t + \varepsilon_{i,x,t}, \quad i = 1, \cdots, r, \quad x = x_L, \cdots, x_U, \quad t = t_L, \cdots, t_U,$$

subject to two constraints, $\sum_{t=t_L}^{t_U} k_t = 0$ and $\sum_{i=1}^{r} \sum_{x=x_L}^{x_U} w_i \, \beta_x = 1$, where w_i, set to be $1/r$ in this paper, is the weight for population i and $\sum_{i=1}^{r} w_i = 1$. We can similarly estimate $\alpha_{i,x}$ by

$$\hat{\alpha}_{i,x} = \frac{\sum_{t=t_L}^{t_U} \ln(m_{i,x,t})}{t_U - t_L + 1}, \quad x = x_L, \cdots, x_U,$$

and k_t as

$$\hat{k}_t = \sum_{i=1}^{r} \sum_{x=x_L}^{x_U} w_i \times [\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}], \quad t = t_L, \cdots, t_U.$$

Then $\hat{\beta}_x$ can be similarly obtained by regressing $\sum_{i=1}^{r} w_i \times [\ln(m_{i,x,t}) - \hat{\alpha}_{i,x}]$ on $\hat{k}_t$ without the constant term for each age x.

To include the individual differences in the trends, Li and Lee (2005) added a factor $\beta'_{i,x} \times k'_{i,t}$ to the common factor model to get

$$\ln(m_{i,x,t}) = \alpha_{i,x} + \beta_x \times k_t + \beta'_{i,x} \times k'_{i,t} + \varepsilon_{i,x,t},$$

with an extra constraint $\sum_{x=x_L}^{x_U} \beta'_{i,x} = 1$, which is called the augmented common factor model. The extra constraint implies $\hat{k}'_{i,t} = \sum_{x=x_L}^{x_U} [\ln(m_{i,x,t}) - \hat{\alpha}_{i,x} - \hat{\beta}_x \times \hat{k}_t]$, and $\hat{\beta}'_{i,x}$ can be obtained by regressing $[\ln(m_{i,x,t}) - \hat{\alpha}_{i,x} - \hat{\beta}_x \times \hat{k}_t]$ on $\hat{k}'_{i,t}$ without the constant term for each age x.

Similarly, we assume that both time trends $\hat{k}_t$ and $\hat{k}'_{i,t}$ follow a random walk with drifts θ and θ'_i, respectively. Specifically, $\hat{k}_t = \hat{k}_{t-1} + \theta + \epsilon_t$, and $\hat{k}'_{i,t} = \hat{k}'_{i,t-1} + \theta'_i + \epsilon_{i,t}$, where each of the time trend errors ϵ_t and $\epsilon_{i,t}$, $t = t_L + 1, \cdots, t_U$, are assumed independent and identically distributed, and all of the three error terms, $\varepsilon_{i,x,t}$, ϵ_t and $\epsilon_{i,t}$, are assumed to be independent. Again, the drift parameters θ and θ'_i can be similarly estimated by $\hat{\theta} = (\hat{k}_{t_U} - \hat{k}_{t_L})/(t_U - t_L)$ and $\hat{\theta}'_i = (\hat{k}'_{i,t_U} - \hat{k}'_{i,t_L})/(t_U - t_L)$.

Finally, the logarithm of the predicted central death rates for lives aged x in year $t_U + \tau$ and population i can be expressed as

$$\begin{aligned} \ln(\hat{m}_{i,x,t_U+\tau}) &= \hat{\alpha}_{i,x} + \hat{\beta}_x \times (\hat{k}_{t_U} + \tau \times \hat{\theta}) + \hat{\beta}'_{i,x} \times (\hat{k}'_{i,t_U} + \tau \times \hat{\theta}'_i) \\ &= \ln(\hat{m}_{i,x,t_U}) + (\hat{\beta}_x \times \hat{\theta} + \hat{\beta}'_{i,x} \times \hat{\theta}'_i) \times \tau, \quad \tau = 1, 2, \cdots, \end{aligned}$$

a linear function of τ with intercept $\ln(\hat{m}_{i,x,t_U})$ and slope $(\hat{\beta}_x \cdot \hat{\theta} + \hat{\beta}'_{i,x} \cdot \hat{\theta}'_i)$, where $\ln(\hat{m}_{i,x,t_U}) = \hat{\alpha}_{i,x} + \hat{\beta}_x \times \hat{k}_{t_U} + \hat{\beta}'_{i,x} \times \hat{k}'_{i,t_U}$.

Appendix B

Estimation of structural parameters and credibility factors for the hierarchical credibility model

Credibility estimators

From Section 6.4 of Bühlmann and Gisler (2005), to find the hierarchical credibility estimator $\hat{Y}_{c,g,x,T+1}$ for age x and gender g and country c in year $T+1$, we need to first find the credibility estimators $\hat{Y}_{c,g}$ and $\hat{Y}_{c,g,x}$. All three credibility estimators $\hat{Y}_{c,g}$, $\hat{Y}_{c,g,x}$ and $\hat{Y}_{c,g,x,T+1}$ can be seen as weighted averages of Bs ($B_c^{(3)}$, $B_{c,g}^{(2)}$ and $B_{c,g,x}^{(1)}$) and the overall mean $\hat{\mu}_4$ as follows:

$$\hat{Y}_{c,g} = \hat{\alpha}_c^{(3)} \cdot B_c^{(3)} + (1 - \hat{\alpha}_c^{(3)}) \cdot \hat{\mu}_4, \tag{A.1}$$

$$\begin{aligned} \hat{Y}_{c,g,x} &= \hat{\alpha}_{c,g}^{(2)} \cdot B_{c,g}^{(2)} + (1 - \hat{\alpha}_{c,g}^{(2)}) \cdot \hat{Y}_{c,g} \qquad \text{(A.2)} \\ &= \hat{\alpha}_{c,g}^{(2)} \cdot B_{c,g}^{(2)} + [(1 - \hat{\alpha}_{c,g}^{(2)}) \cdot \hat{\alpha}_c^{(3)}] \cdot B_c^{(3)} + [(1 - \hat{\alpha}_{c,g}^{(2)}) \cdot (1 - \hat{\alpha}_c^{(3)})] \cdot \hat{\mu}_4, \end{aligned}$$

and

$$
\begin{aligned}
& \hat{Y}_{c,\,g,\,x,\,T+1} \\
= \;& \hat{\alpha}^{(1)}_{c,\,g,\,x} \cdot B^{(1)}_{c,\,g,\,x} + (1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot \hat{Y}_{c,\,g,\,x} \qquad \text{(A.3)} \\
= \;& \hat{\alpha}^{(1)}_{c,\,g,\,x} \cdot B^{(1)}_{c,\,g,\,x} + [(1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot \hat{\alpha}^{(2)}_{c,\,g}] \cdot B^{(2)}_{c,\,g} + [(1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot (1-\hat{\alpha}^{(2)}_{c,\,g})] \cdot \hat{Y}_{c,\,g} \\
= \;& \hat{\alpha}^{(1)}_{c,\,g,\,x} \cdot B^{(1)}_{c,\,g,\,x} + [(1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot \hat{\alpha}^{(2)}_{c,\,g}] \cdot B^{(2)}_{c,\,g} + [(1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot (1-\hat{\alpha}^{(2)}_{c,\,g}) \cdot \hat{\alpha}^{(3)}_{c}] \cdot B^{(3)}_{c} \\
& + [(1-\hat{\alpha}^{(1)}_{c,\,g,\,x}) \cdot (1-\hat{\alpha}^{(2)}_{c,\,g}) \cdot (1-\hat{\alpha}^{(3)}_{c})] \cdot \hat{\mu}_4,
\end{aligned}
$$

where $\hat{\alpha}^{(1)}_{c,\,g,\,x}$, $\hat{\alpha}^{(2)}_{c,\,g}$ and $\hat{\alpha}^{(3)}_{c}$ are the corresponding credibility factors for levels one, two and three, respectively. The expressions for $B^{(1)}_{c,\,g,\,x}$, $B^{(2)}_{c,\,g}$ and $B^{(3)}_{c}$, the credibility factors ($\hat{\alpha}^{(1)}_{c,\,g,\,x}$, $\hat{\alpha}^{(2)}_{c,\,g}$ and $\hat{\alpha}^{(3)}_{c}$), and $\hat{\mu}_4$ by theorem 6.4 of Bühlmann and Gisler (2005) are given in Table B.1.

