Canadian Institute of Actuaries
Institut canadien des actuaires

Research Paper

A Hierarchical Credibility Approach to Modelling Mortality Rates for Multiple Populations

Academic Research Subcommittee – Research Executive Committee

April 2018

Document 218042

Ce document est disponible en français

A Hierarchical Credibility Approach to Modelling Mortality Rates for Multiple Populations

Cary Chi-Liang Tsai and Adelaide Di Wu

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC

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 paper 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 five/four/three-level hierarchical credibility models are compared with those of the classical Lee-Carter model and its three extensions for multiple populations (joint-k, co-integrated and augmented common factor Lee-Carter models). Numerical illustrations based on mortality data for both genders of the U.S.A, 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 the predicted multi-population mortality rates can be used to construct a mortality index for better pricing mortality-indexed securities.

Keywords: hierarchical credibility theory; Bühlmann credibility theory; Lee-Carter model; multi-population mortality model.

1. Introduction

Mortality is one of the key factors in determining the premiums and reserves of life insurance and annuity products and pricing mortality-linked securities, (e.g., q-forwards, longevity bonds, survivor swaps, annuity futures, mortality options, and survivor caps). Longevity risk has become a substantial issue in the human society. Over recent decades, mortality rates have displayed a dramatic improvement. Since life annuity contracts and pension plans often last for decades, with a gradual increase in life expectancy of most developed countries, annuity providers, retirement programs and pension/long-term care systems face a significant risk of paying more survival benefits than expected to retirees and annuitants, which is called longevity risk. Such longevity risk could lead to financial distress or insolvency for annuity providers, retirement programs and social security systems. At the same time, life insurers face mortality risk due to catastrophic mortality deterioration (e.g., the 1918 Spanish flu pandemic, and the 2004 Indian Ocean earthquake and tsunami). To reduce mortality risk, Swiss Re issued a three-year catastrophe mortality bond of \$400 million in December 2003. The principle paid to the investor at maturity depends on a mortality index—a specifically constructed index of mortality rates across both genders of five countries (the US, the UK, France, Italy and Switzerland). Therefore, building effective mortality models in order to provide accurate mortality rates for better pricing life insurance policies, annuity products, pension plans, social security systems and mortality-linked securities has been a matter of great urgency.

An effective multi-population mortality model can help construct an accurate mortality index on which pricing of mortality-indexed securities and derivatives can be based to hedge mortality and longevity risks for life insurers and annuity providers. When modelling multi-population mortality rates, although traditional mortality models can project mortality rates separately for each population, these models do not consider a correlation among populations. The mortality rates between females and males in a country are highly correlated because they are exposed to the common conditions such as public health, education systems, medical services, and living environments. The mortality rates among developed countries or provinces/states in a country also have a different degree of correlation due to similar exposures.

Projecting mortality rates and modelling their randomness have garnered much attention in the past decades. The Lee-Carter (1992) model is the most widely cited model in mortality prediction and applications. The CBD model, proposed by Cairns et al. (2006), is another well-known model designed for modelling mortality rates for seniors. Tsai and Yang (2015) related each of a series of period mortality rate sequences for consecutive years to that for a base year with a linear regression, and model each of the two resulting sequences of intercept and slope parameters with a random walk with drift for predicting mortality rates. 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. Numerous extensions of the Lee-Carter and CBD models were developed. For example, 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.

There are some extensions to the Lee-Carter model, which incorporate the dependence between the mortality rates for multiple populations. Carter and Lee (1992) introduced the joint-k Lee-Carter model, assuming that the mortality rates for two populations are jointly driven by a common time-varying index. Li and Lee (2005) proposed the augmented common factor Lee-Carter model involving common factors and population-specific ones. Li and Hardy (2011) showed the existence of cointegration in the time trends of mortality rates for two populations, and modelled the time trends of mortality rates by a linear relationship, called the cointegrated Lee-Carter model.