$B^{(1)}_{c,\,g,\,x} = \sum_{t=1}^{T} \frac{w_{c,\,g,\,x,\,t}}{w^{(1)}_{c,\,g,\,x}} \cdot Y_{c,\,g,\,x,\,t}$	$w^{(1)}_{c,\,g,\,x} = \sum_{t=1}^{T} w_{c,\,g,\,x,\,t}$	$\hat{\alpha}^{(1)}_{c,\,g,\,x} = \frac{w^{(1)}_{c,\,g,\,x} \cdot \hat{\sigma}_1^2}{w^{(1)}_{c,\,g,\,x} \cdot \hat{\sigma}_1^2 + \hat{\sigma}_0^2}$
$B^{(2)}_{c,\,g} = \sum_{x=1}^{X} \frac{\hat{\alpha}^{(1)}_{c,\,g,\,x}}{w^{(2)}_{c,\,g}} \cdot B^{(1)}_{c,\,g,\,x}$	$w^{(2)}_{c,\,g} = \sum_{x=1}^{X} \hat{\alpha}^{(1)}_{c,\,g,\,x}$	$\hat{\alpha}^{(2)}_{c,\,g} = \frac{w^{(2)}_{c,\,g} \cdot \hat{\sigma}_2^2}{w^{(2)}_{c,\,g} \cdot \hat{\sigma}_2^2 + \hat{\sigma}_1^2}$
$B^{(3)}_{c} = \sum_{g=1}^{G} \frac{\hat{\alpha}^{(2)}_{c,\,g}}{w^{(3)}_{c}} \cdot B^{(2)}_{c,\,g}$	$w^{(3)}_{c} = \sum_{g=1}^{G} \hat{\alpha}^{(2)}_{c,\,g}$	$\hat{\alpha}^{(3)}_{c} = \frac{w^{(3)}_{c} \cdot \hat{\sigma}_3^2}{w^{(3)}_{c} \cdot \hat{\sigma}_3^2 + \hat{\sigma}_2^2}$
$\hat{\mu}_3 = \sum_{c=1}^{C} \frac{\hat{\alpha}^{(3)}_{c}}{w^{(4)}} \cdot B^{(3)}_{c}$	$w^{(4)} = \sum_{c=1}^{C} \hat{\alpha}^{(3)}_{c}$	

Table B.1: Formulas for $B^{(1)}_{c,\,g,\,x}$, $B^{(2)}_{c,\,g}$, $B^{(3)}_{c}$, $\hat{\alpha}^{(1)}_{c,\,g,\,x}$, $\hat{\alpha}^{(2)}_{c,\,g}$, $\hat{\alpha}^{(3)}_{c}$ and $\hat{\mu}_4$

Estimation of the structural parameters

We observe from Table B.1 that

- $B^{(1)}_{c,\,g,\,x} = \sum_{t=1}^{T} \frac{w_{c,\,g,\,x,\,t}}{w^{(1)}_{c,\,g,\,x}} \cdot Y_{c,\,g,\,x,\,t}$ for level 1 is an exposure-unit-weighted average of $Y_{c,\,g,\,x,\,t}$s over all years t at level 0 under age x, gender g and country c with the weight $[w_{c,\,g,\,x,\,t}/w^{(1)}_{c,\,g,\,x}]$, where $w^{(1)}_{c,\,g,\,x}$ is the sum of the level-zero exposure units $w_{c,\,g,\,x,\,t}$ for $t = 1, \cdots, T$.
- $B^{(2)}_{c,\,g} = \sum_{x=1}^{X} \frac{\hat{\alpha}^{(1)}_{c,\,g,\,x}}{w^{(2)}_{c,\,g}} \cdot B^{(1)}_{c,\,g,\,x}$ for level 2 is a credibility-factor-weighted average of $B^{(1)}_{c,\,g,\,x}$s at level 1 with weight $[\hat{\alpha}^{(1)}_{c,\,g,\,x}/w^{(2)}_{c,\,g}]$, where $w^{(2)}_{c,\,g}$ is the sum of the level-one credibility factors $\hat{\alpha}^{(1)}_{c,\,g,\,x}$ for $x = 1, \cdots, X$; since $B^{(2)}_{c,\,g} = \sum_{x=1}^{X}\sum_{t=1}^{T} \frac{\hat{\alpha}^{(1)}_{c,\,g,\,x}}{w^{(2)}_{c,\,g}} \cdot \frac{w_{c,\,g,\,x,\,t}}{w^{(1)}_{c,\,g,\,x}} \cdot Y_{c,\,g,\,x,\,t}$, it is also a weighted average of $Y_{c,\,g,\,x,\,t}$s for all ages x and years t at levels 0–1 under gender g and country c.

- $B_c^{(3)} = \sum_{g=1}^{G} \frac{\hat{\alpha}_{c,g}^{(2)}}{w_c^{(3)}} \cdot B_{c,g}^{(2)}$ for level 3 is a credibility-factor-weighted average of $B_{c,g}^{(2)}$s at level 2 with weight $[\hat{\alpha}_{c,g}^{(2)}/w_c^{(3)}]$, where $w_c^{(3)}$ is the sum of the level-two credibility factors $\hat{\alpha}_{c,g}^{(2)}$ for $g = 1, \cdots, G$; because $B_c^{(3)} = \sum_{g=1}^{G}\sum_{x=1}^{X}\sum_{t=1}^{T} \frac{\hat{\alpha}_{c,g}^{(2)}}{w_c^{(3)}} \cdot \frac{\hat{\alpha}_{c,g,x}^{(1)}}{w_{c,g}^{(2)}} \cdot \frac{w_{c,g,x,t}}{w_{c,g,x}^{(1)}} \cdot Y_{c,g,x,t}$, it is also a weighted average of $Y_{c,g,x,t}$s for all genders g, ages x and years t at levels 0–2 under country c.
- $\hat{\mu}_4 = \sum_{c=1}^{C} \frac{\hat{\alpha}_c^{(3)}}{w^{(4)}} \cdot B_c^{(3)}$ for level 4 is a credibility-factor-weighted average of $B_c^{(3)}$s at level 3 with weight $[\hat{\alpha}_c^{(3)}/w^{(4)}]$, where $w^{(4)}$ is the sum of the level-three credibility factors $\hat{\alpha}_c^{(3)}$ for $c = 1, \cdots, C$; furthermore, $\hat{\mu}_4 = \sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\sum_{t=1}^{T} \frac{\hat{\alpha}_c^{(3)}}{w^{(4)}} \cdot \frac{\hat{\alpha}_{c,g}^{(2)}}{w_c^{(3)}} \cdot \frac{\hat{\alpha}_{c,g,x}^{(1)}}{w_{c,g}^{(2)}} \cdot \frac{w_{c,g,x,t}}{w_{c,g,x}^{(1)}} \cdot Y_{c,g,x,t}$ is also a weighted average of $Y_{c,g,x,t}$s for all countries c, genders g, ages x and years t at levels 0–3.

$\sigma_0^2 = E[\sigma_1^2(\Theta_{c,g,x})]$	$\sigma_1^2(\Theta_{c,g,x}) = w_{c,g,x,t} \cdot Var[Y_{c,g,x,t} \vert \Theta_{c,g,x}]$
$\hat{\sigma}_0^2 = \frac{1}{C \cdot G \cdot X} \sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X} \hat{\sigma}_1^2(\Theta_{c,g,x})$	$\hat{\sigma}_1^2(\Theta_{c,g,x}) \triangleq \frac{1}{T-1} \sum_{t=1}^{T} w_{c,g,x,t} \cdot [Y_{c,g,x,t} - B_{c,g,x}^{(1)}]^2$
$\sigma_1^2 = E\{Var[\mu_1(\Theta_{c,g,x}) \vert \Phi_{c,g}]\}$	$z_{c,g}^{(1)} = \sum_{x=1}^{X} w_{c,g,x}^{(1)}, \quad \overline{B}_{c,g,\bullet}^{(1)} = \sum_{x=1}^{X} \frac{w_{c,g,x}^{(1)}}{z_{c,g}^{(1)}} \cdot B_{c,g,x}^{(1)}$ $c_{c,g}^{(1)} = \frac{X-1}{X} \cdot \left\{ \sum_{x=1}^{X} \frac{w_{c,g,x}^{(1)}}{z_{c,g}^{(1)}} \cdot \left[1 - \frac{w_{c,g,x}^{(1)}}{z_{c,g}^{(1)}} \right] \right\}^{-1}$
$\hat{\sigma}_1^2 = \frac{1}{C \cdot G} \sum_{c=1}^{C}\sum_{g=1}^{G} \max[\hat{T}_{c,g}^{(1)}, 0]$	$\hat{T}_{c,g}^{(1)} = c_{c,g}^{(1)} \cdot \left\{ \frac{X}{X-1} \sum_{x=1}^{X} \frac{w_{c,g,x}^{(1)}}{z_{c,g}^{(1)}} \left[B_{c,g,x}^{(1)} - \overline{B}_{c,g,\bullet}^{(1)} \right]^2 - \frac{X}{z_{c,g}^{(1)}} \cdot \hat{\sigma}_0^2 \right\}$
$\sigma_2^2 = E\{Var[\mu_2(\Phi_{c,g}) \vert \Psi_c]\}$	$z_c^{(2)} = \sum_{g=1}^{G} w_{c,g}^{(2)}, \quad \overline{B}_{c,\bullet}^{(2)} = \sum_{g=1}^{G} \frac{w_{c,g}^{(2)}}{z_c^{(2)}} \cdot B_{c,g}^{(2)}$ $c_c^{(2)} = \frac{G-1}{G} \cdot \left\{ \sum_{g=1}^{G} \frac{w_{c,g}^{(2)}}{z_c^{(2)}} \cdot \left[1 - \frac{w_{c,g}^{(2)}}{z_c^{(2)}} \right] \right\}^{-1}$
$\hat{\sigma}_2^2 = \frac{1}{C} \sum_{c=1}^{C} \max[\hat{T}_c^{(2)}, 0]$	$\hat{T}_c^{(2)} = c_c^{(2)} \cdot \left\{ \frac{G}{G-1} \sum_{g=1}^{G} \frac{w_{c,g}^{(2)}}{z_c^{(2)}} \left[B_{c,g}^{(2)} - \overline{B}_{c,\bullet}^{(2)} \right]^2 - \frac{G}{z_c^{(2)}} \cdot \hat{\sigma}_1^2 \right\}$
$\sigma_3^2 = Var[\mu_3(\Psi_c)]$	$z^{(3)} = \sum_{c=1}^{C} w_c^{(3)}, \quad \overline{B}_{\bullet}^{(3)} = \sum_{c=1}^{C} \frac{w_c^{(3)}}{z^{(3)}} \cdot B_c^{(3)}$ $c^{(3)} = \frac{C-1}{C} \cdot \left\{ \sum_{c=1}^{C} \frac{w_c^{(3)}}{z^{(3)}} \cdot \left[1 - \frac{w_c^{(3)}}{z^{(3)}} \right] \right\}^{-1}$
$\hat{\sigma}_3^2 = \max[\hat{T}^{(3)}, 0]$	$\hat{T}^{(3)} = c^{(3)} \cdot \left\{ \frac{C}{C-1} \sum_{c=1}^{C} \frac{w_c^{(3)}}{z^{(3)}} \left[B_c^{(3)} - \overline{B}_{\bullet}^{(3)} \right]^2 - \frac{C}{z^{(3)}} \cdot \hat{\sigma}_2^2 \right\}$

Table B.2: Hierarchical credibility estimation of σ_0^2, σ_1^2, σ_2^2 and σ_3^2

It is possible that $\hat{T}^{(1)}_{c,g}$, $\hat{T}^{(2)}_{c}$ or $\hat{T}^{(3)}$ given in Table B.2 is negative because of a subtraction. In this case, we set it to zero. Note that for a four-level tree structure, we can set $\hat{\alpha}^{(3)}_{c} = 1$ and do not need to calculate $\hat{\mu}_4$ and $w^{(4)}$ in Table B.1. Similarly, to obtain the corresponding formula (A.3) for a three-level tree structure which is the Bühlmann-Straub credibility model, we let $\hat{\alpha}^{(3)}_{c} = \hat{\alpha}^{(2)}_{c,g} = 1$ and it is not necessary to calculate $B^{(3)}_{c}$, $w^{(3)}_{c}$, $\hat{\mu}_4$ and $w^{(4)}$.

To get the credibility estimate $\hat{Y}_{c,g,x,T+1}$ in (A.3), we need the estimates $\hat{\sigma}^2_0$, $\hat{\sigma}^2_1$, $\hat{\sigma}^2_2$ and $\hat{\sigma}^2_3$. The estimation of σ^2_0, σ^2_1, σ^2_2 and σ^2_3 is given in Table B.2. For the detailed estimation, please refer to Section 6.6 of Bühlmann and Gisler (2005). From the formulas in Table B.2, we notice that the values of $\hat{\sigma}^2_1$, $\hat{\sigma}^2_2$ and $\hat{\sigma}^2_3$ can be zero, which leads to the values of $\hat{\alpha}^{(1)}_{c,g,x}$, $\hat{\alpha}^{(2)}_{c,g}$ and $\hat{\alpha}^{(3)}_{c}$ being zero, respectively. Since the structural parameters are estimated from the bottom to the top of the tree structure, it is easy to extend them to a tree structure with higher levels. However, a hierarchical tree structure with higher levels has more structural parameters. As suggested by Bühlmann and Gisler (2005), one should be careful choosing the number of levels in the hierarchical credibility model.

It is obvious that the tree structure of the hierarchical credibility model covers that of the Bühlmann-Straub credibility model, which can be obtained by applying a three-level tree structure with only levels zero, one and two in Figure 3.1 for a specific population of country c and gender g. Therefore, the five-level hierarchical credibility model is a generalization of the Bühlmann-Straub credibility model. Given a specific population of country c and gender g for a three-level hierarchical tree structure, the Bühlmann-Straub credibility model reduces to the Bühlmann one if we further set $w_{c,g,x,t} = 1$ for $x = 1, \cdots, X$ and $t = 1, \cdots, T$.

Next, we give a special case where all of the exposure units $w_{c,g,x,t}$s are set to 1 (equal exposure units). We will use this special case for our hierarchical credibility mortality model.

A special case

If $w_{c,g,x,t} = 1$ for all $c = 1, \cdots, C$, $g = 1, \cdots, G$, $x = 1, \cdots, X$, and $t = 1, \cdots, T$, then the quantities in Tables B.1 and B.2 simplify to those in Tables B.3 and B.4. Moreover, the credibility factors $\hat{\alpha}^{(1)}_{c,g,x}$, $\hat{\alpha}^{(2)}_{c,g}$ and $\hat{\alpha}^{(3)}_{c}$ become subscript-free ones $\hat{\alpha}^{(1)}$, $\hat{\alpha}^{(2)}$ and $\hat{\alpha}^{(3)}$, and $B^{(1)}_{c,g,x}$, $B^{(2)}_{c,g}$, $B^{(3)}_{c}$, and $\hat{\mu}_4$ simplify to $\overline{Y}_{c,g,x,\bullet}$, $\overline{Y}_{c,g,\bullet,\bullet}$, $\overline{Y}_{c,\bullet,\bullet,\bullet}$ and $\overline{Y}_{\bullet,\bullet,\bullet,\bullet}$, respectively. Note that

Table B.3: Formulas for $B^{(1)}_{c,g,x}$, $B^{(2)}_{c,g}$, $B^{(3)}_{c}$, $\hat{\alpha}^{(1)}_{c,g,x}$, $\hat{\alpha}^{(2)}_{c,g}$, $\hat{\alpha}^{(3)}_{c}$ and $\hat{\mu}_4$ under all $w_{c,g,x,t} = 1$