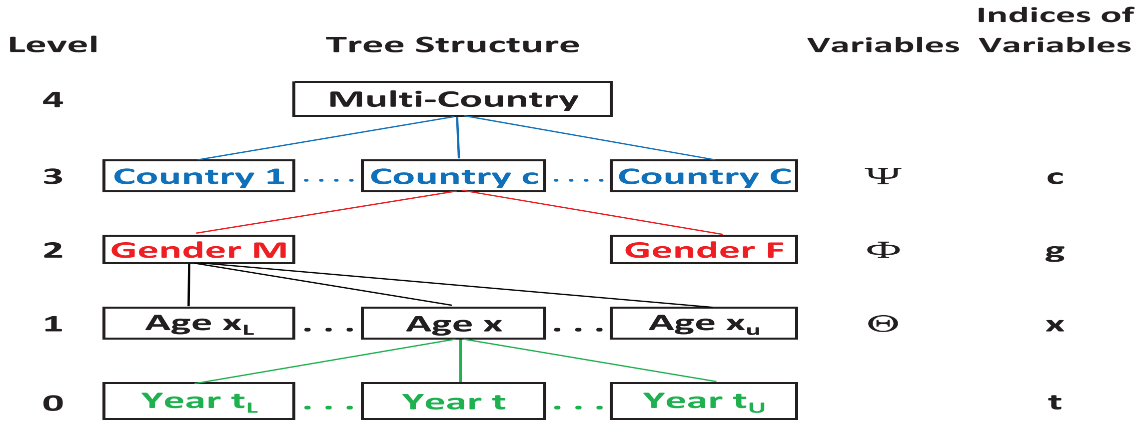
Also, there are articles, not based on the Lee-Carter model, studying the correlation of multi-population mortality rates. For example, Cairns et al. (2011) introduced a new framework for modelling the joint development over time of mortality rates in a pair of related populations with the primary aim of producing consistent mortality forecasts for two populations. Wang et al. (2015) used time-varying copula models to capture the mortality dependence structure across countries, examining both symmetric and asymmetric dependence structures. Chen et al. (2015) first filtered the mortality dynamics of each population using an ARMA-GARCH process with heavy-tailed innovations, and then modelled the residual risk using a one-factor copula model that is widely applicable to high dimension data and very flexible in terms of model specification. Li et al. (2015) proposed a systematic process for constructing a two-population mortality model and developed two-population generalizations for each of the seven single-population models studied in Cairns et al. (2009). Li et al. (2017) introduced a new concept called semi-coherence to produce semi-coherent mortality forecasts using a vector threshold autoregression. Enchev et al. (2016) reviewed a number of multi-population mortality models and developed forecasting models that produce non-diverging and joint mortality rate scenarios.

Credibility theory is widely applied in property and casualty insurance. Bühlmann (1967) proposed a credibility formula to determine the credibility estimate which is calculated by a weighted average of the sample mean of the past claim data of a policyholder and the true mean of claims. Bühlmann and Straub (1970) extended the Bühlmann model by allowing unequal number of exposure units for each risk. 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. Tsai and Lin (2017a) incorporated a Bühlmann credibility approach to the Lee-Carter, CBD, and Tsai and Yang (2015) models to improve mortality forecasting performances. Tsai and Lin (2017b) proposed a non-parametric Bühlmann credibility approach to modelling mortality rates and showed that the approach outperforms the Lee-Carter model in forecasting performance.

Bühlmann and Gisler (2005) incorporated the idea of hierarchical structure to the credibility theory to achieve the hierarchical credibility model. It can be seen that the hierarchical credibility is a generalization of the Bühlmann model and the Bühlmann-Straub model with a tree structure of more levels. Traditional mortality models for single population group mortality data into three levels (population, age and year). In this paper, we would like to propose a hierarchical credibility approach to modelling multi-population mortality rates. We group the mortality data for both genders of three developed countries into four levels (country, gender, age and year) or five levels (multi-country, country, gender, age and year), apply the hierarchical credibility to each level to better reflect the correlation of mortality rates among populations or genders/countries, and compare the forecasting performances of the underlying mortality models.

The remainder of this paper proceeds as follows. Section 2 introduces the four-level and five-level hierarchical credibility approaches to modelling multi-population mortality rates and applies them to projecting mortality rates for six populations (both genders of the US, the UK and Japan). Section 3 compares the forecasting performances of the proposed three/four/five-level hierarchical mortality models and the Lee-Carter model and its three variations (the joint-k, the cointegrated and the augmented common factor models; see Appendix A) for multiple populations. Section 4 concludes this paper.