$B^{(1)}_{c,g,x} = \frac{1}{T}\sum_{t=1}^{T} Y_{c,g,x,t} \triangleq \overline{Y}_{c,g,x,\bullet}$	$w^{(1)}_{c,g,x} = T \triangleq w^{(1)}$	$\hat{\alpha}^{(1)}_{c,g,x} = \frac{T \cdot \hat{\sigma}^2_1}{T \cdot \hat{\sigma}^2_1 + \hat{\sigma}^2_0} \triangleq \hat{\alpha}^{(1)}$
$B^{(2)}_{c,g} = \frac{1}{X}\sum_{x=1}^{X} B^{(1)}_{c,g,x} \triangleq \overline{Y}_{c,g,\bullet,\bullet}$	$w^{(2)}_{c,g} = X \cdot \hat{\alpha}^{(1)}$	$\hat{\alpha}^{(2)}_{c,g} = \frac{X \cdot \hat{\alpha}^{(1)} \cdot \hat{\sigma}^2_2}{X \cdot \hat{\alpha}^{(1)} \cdot \hat{\sigma}^2_2 + \hat{\sigma}^2_1} \triangleq \hat{\alpha}^{(2)}$
$B^{(3)}_{c} = \frac{1}{G}\sum_{g=1}^{G} B^{(2)}_{c,g} \triangleq \overline{Y}_{c,\bullet,\bullet,\bullet}$	$w^{(3)}_{c} = G \cdot \hat{\alpha}^{(2)}$	$\hat{\alpha}^{(3)}_{c} = \frac{G \cdot \hat{\alpha}^{(2)} \cdot \hat{\sigma}^2_3}{G \cdot \hat{\alpha}^{(2)} \cdot \hat{\sigma}^2_3 + \hat{\sigma}^2_2} \triangleq \hat{\alpha}^{(3)}$
$\hat{\mu}_3 = \frac{1}{C}\sum_{c=1}^{C} B^{(3)}_{c} \triangleq \overline{Y}_{\bullet,\bullet,\bullet,\bullet}$	$w^{(4)} = C \cdot \hat{\alpha}^{(3)}$	

- $\overline{Y}_{c,g,x,\bullet}$ (called the age sample mean) for level 1 is the average of $Y_{c,g,x,t}$s at level 0 for $t = 1, \cdots, T$, or the average of all entries $Y_{c,g,x,t}$s ($t = 1, \cdots, T$) in a $1 \times T$ row vector, for age x under gender g and country c;
- $\overline{Y}_{c,g,\bullet,\bullet}$ (called the gender sample mean) for level 2 is the average of $\overline{Y}_{c,g,x,\bullet}$ at level 1 for $x = 1, \cdots, X$, or the average of all entries $Y_{c,g,x,t}$s ($x = 1, \cdots, X$ and $t = 1, \cdots, T$) in an $(X \times T)$ matrix or rectangle, for gender g under country c,
- $\overline{Y}_{c,\bullet,\bullet,\bullet}$ (called the country sample mean) for level 3 is the average of $\overline{Y}_{c,g,\bullet,\bullet}$ at level 2 for $g = 1, \cdots, G$, or the average of all entries $Y_{c,g,x,t}$s ($g = 1, \cdots, G$, $x = 1, \cdots, X$, and $t = 1, \cdots, T$) in a $(G \times X \times T)$ matrix or solid, for country c; and
- $\overline{Y}_{\bullet,\bullet,\bullet,\bullet}$ (called the multi-country sample mean) for level 4 is the average of $\overline{Y}_{c,\bullet,\bullet,\bullet}$ at level 3 for $c = 1, \cdots, C$, or the average of all entries $Y_{c,g,x,t}$s ($c = 1, \cdots, C$, $g = 1, \cdots, G$, $x = 1, \cdots, X$, and $t = 1, \cdots, T$) in a $C \times G \times X \times T$ matrix or in C $(G \times X \times T)$ solids for the multi-country population.

Under this special case that all $w_{c,g,x,t} = 1$, (A.1)–(A.3) turn out to

$$\hat{Y}_{c,g} = \hat{\alpha}^{(3)} \cdot \overline{Y}_{c,\bullet,\bullet,\bullet} + (1 - \hat{\alpha}^{(3)}) \cdot \overline{Y}_{\bullet,\bullet,\bullet,\bullet}, \tag{A.4}$$

$$\begin{aligned} \hat{Y}_{c,g,x} &= \hat{\alpha}^{(2)} \cdot \overline{Y}_{c,g,\bullet,\bullet} + (1 - \hat{\alpha}^{(2)}) \cdot \hat{Y}_{c,g} \\ &= \hat{\alpha}^{(2)} \cdot \overline{Y}_{c,g,\bullet,\bullet} + [(1 - \hat{\alpha}^{(2)}) \cdot \hat{\alpha}^{(3)}] \cdot \overline{Y}_{c,\bullet,\bullet,\bullet} + [(1 - \hat{\alpha}^{(2)}) \cdot (1 - \hat{\alpha}^{(3)})] \cdot \overline{Y}_{\bullet,\bullet,\bullet,\bullet}, \end{aligned} \tag{A.5}$$

and

$$\begin{aligned} \hat{Y}_{c,g,x,T+1} &= \hat{\alpha}^{(1)} \cdot \overline{Y}_{c,g,x,\bullet} + (1 - \hat{\alpha}^{(1)}) \cdot \hat{Y}_{c,g,x} \\ &= \hat{\alpha}^{(1)} \cdot \overline{Y}_{c,g,x,\bullet} + [(1 - \hat{\alpha}^{(1)}) \cdot \hat{\alpha}^{(2)}] \cdot \overline{Y}_{c,g,\bullet,\bullet} \\ &\quad + [(1 - \hat{\alpha}^{(1)}) \cdot (1 - \hat{\alpha}^{(2)}) \cdot \hat{\alpha}^{(3)}] \cdot \overline{Y}_{c,\bullet,\bullet,\bullet} \\ &\quad + [(1 - \hat{\alpha}^{(1)}) \cdot (1 - \hat{\alpha}^{(2)}) \cdot (1 - \hat{\alpha}^{(3)})] \cdot \overline{Y}_{\bullet,\bullet,\bullet,\bullet}. \end{aligned} \tag{A.6}$$

$\sigma_0^2 = E[\sigma_1^2(\Theta_{c,\,g,\,x})]$	$\sigma_1^2(\Theta_{c,\,g,\,x}) = Var[Y_{c,\,g,\,x,\,t} \vert \Theta_{c,\,g,\,x}]$
$\hat{\sigma}_0^2 = \dfrac{1}{C \cdot G \cdot X} \sum_{c=1}^{C} \sum_{g=1}^{G} \sum_{x=1}^{X} \hat{\sigma}_1^2(\Theta_{c,\,g,\,x})$	$\hat{\sigma}_1^2(\Theta_{c,\,g,\,x}) \triangleq \dfrac{1}{T-1} \sum_{t=1}^{T} \left[Y_{c,\,g,\,x,\,t} - \overline{Y}_{c,\,g,\,x,\,\bullet} \right]^2$
$\sigma_1^2 = E\{Var[\mu_1(\Theta_{c,\,g,\,x}) \vert \Phi_{c,\,g}]\}$	$z_{c,\,g}^{(1)} = X \cdot T, \quad \overline{B}_{c,\,g,\,\bullet}^{(1)} = \overline{Y}_{c,\,g,\,\bullet,\,\bullet}, \quad c_{c,\,g}^{(1)} = 1$
$\hat{\sigma}_1^2 = \dfrac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \max[\hat{T}_{c,\,g}^{(1)},\, 0]$	$\hat{T}_{c,\,g}^{(1)} = \dfrac{1}{X-1} \sum_{x=1}^{X} \left[\overline{Y}_{c,\,g,\,x,\,\bullet} - \overline{Y}_{c,\,g,\,\bullet,\,\bullet} \right]^2 - \dfrac{\hat{\sigma}_0^2}{T}$
$\sigma_2^2 = E\{Var[\mu_2(\Phi_{c,\,g}) \vert \Psi_c]\}$	$z_c^{(2)} = G \cdot X \cdot \hat{\alpha}^{(1)}, \quad \overline{B}_{c,\,\bullet}^{(2)} = \overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}, \quad c_c^{(2)} = 1$
$\hat{\sigma}_2^2 = \dfrac{1}{C} \sum_{c=1}^{C} \max[\hat{T}_c^{(2)},\, 0]$	$\hat{T}_c^{(2)} = \dfrac{1}{G-1} \sum_{g=1}^{G} \left[\overline{Y}_{c,\,g,\,\bullet,\,\bullet} - \overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet} \right]^2 - \left[\dfrac{\hat{\sigma}_1^2}{X} + \dfrac{\hat{\sigma}_0^2}{X \cdot T} \right]$
$\sigma_3^2 = Var[\mu_3(\Psi_c)]$	$z^{(3)} = C \cdot G \cdot \hat{\alpha}^{(2)}, \quad \overline{B}_{\bullet}^{(3)} = \overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}, \quad c^{(3)} = 1$
$\hat{\sigma}_3^2 = \max[\hat{T}^{(3)},\, 0]$	$\hat{T}^{(3)} = \dfrac{1}{C-1} \sum_{c=1}^{C} \left[\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet} - \overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet} \right]^2 - \left[\dfrac{\hat{\sigma}_2^2}{G} + \dfrac{\hat{\sigma}_1^2}{G \cdot X} + \dfrac{\hat{\sigma}_0^2}{G \cdot X \cdot T} \right]$