Figure 1: A five-level hierarchical tree structure



2. Hierarchical Bühlmann credibility mortality model

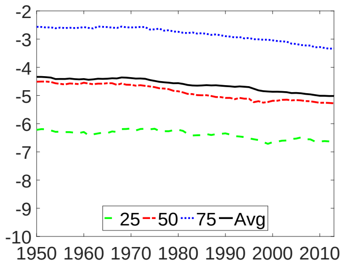
This section applies the hierarchical credibility idea from Bühlmann and Gisler (2005) to propose a hierarchical credibility mortality model for multiple populations, which is a generalization of the non-parametric Bühlmann credibility mortality model for single population proposed by Tsai and Lin (2017b).

Figure 1 structures mortality data in a five-level hierarchical tree with the top level 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. 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 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 $x = 25, 50, 75$ and Avg, where the Avg curve is the average of $\ln(m_{c,g,x,t})$s over $x = 20, \cdots, 84$. As with the Bühlmann credibility mortality model proposed by Tsai and Lin (2017a, b), we apply the hierarchical credibility approach to modelling $Y_{c,g,x,t} \triangleq \ln(m_{c,g,x,t}) - \ln(m_{c,g,x,t-1})$, the 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). 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, $Y_{c,g,x,t} = \ln(m_{c,g,x_L+x-1,t_L+t}) - \ln(m_{c,g,x_L+x-1,t_L+t-1})$ for $x = 1, \cdots, X$ and $t = 1, \cdots, T$, where $X - 1 = x_U - x_L$ and $T = t_U - t_L$.

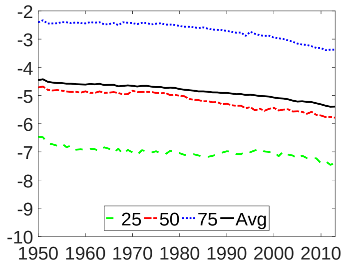
2.1 Assumptions and notations

This subsection first gives the assumptions and then the notations for conditional means, mean, conditional variances and variances for a hierarchical tree. The following are the

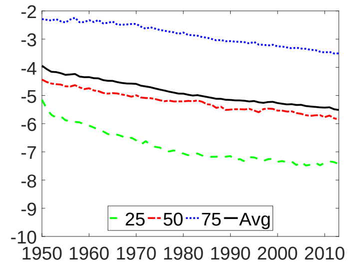
Figure 2: $\ln(m_{x,t,i})$ against t



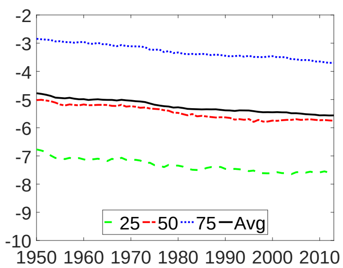
(a) US Males



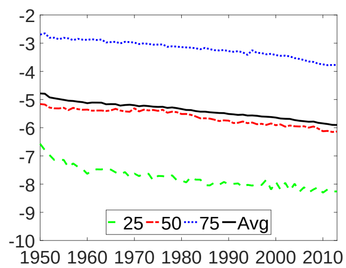
(b) UK Males



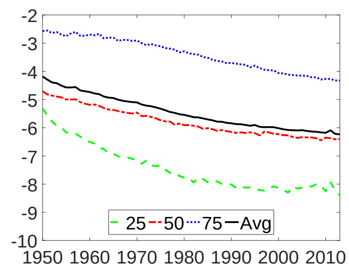
(c) Japan Males



(d) US Females

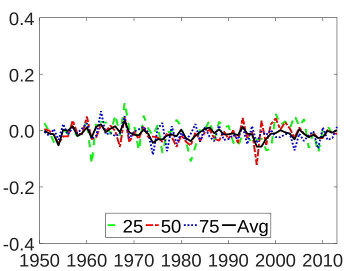


(e) UK Females

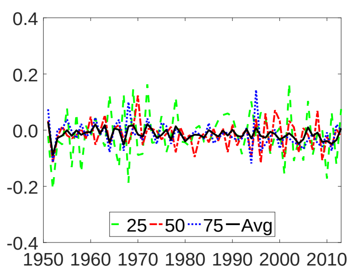


(f) Japan Females

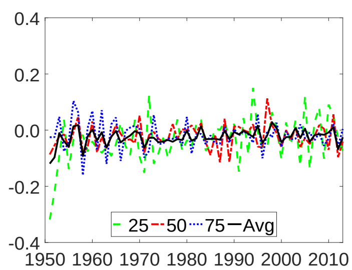
Figure 3: $Y_{x,t,i} = \ln(m_{x,t,i}) - \ln(m_{x,t-1,i})$ against t