Table B.4: Hierarchical credibility estimation of σ_0^2, σ_1^2, σ_2^2 and σ_3^2 under all $w_{c,\,g,\,x,\,t} = 1$

From (A.4)–(A.6), we observe that

- the common credibility estimate $\hat{Y}_{c,\,g}$ for all genders g ($g = 1, \cdots, G$) at level 2 under country c at level 3 is a weighted average of $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$;
- the common credibility estimate $\hat{Y}_{c,\,g,\,x}$ for all ages x ($x = 1, \cdots, X$) at level 1 under gender g at level 2 and country c at level 3 is a weighted average of $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}$ and $\hat{Y}_{c,\,g}$, which results in a weighted average of $\overline{Y}_{c,\,g,\,\bullet\bullet}$, $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$; and
- the credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+1}$ for year $T+1$ at level 0 under age x at level 1, gender g at level 2, and country c at level 3 is a weighted average of $\overline{Y}_{c,\,g,\,x,\,\bullet}$ and $\hat{Y}_{c,\,g,\,x}$, which leads to a weighted average of $\overline{Y}_{c,\,g,\,x,\,\bullet}$, $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}$, $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$.

We use subscripts (g, x, t) for gender g and age x in year t for a four-level hierarchical structure being applied to a country. The structural parameters for a four-level hierarchical tree can also be estimated from those for the five-level hierarchical tree by setting $C = 1$ (for a specific country) and $\hat{\alpha}^{(3)} = 1$, and we do not need to calculate $\hat{T}^{(3)}$ and $\hat{\sigma}_3^2$ in Table B.4. Then the hierarchical credibility estimate of the decrement in the logarithm of central death rate over $[T,\, T+1]$ for gender g and age x under the special case is

$$\hat{Y}_{g,\,x,\,T+1} = \hat{\alpha}^{(1)} \cdot \overline{Y}_{g,\,x,\,\bullet} + [(1-\hat{\alpha}^{(1)}) \cdot \hat{\alpha}^{(2)}] \cdot \overline{Y}_{g,\,\bullet,\,\bullet} + [(1-\hat{\alpha}^{(1)}) \cdot (1-\hat{\alpha}^{(2)})] \cdot \overline{Y}_{\bullet,\,\bullet,\,\bullet}, \tag{A.7}$$

where $\overline{Y}_{g,\,x,\,\bullet} = \frac{1}{T} \sum_{t=1}^{T} Y_{g,\,x,\,t}$, $\overline{Y}_{g,\,\bullet,\,\bullet} = \frac{1}{X} \sum_{x=1}^{X} \overline{Y}_{g,\,x,\,\bullet} = \frac{1}{X \cdot T} \sum_{x=1}^{X} \sum_{t=1}^{T} Y_{g,\,x,\,t}$, and

$$\overline{Y}_{\bullet,\,\bullet,\,\bullet} = \frac{1}{G} \sum_{g=1}^{G} \overline{Y}_{g,\,\bullet,\,\bullet} = \frac{1}{G \cdot X \cdot T} \sum_{g=1}^{G} \sum_{x=1}^{X} \sum_{t=1}^{T} Y_{g,\,x,\,t}.$$

Note that $\hat{\sigma}_0^2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ in the expressions for $\hat{\alpha}^{(1)}$ and $\hat{\alpha}^{(2)}$ (see Tables B.3 and B.4) become

$$\hat{\sigma}_0^2 = \frac{1}{G \cdot X} \sum_{g=1}^{G} \sum_{x=1}^{X} \left[\frac{1}{T-1} \sum_{t=1}^{T} (Y_{g,\,x,\,t} - \overline{Y}_{g,\,x,\,\bullet})^2 \right],$$

$\hat{\sigma}_1^2 = \frac{1}{G} \sum_{g=1}^{G} \max[\hat{T}_g^{(1)}, 0]$ and $\hat{\sigma}_2^2 = \max[\hat{T}^{(2)}, 0]$, where

$$\hat{T}_g^{(1)} = \frac{1}{X-1} \sum_{x=1}^{X} \left[\overline{Y}_{g,\,x,\,\bullet} - \overline{Y}_{g,\,\bullet,\,\bullet} \right]^2 - \frac{\hat{\sigma}_0^2}{T},$$

and

$$\hat{T}^{(2)} = \frac{1}{G-1} \sum_{g=1}^{G} \left[\overline{Y}_{g,\,\bullet,\,\bullet} - \overline{Y}_{\bullet,\,\bullet,\,\bullet} \right]^2 - \left[\frac{\hat{\sigma}_1^2}{X} + \frac{\hat{\sigma}_0^2}{X \cdot T} \right].$$

If $C = G = 1$ (mortality data are applied to a specific population of gender and country), the five-level hierarchical structure reduces to three-level one, and the non-parametric classical Bühlmann credibility model recovers. In this case, $\mu_1(\Theta_{c,\,g,\,x}) = E[Y_{c,\,g,\,x,\,t}|\Theta_{c,\,g,\,x}]$, $\sigma_1^2(\Theta_{c,\,g,\,x}) = Var[Y_{c,\,g,\,x,\,t}|\Theta_{c,\,g,\,x}]$, $\mu_2(\Phi_{c,\,g}) = E[E(Y_{c,\,g,\,x,\,t}|\Theta_{c,\,g,\,x})|\Phi_{c,\,g}]$, $\sigma_0^2 = E[\sigma_1^2(\Theta_{c,\,g,\,x})]$, and $\sigma_1^2 = E\{Var[\mu_1(\Theta_{c,\,g,\,x})|\Phi_{c,\,g}]\}$ become (we use subscripts $(x,\,t)$ for age x in year t under a three-level hierarchical structure, and ignore subscripts $(c,\,g) = (1,\,1)$ for simplifying notations)

- $\mu_1(\Theta_x) = E[Y_{x,\,t}|\Theta_x]$, the hypothetical mean;
- $\sigma_1^2(\Theta_x) = Var[Y_{x,\,t}|\Theta_x]$, the process variance;
- $\mu_2 = E[\mu_1(\Theta_x)] = E[E(Y_{x,\,t}|\Theta_x)] = E[Y_{x,\,t}]$, the expected value of the hypothetical means;
- $\sigma_0^2 = E[\sigma_1^2(\Theta_x)]$, the expected value of the process variance; and
- $\sigma_1^2 = Var[\mu_1(\Theta_x)]$, the variance of the hypothetical mean.

Then the non-parametric Bühlmann credibility estimate of the decrement in the logarithm of central death rate for age x over $[T,\,T{+}1]$, $\hat{Y}_{x,\,T+1}$, can be obtained by setting $\hat{\alpha}^{(2)} = \hat{\alpha}^{(3)} = 1$ in (A.6) as (see also Tsai and Lin (2017b) and Klugman et al. (2012))

$$\hat{Y}_{x,\,T+1} = \hat{\alpha}^{(1)} \cdot \overline{Y}_{x,\,\bullet} + (1 - \hat{\alpha}^{(1)}) \cdot \overline{Y}_{\bullet,\,\bullet}, \tag{A.8}$$

where $\overline{Y}_{x,\,\bullet} = \frac{1}{T} \sum_{t=1}^{T} Y_{x,\,t}$ and $\hat{\mu}_2 = \overline{Y}_{\bullet,\,\bullet} = \frac{1}{X} \sum_{x=1}^{X} \overline{Y}_{x,\,\bullet} = \frac{1}{X \cdot T} \sum_{x=1}^{X} \sum_{t=1}^{T} Y_{x,\,t}$. Note that $\hat{\sigma}_0^2$ and $\hat{\sigma}_1^2$ (given in Table B.4) in the expression for $\hat{\alpha}^{(1)}$ (see Table B.3) become

$$\hat{\sigma}_0^2 = \frac{1}{X} \sum_{x=1}^{X} \left[\frac{1}{T-1} \sum_{t=1}^{T} \left(Y_{x,\,t} - \overline{Y}_{x,\,\bullet} \right)^2 \right],$$

and $\hat{\sigma}_1^2 = \max[\hat{T}^{(1)}, 0]$, where

$$\hat{T}^{(1)} = \frac{1}{X-1} \sum_{x=1}^{X} \left[\overline{Y}_{x,\,\bullet} - \overline{Y}_{\bullet,\,\bullet} \right]^2 - \frac{\hat{\sigma}_0^2}{T}.$$