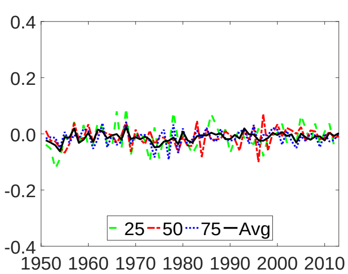
(a) US Males



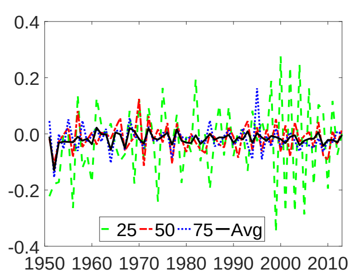
(b) UK Males



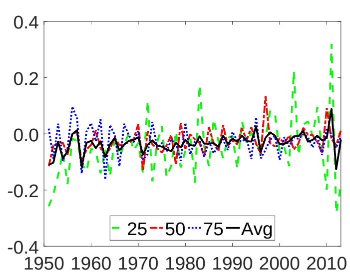
(c) Japan Males



(d) US Females



(e) UK Females



(f) Japan Females

assumptions for the hierarchical model:

- Level 0: $Y_{c,g,x,t}|\Theta_{c,g,x}$, $t = 1, \cdots, T$, are independent and identically distributed for fixed c, g and x;
- Level 1: $\Theta_{c,g,x}|\Phi_{c,g}$, $x = 1, \cdots, X$, are independent and identically distributed for fixed c and g;
- Level 2: $\Phi_{c,g}|\Psi_c$, $g = 1, \cdots, G$, are independent and identically distributed for fixed c; and
- Level 3: Ψ_c, $c = 1, \cdots, C$, are independent and identically distributed.

First, we denote the conditional means and the mean for levels 0–3 as follows:

- Level 0: $\mu_1(\Theta_{c,g,x}) \triangleq E(Y_{c,g,x,t}|\Theta_{c,g,x})$, the conditional expectation of $Y_{c,g,x,t}$ given $\Theta_{c,g,x}$;
- Level 1: $\mu_2(\Phi_{c,g}) \triangleq E[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}] = E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}]$, the conditional expectation of $\mu_1(\Theta_{c,g,x})$ given $\Phi_{c,g}$;
- Level 2: $\mu_3(\Psi_c) \triangleq E[\mu_2(\Phi_{c,g})|\Psi_c] = E\{E[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\}$ $= E\{E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\}$, the conditional expectation of $\mu_2(\Phi_{c,g})$ given Ψ_c; and
- Level 3: $\mu_4 \triangleq E[\mu_3(\Psi_c)] = E\{E[\mu_2(\Phi_{c,g})|\Psi_c]\} = E\{E\{E[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\}\}$ $= E\{E\{E[E(Y_{c,g,x,t}|\Theta_{c,g,x})|\Phi_{c,g}]|\Psi_c\}\}$, the expectation of $\mu_3(\Psi_c)$.

By the law of total expectation, it is easy to show that $\mu_4 = E[Y_{c,g,x,t}]$, $\mu_3(\Psi_c) = E[Y_{c,g,x,t}|\Psi_c]$, and $\mu_2(\Phi_{c,g}) = E[Y_{c,g,x,t}|\Phi_{c,g}]$.

Next, we denote the following conditional variances for levels 0–2:

- Level 0: $\dfrac{\sigma_1^2(\Theta_{c,g,x})}{w_{c,g,x,t}} \triangleq Var[Y_{c,g,x,t}|\Theta_{c,g,x}] = E\{[Y_{c,g,x,t} - \mu_1(\Theta_{c,g,x})]^2|\Theta_{c,g,x}\}$, the conditional variance of $Y_{c,g,x,t}$ given $\Theta_{c,g,x}$ where $w_{c,g,x,t}$ is a known exposure unit;
- Level 1: $\sigma_2^2(\Phi_{c,g}) \triangleq Var[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}] = E\{[\mu_1(\Theta_{c,g,x}) - \mu_2(\Phi_{c,g})]^2|\Phi_{c,g}\}$, the conditional variance of $\mu_1(\Theta_{c,g,x})$ given $\Phi_{c,g}$; and
- Level 2: $\sigma_3^2(\Psi_c) \triangleq Var[\mu_2(\Phi_{c,g})|\Psi_c] = E\{[\mu_2(\Phi_{c,g}) - \mu_3(\Psi_c)]^2|\Psi_c\}$, the conditional variance of $\mu_2(\Phi_{c,g})$ given Ψ_c.

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

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

- Level 3: $\sigma_4^2 \triangleq Var[\mu_3(\Psi_c)] = E\{[\mu_3(\Psi_c) - \mu_4]^2\}$, the variance of $\mu_3(\Psi_c)$.

Note that by the law of total variance, we have

$$\begin{aligned} Var[\mu_2(\Phi_{c,g})] &= E\{Var[\mu_2(\Phi_{c,g})|\Psi_c]\} + Var\{E[\mu_2(\Phi_{c,g})|\Psi_c]\} \\ &= \sigma_3^2 + Var[\mu_3(\Psi_c)] = \sigma_3^2 + \sigma_4^2, \end{aligned}$$

and

$$\begin{aligned} Var[\mu_1(\Theta_{c,g,x})] &= E\{Var[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}]\} + Var\{E[\mu_1(\Theta_{c,g,x})|\Phi_{c,g}]\} \\ &= \sigma_2^2 + Var[\mu_2(\Phi_{c,g})] = \sigma_2^2 + \sigma_3^2 + \sigma_4^2. \end{aligned}$$

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 1 for a single 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 population of country c and gender g, if further $w_{c,g,x,t} = 1$ for $x = 1, \cdots, X$ and $t = 1, \cdots, T$, then the Bühlmann-Straub credibility model reduces to the Bühlmann credibility model.

2.2 Parameter estimations

Based on Section 6.4 of Bühlmann and Gisler (2005), the compressed data B and credibility factors α are calculated from the bottom level to the top level by minimizing the associated projection functions from the top to the bottom.

Credibility estimators

To find the 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 $\hat{Y}_{c,g,x}$ and $\hat{Y}_{c,g}$. All the credibility estimators can be seen as linear combinations of the overall mean $\hat{\mu}_4$ and the compressed data B as follows: $\hat{Y}_{c,g} = \alpha_c^{(3)} \cdot B_c^{(3)} + (1 - \alpha_c^{(3)}) \cdot \hat{\mu}_4$,

$$\begin{aligned} \hat{Y}_{c,g,x} &= \alpha_{c,g}^{(2)} \cdot B_{c,g}^{(2)} + (1 - \alpha_{c,g}^{(2)}) \cdot \hat{Y}_{c,g} \\ &= \alpha_{c,g}^{(2)} \cdot B_{c,g}^{(2)} + [(1 - \alpha_{c,g}^{(2)}) \cdot \alpha_c^{(3)}] \cdot B_c^{(3)} + [(1 - \alpha_{c,g}^{(2)}) \cdot (1 - \alpha_c^{(3)})] \cdot \hat{\mu}_4, \end{aligned}$$

and

$$\begin{aligned} &\hat{Y}_{c,g,x,T+1} \\ = \ &\alpha_{c,g,x}^{(1)} \cdot B_{c,g,x}^{(1)} + (1 - \alpha_{c,g,x}^{(1)}) \cdot \hat{Y}_{c,g,x} \\ = \ &\alpha_{c,g,x}^{(1)} \cdot B_{c,g,x}^{(1)} + [(1 - \alpha_{c,g,x}^{(1)}) \cdot \alpha_{c,g}^{(2)}] \cdot B_{c,g}^{(2)} + [(1 - \alpha_{c,g,x}^{(1)}) \cdot (1 - \alpha_{c,g}^{(2)})] \cdot \hat{Y}_{c,g} \\ = \ &\alpha_{c,g,x}^{(1)} \cdot B_{c,g,x}^{(1)} + [(1 - \alpha_{c,g,x}^{(1)}) \cdot \alpha_{c,g}^{(2)}] \cdot B_{c,g}^{(2)} + [(1 - \alpha_{c,g,x}^{(1)}) \cdot (1 - \alpha_{c,g}^{(2)}) \cdot \alpha_c^{(3)}] \cdot B_c^{(3)} \\ &+ [(1 - \alpha_{c,g,x}^{(1)}) \cdot (1 - \alpha_{c,g}^{(2)}) \cdot (1 - \alpha_c^{(3)})] \cdot \hat{\mu}_4, \end{aligned} \tag{2.1}$$