Appendix C

Proof of Proposition 1

C.1 Proof of Proposition 1 (a)

Proof: By (3.2),

$$\begin{aligned}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{c,\,g,\,x,\,T+\tau} &= \hat{\alpha}_{\tau}^{(1)}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau}+[(1-\hat{\alpha}_{\tau}^{(1)})\cdot\hat{\alpha}_{\tau}^{(2)}]\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad +[(1-\hat{\alpha}_{\tau}^{(1)})\cdot(1-\hat{\alpha}_{\tau}^{(2)})\cdot\hat{\alpha}_{\tau}^{(3)}]\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad +[(1-\hat{\alpha}_{\tau}^{(1)})\cdot(1-\hat{\alpha}_{\tau}^{(2)})\cdot(1-\hat{\alpha}_{\tau}^{(3)})]\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &= \hat{\alpha}_{\tau}^{(1)}\cdot C\cdot G\cdot X\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}+[(1-\hat{\alpha}_{\tau}^{(1)})\cdot\hat{\alpha}_{\tau}^{(2)}]\cdot C\cdot G\cdot X\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad +[(1-\hat{\alpha}_{\tau}^{(1)})\cdot(1-\hat{\alpha}_{\tau}^{(2)})\cdot\hat{\alpha}_{\tau}^{(3)}]\cdot C\cdot G\cdot X\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad +[(1-\hat{\alpha}_{\tau}^{(1)})\cdot(1-\hat{\alpha}_{\tau}^{(2)})\cdot(1-\hat{\alpha}_{\tau}^{(3)})]\cdot C\cdot G\cdot X\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &= C\cdot G\cdot X\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}.\end{aligned}$$

Dividing $(C\cdot G\cdot X)$ on both sides, we have for $\tau=1,2,\cdots,$

$$\frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{c,\,g,\,x,\,T+\tau}=\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}=\frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau}.$$

C.2 Proof of Proposition 1 (b)

Proof: This proposition is proved by mathematical induction on τ. First, for $\tau=2$, by definition, (3.3) and Proposition 1,

$$
\begin{aligned}
\overline{Y}^{T+2}_{\bullet,\bullet,\bullet,\bullet} &= \frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}^{T+2}_{c,\,g,\,x,\,\bullet} \\
&= \frac{1}{C\cdot G\cdot X\cdot (T+1)}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\left[\sum_{t=1}^{T}Y_{c,\,g,\,x,\,t}+\hat{Y}_{c,\,g,\,x,\,T+1}\right] \\
&= \frac{T}{T+1}\cdot\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}+\frac{1}{T+1}\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}.
\end{aligned}
$$

Next, assume that $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}$ holds. Then (3.3) and Proposition 1 lead to

$$
\begin{aligned}
\overline{Y}^{T+\tau+1}_{\bullet,\bullet,\bullet,\bullet} &= \frac{1}{C\cdot G\cdot X}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\overline{Y}^{T+\tau+1}_{c,\,g,\,x,\,\bullet} \\
&= \frac{1}{C\cdot G\cdot X\cdot (T+\tau)}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\left[\left(\sum_{t=1}^{T}Y_{c,\,g,\,x,\,t}+\sum_{t=T+1}^{T+\tau-1}\hat{Y}_{c,\,g,\,x,\,t}\right)+\hat{Y}_{c,\,g,\,x,\,T+\tau}\right] \\
&= \frac{1}{C\cdot G\cdot X\cdot (T+\tau)}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\left[(T+\tau-1)\cdot\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}+\hat{Y}_{c,\,g,\,x,\,T+\tau}\right] \\
&= \frac{T+\tau-1}{T+\tau}\cdot\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}+\frac{1}{T+\tau}\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}.
\end{aligned}
$$

Therefore, we prove that $\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet}$ for $\tau=2,3,\cdots$ under the EW strategy.

C.3 Proof of Proposition 1 (c)

Proof: Let $f_\tau=(T+\tau)\cdot\hat{\sigma}_1^2+\hat{\sigma}_0^2$, $g_\tau=X\cdot(T+\tau)\cdot\hat{\sigma}_2^2+f_\tau$ and $h_\tau=G\cdot X\cdot(T+\tau)\cdot\hat{\sigma}_3^2+g_\tau$. Then by (3.7), (3.8) and (3.9), we have

$$
\begin{array}{lll}
\hat{\alpha}^{(1)}_\tau=\frac{f_{\tau-1}-\hat{\sigma}_0^2}{f_{\tau-1}}, & 1-\hat{\alpha}^{(1)}_\tau=\frac{\hat{\sigma}_0^2}{f_{\tau-1}}, & (1-\hat{\alpha}^{(1)}_\tau)\,\hat{\alpha}^{(2)}_\tau=\frac{\hat{\sigma}_0^2(g_{\tau-1}-f_{\tau-1})}{f_{\tau-1}\cdot g_{\tau-1}}, \\
\hat{\alpha}^{(2)}_\tau=\frac{g_{\tau-1}-f_{\tau-1}}{g_{\tau-1}}, & 1-\hat{\alpha}^{(2)}_\tau=\frac{f_{\tau-1}}{g_{\tau-1}}, & (1-\hat{\alpha}^{(1)}_\tau)\,(1-\hat{\alpha}^{(2)}_\tau)\,\hat{\alpha}^{(3)}_2=\frac{\hat{\sigma}_0^2(h_{\tau-1}-g_{\tau-1})}{g_{\tau-1}\cdot h_{\tau-1}}, \\
\hat{\alpha}^{(3)}_\tau=\frac{h_{\tau-1}-g_{\tau-1}}{h_{\tau-1}}, & 1-\hat{\alpha}^{(3)}_\tau=\frac{g_{\tau-1}}{h_{\tau-1}}, & (1-\hat{\alpha}^{(1)}_\tau)\,(1-\hat{\alpha}^{(2)}_\tau)\,(1-\hat{\alpha}^{(3)}_2)=\frac{\hat{\sigma}_0^2}{h_{\tau-1}}.
\end{array}
$$

From (3.2), we can express $\hat{Y}_{c,\,g,\,x,\,T+\tau}$ as

$$
\begin{aligned}
&\hat{Y}_{c,\,g,\,x,\,T+\tau} \\
=\ & \frac{1}{f_{\tau-1}\,g_{\tau-1}\,h_{\tau-1}}\Big[g_{\tau-1}\,h_{\tau-1}\,(f_{\tau-1}-\hat{\sigma}_0^2)\cdot\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}+\hat{\sigma}_0^2\,(g_{\tau-1}-f_{\tau-1})\,h_{\tau-1}\cdot\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} \\
&+\hat{\sigma}_0^2\,(h_{\tau-1}-g_{\tau-1})\,f_{\tau-1}\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}+\hat{\sigma}_0^2\,f_{\tau-1}\,g_{\tau-1}\cdot\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}\Big]. \qquad \text{(C.1)}
\end{aligned}
$$

Our goal is to show $f_\tau\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau+1} = f_\tau\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau}$, which implies $\hat{Y}_{c,\,g,\,x,\,T+\tau+1} = \hat{Y}_{c,\,g,\,x,\,T+\tau}$, $\tau = 1,\, 2,\, \cdots$, and thus $\hat{Y}_{c,\,g,\,x,\,T+\tau} = \hat{Y}_{c,\,g,\,x,\,T+1}$, $\tau = 2,\, 3,\, \cdots$. First, we let

$$DIFF = f_\tau\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau+1} - f_{\tau-1}\, g_{\tau-1}\, h_{\tau-1} \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau}.$$

Then our goal changes to prove

$$f_\tau\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau} = f_\tau\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau+1} = f_{\tau-1}\, g_{\tau-1}\, h_{\tau-1} \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau} + DIFF,$$

or equivalently, $DIFF = (f_\tau \cdot g_\tau \cdot h_\tau - f_{\tau-1} \cdot g_{\tau-1} \cdot h_{\tau-1}) \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau}$.

By (C.1), $DIFF$ is the sum of the following four expressions (C.2)−(C.5):

$$\begin{aligned}
&g_\tau\, h_\tau\, (f_\tau - \hat{\sigma}_0^2) \cdot \overline{Y}^{T+\tau+1}_{c,\,g,\,x,\,\bullet} - g_{\tau-1}\, h_{\tau-1}\, (f_{\tau-1} - \hat{\sigma}_0^2) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} \\
&= \hat{\sigma}_1^2 \Bigg[g_\tau h_\tau \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) - g_{\tau-1} h_{\tau-1} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) \Bigg] \\
&= \hat{\sigma}_1^2 \Bigg[(g_\tau\, h_\tau - g_{\tau-1}\, h_{\tau-1}) \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) + g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau} \Bigg] \\
&= (g_\tau\, h_\tau - g_{\tau-1}\, h_{\tau-1})\, (f_{\tau-1} - \hat{\sigma}_0^2) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} + \hat{\sigma}_1^2\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau},
\end{aligned} \tag{C.2}$$

$$\begin{aligned}
&\hat{\sigma}_0^2 (g_\tau - f_\tau)\, h_\tau \cdot \overline{Y}^{T+\tau+1}_{c,\,g,\,\bullet,\,\bullet} - \hat{\sigma}_0^2\, (g_{\tau-1} - f_{\tau-1})\, h_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} \\
&= \hat{\sigma}_0^2\, \hat{\sigma}_2^2 \Bigg[h_\tau \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) - h_{\tau-1} \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) \Bigg] \\
&= \hat{\sigma}_0^2 \hat{\sigma}_2^2 \Bigg[(h_\tau - h_{\tau-1}) \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) + h_\tau \sum_{x=1}^{X} \hat{Y}_{c,\,g,\,x,\,T+\tau} \Bigg] \\
&= \hat{\sigma}_0^2\, (h_\tau - h_{\tau-1})\, (g_{\tau-1} - f_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} + \hat{\sigma}_0^2\, \hat{\sigma}_2^2\, h_\tau \sum_{x=1}^{X} \hat{Y}_{c,\,g,\,x,\,T+\tau},
\end{aligned} \tag{C.3}$$

$$
\begin{aligned}
&\hat{\sigma}_0^2\,(h_\tau - g_\tau)\,f_\tau \cdot \overline{Y}^{T+\tau+1}_{c,\bullet,\bullet,\bullet} - \hat{\sigma}_0^2\,(h_{\tau-1} - g_{\tau-1})\,f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} \\
&= \hat{\sigma}_0^2\,\hat{\sigma}_3^2 \Bigg[f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,g,x,t}\Bigg) \\
&\quad - f_{\tau-1} \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,g,x,t}\Bigg)\Bigg] \\
&= \hat{\sigma}_0^2\,\hat{\sigma}_3^2 \Bigg[(f_\tau - f_{\tau-1}) \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,g,x,t}\Bigg) + f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X} \hat{Y}_{c,g,x,T+\tau}\Bigg] \\
&= \hat{\sigma}_0^2\,(f_\tau - f_{\tau-1})\,(h_{\tau-1} - g_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} + \hat{\sigma}_0^2\,\hat{\sigma}_3^2 f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X} \hat{Y}_{c,g,x,T+\tau},
\end{aligned}
\tag{C.4}
$$

$$
\hat{\sigma}_0^2\, f_\tau g_\tau \cdot \overline{Y}^{T+\tau+1}_{\bullet,\bullet,\bullet,\bullet} - \hat{\sigma}_0^2\, f_{\tau-1}\, g_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet} = \hat{\sigma}_0^2\,(f_\tau\, g_\tau - f_{\tau-1}\, g_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}. \tag{C.5}
$$

Next,

$$
f_\tau g_\tau h_\tau - f_{\tau-1} g_{\tau-1} h_{\tau-1} = (f_{\tau-1} + \hat{\sigma}_1^2) g_\tau h_\tau - f_{\tau-1} g_{\tau-1} h_{\tau-1} = \hat{\sigma}_1^2 g_\tau h_\tau + f_{\tau-1}(g_\tau h_\tau - g_{\tau-1} h_{\tau-1}),
$$

and

$$
(f_\tau g_\tau h_\tau - f_{\tau-1} g_{\tau-1} h_{\tau-1}) \cdot \hat{Y}_{c,g,x,T+\tau} = \hat{\sigma}_1^2 g_\tau h_\tau \cdot \hat{Y}_{c,g,x,T+\tau} + f_{\tau-1}(g_\tau h_\tau - g_{\tau-1} h_{\tau-1}) \cdot \hat{Y}_{c,g,x,T+\tau}. \tag{C.6}
$$

The first term of (C.6) cancels out the second term of (C.2), and the second term of (C.6) by (C.1) gives

$$
\begin{aligned}
&(g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(f_{\tau-1} - \hat{\sigma}_0^2) \cdot \overline{Y}^{T+\tau}_{c,g,x,\bullet} + \frac{\hat{\sigma}_0^2}{g_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(g_{\tau-1} - f_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,g,\bullet,\bullet} \\
&+ \frac{\hat{\sigma}_0^2}{g_{\tau-1} h_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(h_{\tau-1} - g_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} \\
&+ \frac{\hat{\sigma}_0^2}{h_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}.
\end{aligned}
\tag{C.7}
$$

The first term of (C.7) cancels out the first term of (C.2), so the remaining task is to prove that the sum of the last three terms of (C.7) equals the sum of (C.3), (C.4) and (C.5). Since

$$\begin{aligned}\sum_{x=1}^{X}\hat{Y}_{c,\,g,\,x,\,T+\tau} =&\frac{X}{f_{\tau-1}g_{\tau-1}h_{\tau-1}}\Big\{\Big[g_{\tau-1}h_{\tau-1}(f_{\tau-1}-\hat{\sigma}_0^2)+\hat{\sigma}_0^2(g_{\tau-1}-f_{\tau-1})h_{\tau-1}\Big]\cdot\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}\\ &+\hat{\sigma}_0^2(h_{\tau-1}-g_{\tau-1})f_{\tau-1}\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}+\hat{\sigma}_0^2f_{\tau-1}g_{\tau-1}\cdot\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}\Big\}\\ =&\frac{X}{f_{\tau-1}g_{\tau-1}h_{\tau-1}}\Big[f_{\tau-1}h_{\tau-1}(g_{\tau-1}-\hat{\sigma}_0^2)\cdot\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}\\ &+\hat{\sigma}_0^2(h_{\tau-1}-g_{\tau-1})f_{\tau-1}\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}+\hat{\sigma}_0^2f_{\tau-1}g_{\tau-1}\cdot\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}\Big]\\ =&\frac{X(g_{\tau-1}-\hat{\sigma}_0^2)}{g_{\tau-1}}\cdot\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}+\frac{X\hat{\sigma}_0^2(h_{\tau-1}-g_{\tau-1})}{g_{\tau-1}h_{\tau-1}}\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}+\frac{X\hat{\sigma}_0^2}{h_{\tau-1}}\cdot\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}\end{aligned} \tag{C.8}$$

and

$$\begin{aligned}\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{c,\,g,\,x,\,T+\tau} =&\frac{GX}{f_{\tau-1}g_{\tau-1}h_{\tau-1}}\Big\{\Big[f_{\tau-1}h_{\tau-1}(g_{\tau-1}-\hat{\sigma}_0^2)+\hat{\sigma}_0^2(h_{\tau-1}-g_{\tau-1})f_{\tau-1}\Big]\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}\\ &+\hat{\sigma}_0^2f_{\tau-1}g_{\tau-1}\cdot\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}\Big\}\\ =&\frac{GX(h_{\tau-1}-\hat{\sigma}_0^2)}{h_{\tau-1}}\cdot\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}+\frac{GX\hat{\sigma}_0^2}{h_{\tau-1}}\cdot\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet},\end{aligned} \tag{C.9}$$

comparing the coefficients of $\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}$, $\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$ in the last three terms of (C.7) with those in (C.3), (C.4) and (C.5) associated with (C.8) and (C.9), it is sufficient to show