where $B_{c,g,x}^{(1)}$, $B_{c,g}^{(2)}$, and $B_c^{(3)}$ represent compressed data; $\alpha_{c,g,x}^{(1)}$, $\alpha_{c,g}^{(2)}$ and $\alpha_c^{(3)}$ are the corresponding credibility factors for levels one, two and three, respectively. The expressions for the compressed data ($B_{c,g,x}^{(1)}$, $B_{c,g}^{(2)}$ and $B_c^{(3)}$), the credibility factors ($\alpha_{c,g,x}^{(1)}$, $\alpha_{c,g}^{(2)}$

Table 1: Hierarchical credibility estimation of σ_1^2, σ_2^2, σ_3^2 and σ_4^2

Panel A: Formulas for $B^{(1)}_{c,g,x}$, $B^{(2)}_{c,g}$, $B^{(3)}_{c}$, $\alpha^{(1)}_{c,g,x}$, $\alpha^{(2)}_{c,g}$, $\alpha^{(3)}_{c}$ and $\hat{\mu}_4$		
$B^{(1)}_{c,g,x} = \sum_{t=1}^{T} \frac{w_{c,g,x,t}}{w^{(1)}_{c,g,x}} \cdot Y_{c,g,x,t},$	$w^{(1)}_{c,g,x} = \sum_{t=1}^{T} w_{c,g,x,t},$	$\alpha^{(1)}_{c,g,x} = \frac{w^{(1)}_{c,g,x} \cdot \sigma_2^2}{w^{(1)}_{c,g,x} \cdot \sigma_2^2 + \sigma_1^2}$
$B^{(2)}_{c,g} = \sum_{x=1}^{X} \frac{\alpha^{(1)}_{c,g,x}}{w^{(2)}_{c,g}} \cdot B^{(1)}_{c,g,x},$	$w^{(2)}_{c,g} = \sum_{x=1}^{X} \alpha^{(1)}_{c,g,x},$	$\alpha^{(2)}_{c,g} = \frac{w^{(2)}_{c,g} \cdot \sigma_3^2}{w^{(2)}_{c,g} \cdot \sigma_3^2 + \sigma_2^2}$
$B^{(3)}_{c} = \sum_{g=1}^{G} \frac{\alpha^{(2)}_{c,g}}{w^{(3)}_{c}} \cdot B^{(2)}_{c,g},$	$w^{(3)}_{c} = \sum_{g=1}^{G} \alpha^{(2)}_{c,g},$	$\alpha^{(3)}_{c} = \frac{w^{(3)}_{c} \cdot \sigma_4^2}{w^{(3)}_{c} \cdot \sigma_4^2 + \sigma_3^2}$
$\hat{\mu}_4 = \sum_{c=1}^{C} \frac{\alpha^{(3)}_{c}}{w^{(4)}} \cdot B^{(3)}_{c},$	$w^{(4)} = \sum_{c=1}^{C} \alpha^{(3)}_{c}.$	