1. coefficient of $\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}$:

$$\frac{\hat{\sigma}_0^2}{g_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})(g_{\tau-1}-f_{\tau-1})=\hat{\sigma}_0^2(h_\tau-h_{\tau-1})(g_{\tau-1}-f_{\tau-1})+\frac{X\hat{\sigma}_0^2\hat{\sigma}_2^2h_\tau(g_{\tau-1}-\hat{\sigma}_0^2)}{g_{\tau-1}}; \tag{C.10}$$

2. coefficient of $\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}$:

$$\begin{aligned}&\frac{\hat{\sigma}_0^2}{g_{\tau-1}\,h_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})(h_{\tau-1}-g_{\tau-1})f_{\tau-1}\\ =\;&\frac{X\hat{\sigma}_0^4\hat{\sigma}_2^2h_\tau(h_{\tau-1}-g_{\tau-1})}{g_{\tau-1}h_{\tau-1}}+\hat{\sigma}_0^2(f_\tau-f_{\tau-1})(h_{\tau-1}-g_{\tau-1})+\frac{GX\hat{\sigma}_0^2\hat{\sigma}_3^2f_\tau(h_{\tau-1}-\hat{\sigma}_0^2)}{h_{\tau-1}};\end{aligned} \tag{C.11}$$

3. coefficient of $\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$:

$$\frac{\hat{\sigma}_0^2}{h_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})f_{\tau-1}=\frac{X\,\hat{\sigma}_0^4\,\hat{\sigma}_2^2\,h_\tau}{h_{\tau-1}}+\frac{G\,X\,\hat{\sigma}_0^4\,\hat{\sigma}_3^2\,f_\tau}{h_{\tau-1}}+\hat{\sigma}_0^2\,(f_\tau\,g_\tau-f_{\tau-1}\,g_{\tau-1}). \tag{C.12}$$

After rearrangement and simplification, the three equations (C.10), (C.11) and (C.12) we need to show become

$$(g_{\tau-1} - f_{\tau-1})(g_\tau - g_{\tau-1}) = X\hat{\sigma}_2^2(g_{\tau-1} - \hat{\sigma}_0^2),$$

$$(h_{\tau-1} - g_{\tau-1})(f_{\tau-1}g_\tau h_\tau - f_\tau g_{\tau-1}h_{\tau-1} - X\hat{\sigma}_0^2\hat{\sigma}_2^2 h_\tau) = GX\hat{\sigma}_3^2 f_\tau g_{\tau-1}(h_{\tau-1} - \hat{\sigma}_0^2),$$

$$g_\tau(f_{\tau-1}h_\tau - f_\tau h_{\tau-1}) = X\hat{\sigma}_0^2(\hat{\sigma}_2^2 h_\tau + G\hat{\sigma}_3^2 f_\tau),$$

respectively, which can be verified directly by the definitions of f_τ, g_τ and h_τ.

Bibliography

Bühlmann, Hans (1967). "Experience rating and credibility". In: *Astin Bulletin* 4(3), pp. 199–207.

Bühlmann, Hans and Gisler, Alois (2005). *A course in credibility theory and its applications.* Springer Science & Business Media.

Bühlmann, Hans and Straub, Erwin (1970). "Glaubwürdigkeit für schadensätze". In: *Bulletin of the Swiss Association of Actuaries* 70(1), pp. 111–133.

Cairns, Andrew J.G., Blake, David, and Dowd, Kevin (2006). "A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration". In: *Journal of Risk and Insurance* 73(4), pp. 687–718.

Cairns, Andrew J.G., Blake, David, Dowd, Kevin, Coughlan, Guy D., and Khalaf-Allah, Marwa (2011). "Bayesian stochastic mortality modelling for two populations". In: *Astin Bulletin* 41(1), pp. 29–59.

Carter, Lawrence R. and Lee, Ronald D. (1992). "Forecasting demographic components: Modeling and forecasting U.S. sex differentials in mortality". In: *International Journal of Forecasting* 8, pp. 393–411.

Chen, Hua, MacMinn, Richard, and Sun, Tao (2015). "Multi-population mortality models: A factor copula approach". In: *Insurance: Mathematics and Economics* 63, pp. 135–146.

Dannenburg, Dennis (1995). "Crossed classification credibility models". In: *Insurance: Mathematics and Economics* 17(1), p. 65.

Goulet, Vincent (2001). "A generalized crossed classification credibility model". In: *Insurance: Mathematics and Economics* 28(2), pp. 205–216.

Jewell, William S. (1975). "The use of collateral data in credibility theory: A hierarchical model". In: *Austria: International Institute for Applied Systems Analysis.*

Kleinow, Torsten (2015). "A common age effect model for the mortality of multiple populations". In: *Insurance: Mathematics and Economics* 63, pp. 147–152.

Klugman, Stuart A., Panjer, Harry H., and Willmot, Gordon E. (2012). *Loss models: From data to decisions, 4th Edition.* John Wiley & Sons.

Lee, Ronald D. and Carter, Lawrence R. (1992). "Modeling and forecasting US mortality". In: *Journal of the American Statistical Association* 87(419), pp. 659–671.

Li, Johnny Siu-Hang and Hardy, Mary R. (2011). "Measuring basis risk in longevity hedges". In: *North American Actuarial Journal* 15(2), pp. 177–200.

Li, Johnny Siu-Hang, Hardy, Mary R., and Tan, Ken Seng (2009). "Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach". In: *Astin Bulletin* 39(1), pp. 137–164.

Li, Nan and Lee, Ronald D. (2005). "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method". In: *Demography* 42(3), pp. 575–594.

Lin, Tzuling, Wang, Chou-Wen, and Tsai, Cary Chi-Liang (2015). "Age-specific copula-AR-GARCH mortality models". In: *Insurance: Mathematics and Economics* 61, pp. 110–124.

Mitchell, Daniel, Brockett, Patrick, Mendoza-Arriaga, Rafael, and Muthuraman, Kumar (2013). "Modeling and forecasting mortality rates". In: *Insurance: Mathematics and Economics* 52(2), pp. 275–285.

Plat, Richard (2009). "On stochastic mortality modeling". In: *Insurance: Mathematics and Economics* 45(3), pp. 393–404.

Poon, Jimmy and Lu, Yi (2015). "A spatial cross-sectional credibility model with dependence among risks". In: *North American Actuarial Journal* 19(4), pp. 289–310.

Renshaw, Arthur E. and Haberman, Steven (2006). "A cohort-based extension to the Lee–Carter model for mortality reduction factors". In: *Insurance: Mathematics and Economics* 38(3), pp. 556–570.

Tsai, Cary Chi-Liang and Lin, Tzuling (2017a). "Incorporating the Bühlmann credibility into mortality models to improve forecasting performances". In: *Scandinavian Actuarial Journal* 2017(5), pp. 419–440.

— (2017b). "A Bühlmann credibility approach to modeling mortality rates". In: *North American Actuarial Journal* 21(2), pp. 204–227.

Tsai, Cary Chi-Liang and Yang, Shuai (2015). "A linear regression approach to modeling mortality rates of different forms". In: *North American Actuarial Journal* 19(1), pp. 1–23.

Wang, Chou-Wen, Yang, Sharon S, and Huang, Hong-Chih (2015). "Modeling multi-country mortality dependence and its application in pricing survivor index swaps-A dynamic copula approach". In: *Insurance: Mathematics and Economics* 63, pp. 30–39.

Wen, Limin and Wu, Xianyi (2011). "The credibility estimator with general dependence structure over risks". In: *Communications in Statistics-Theory and Methods* 40(10), pp. 1893–1910.

Wen, Limin, Wu, Xianyi, and Zhou, Xian (2009). "The credibility premiums for models with dependence induced by common effects". In: *Insurance: Mathematics and Economics* 44(1), pp. 19–25.

Yang, Sharon S and Wang, Chou-Wen (2013). "Pricing and securitization of multi-country longevity risk with mortality dependence". In: *Insurance: Mathematics and Economics* 52(2), pp. 157–169.

Yeo, Keng Leong and Valdez, Emiliano A. (2006). "Claim dependence with common effects in credibility models". In: *Insurance: Mathematics and Economics* 38(3), pp. 609–629.

Zhou, Rui, Wang, Yujiao, Kaufhold, Kai, Li, Johnny Siu-Hang, and Tan, Ken Seng (2014). "Modeling period effects in multi-population mortality models: Applications to Solvency II". In: *North American Actuarial Journal* 18(1), pp. 150–167.