Panel B: Hierarchical credibility estimation of σ_1^2, σ_2^2, σ_3^2 and σ_4^2	
$\sigma_1^2 = E[\sigma_1^2(\Theta_{c,g,x})]$	$\sigma_1^2(\Theta_{c,g,x}) = w_{c,g,x,t} \cdot Var[Y_{c,g,x,t} \| \Theta_{c,g,x}]$
$\hat{\sigma}_1^2 = \frac{1}{C \cdot G \cdot X} \sum_{c=1}^{C} \sum_{g=1}^{G} \sum_{x=1}^{X} \hat{\sigma}_1^2(\Theta_{c,g,x})$	$\hat{\sigma}_1^2(\Theta_{c,g,x}) \triangleq \frac{1}{T-1} \sum_{t=1}^{T} w_{c,g,x,t} \cdot [Y_{c,g,x,t} - B^{(1)}_{c,g,x}]^2$
$\sigma_2^2 = E\{Var[\mu_1(\Theta_{c,g,x}) \| \Phi_{c,g}]\}$	$z^{(1)}_{c,g} = \sum_{x=1}^{X} w^{(1)}_{c,g,x}, \quad \overline{B}^{(1)}_{c,g,\bullet} = \sum_{x=1}^{X} \frac{w^{(1)}_{c,g,x}}{z^{(1)}_{c,g}} \cdot B^{(1)}_{c,g,x},$ $c^{(1)}_{c,g} = \frac{X-1}{X} \cdot \left\{ \sum_{x=1}^{X} \frac{w^{(1)}_{c,g,x}}{z^{(1)}_{c,g}} \cdot \left[1 - \frac{w^{(1)}_{c,g,x}}{z^{(1)}_{c,g}} \right] \right\}^{-1}$
$\hat{\sigma}_2^2 = \frac{1}{C \cdot G} \sum_{c=1}^{C} \sum_{g=1}^{G} \max[\hat{T}^{(1)}_{c,g}, 0]$	$\hat{T}^{(1)}_{c,g} = c^{(1)}_{c,g} \cdot \left\{ \frac{X}{X-1} \sum_{x=1}^{X} \frac{w^{(1)}_{c,g,x}}{z^{(1)}_{c,g}} \left[B^{(1)}_{c,g,x} - \overline{B}^{(1)}_{c,g,\bullet} \right]^2 - \frac{X}{z^{(1)}_{c,g}} \cdot \hat{\sigma}_1^2 \right\}$
$\sigma_3^2 = E\{Var[\mu_2(\Phi_{c,g}) \| \Psi_c]\}$	$z^{(2)}_{c} = \sum_{g=1}^{G} w^{(2)}_{c,g}, \quad \overline{B}^{(2)}_{c,\bullet} = \sum_{g=1}^{G} \frac{w^{(2)}_{c,g}}{z^{(2)}_{c}} \cdot B^{(2)}_{c,g},$ $c^{(2)}_{c} = \frac{G-1}{G} \cdot \left\{ \sum_{g=1}^{G} \frac{w^{(2)}_{c,g}}{z^{(2)}_{c}} \cdot \left[1 - \frac{w^{(2)}_{c,g}}{z^{(2)}_{c}} \right] \right\}^{-1}$
$\hat{\sigma}_3^2 = \frac{1}{C} \sum_{c=1}^{C} \max[\hat{T}^{(2)}_{c}, 0]$	$\hat{T}^{(2)}_{c} = c^{(2)}_{c} \cdot \left\{ \frac{G}{G-1} \sum_{g=1}^{G} \frac{w^{(2)}_{c,g}}{z^{(2)}_{c}} \left[B^{(2)}_{c,g} - \overline{B}^{(2)}_{c,\bullet} \right]^2 - \frac{G}{z^{(2)}_{c}} \cdot \hat{\sigma}_2^2 \right\}$
$\sigma_4^2 = Var[\mu_3(\Psi_c)]$	$z^{(3)} = \sum_{c=1}^{C} w^{(3)}_{c}, \quad \overline{B}^{(3)}_{\bullet} = \sum_{c=1}^{C} \frac{w^{(3)}_{c}}{z^{(3)}} \cdot B^{(3)}_{c},$ $c^{(3)} = \frac{C-1}{C} \cdot \left\{ \sum_{c=1}^{C} \frac{w^{(3)}_{c}}{z^{(3)}} \cdot \left[1 - \frac{w^{(3)}_{c}}{z^{(3)}} \right] \right\}^{-1}$
$\hat{\sigma}_4^2 = \max[\hat{T}^{(3)}, 0]$	$\hat{T}^{(3)} = c^{(3)} \cdot \left\{ \frac{C}{C-1} \sum_{c=1}^{C} \frac{w^{(3)}_{c}}{z^{(3)}} \left[B^{(3)}_{c} - \overline{B}^{(3)}_{\bullet} \right]^2 - \frac{C}{z^{(3)}} \cdot \hat{\sigma}_3^2 \right\}$

and $\alpha_c^{(3)}$), and $\hat{\mu}_4$ by theorem 6.4 of Bühlmann and Gisler (2005) are given in Panel A of Table 1.

Estimation of the structural parameters

Note that for a four-level tree structure, we can set $\alpha_c^{(3)} = 1$ and do not need to calculate $\hat{\mu}_4$ and $w^{(4)}$ in Panel A of Table 1. Similarly, to obtain the corresponding formula (2.1) for a three-level tree structure which is the Bühlmann-Straub credibility model, we let $\alpha_c^{(3)} = \alpha_{c,g}^{(2)} = 1$ and it is not necessary to calculate $B_c^{(3)}$, $w_c^{(3)}$, $\hat{\mu}_4$ and $w^{(4)}$.

To get the credibility estimate $\hat{Y}_{c,g,x,T+1}$ in (2.1), we need the estimates $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$. The estimation of σ_1^2, σ_2^2, σ_3^2 and σ_4^2 is given in Panel B of Table 1. For the detailed estimation, please refer to Section 6.6 of Bühlmann and Gisler (2005). From the formulas in Table 1, we notice that the values of $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$ can be zero, which leads to the values of $\hat{\alpha}_{c,g,x}^{(1)}$, $\hat{\alpha}_{c,g}^{(2)}$ and $\hat{\alpha}_c^{(3)}$ 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.

Next, we will give a special case where all of the known exposure units $w_{c,g,x,t}$s are set to 1. 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 Panel A of Table 1 simplify to

$$w_{c,g,x}^{(1)} = T \triangleq w^{(1)}, \quad B_{c,g,x}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} Y_{c,g,x,t} \triangleq \overline{Y}_{c,g,x,\bullet},$$

$$\hat{\alpha}_{c,g,x}^{(1)} = \frac{T \cdot \hat{\sigma}_2^2}{T \cdot \hat{\sigma}_2^2 + \hat{\sigma}_1^2} \triangleq \hat{\alpha}^{(1)}; \tag{2.2}$$

$$w_{c,g}^{(2)} = X \cdot \hat{\alpha}^{(1)}, \quad B_{c,g}^{(2)} = \frac{1}{X}\sum_{x=1}^{X} B_{c,g,x}^{(1)} = \frac{1}{X \cdot T}\sum_{x=1}^{X}\sum_{t=1}^{T} Y_{c,g,x,t} \triangleq \overline{Y}_{c,g,\bullet,\bullet},$$

$$\hat{\alpha}_{c,g}^{(2)} = \frac{X \cdot \hat{\alpha}^{(1)} \cdot \hat{\sigma}_3^2}{X \cdot \hat{\alpha}^{(1)} \cdot \hat{\sigma}_3^2 + \hat{\sigma}_2^2} \triangleq \hat{\alpha}^{(2)}; \tag{2.3}$$

$$w_c^{(3)} = G \cdot \hat{\alpha}^{(2)}, \quad B_c^{(3)} = \frac{1}{G}\sum_{g=1}^{G} B_{c,g}^{(2)} = \frac{1}{G \cdot X \cdot T}\sum_{g=1}^{G}\sum_{x=1}^{X}\sum_{t=1}^{T} Y_{c,g,x,t} \triangleq \overline{Y}_{c,\bullet,\bullet,\bullet},$$

$$\hat{\alpha}_c^{(3)} = \frac{G \cdot \hat{\alpha}^{(2)} \cdot \hat{\sigma}_4^2}{G \cdot \hat{\alpha}^{(2)} \cdot \hat{\sigma}_4^2 + \hat{\sigma}_3^2} \triangleq \hat{\alpha}^{(3)}, \tag{2.4}$$

$$w^{(4)} = C \cdot \hat{\alpha}^{(3)}, \; \hat{\mu}_4 = \frac{1}{C}\sum_{c=1}^{C} B_c^{(3)} = \frac{1}{C \cdot G \cdot X \cdot T}\sum_{c=1}^{C}\sum_{g=1}^{G}\sum_{x=1}^{X}\sum_{t=1}^{T} Y_{c,g,x,t} \triangleq \overline{Y}_{\bullet,\bullet,\bullet,\bullet}.$$

Also, the quantities simplify to, for estimation of

- σ_1^2: $B^{(1)}_{c,g,x} = \overline{Y}_{c,g,x,\bullet}$, and

$$\hat{\sigma}_1^2(\Theta_{c,g,x}) = \frac{1}{T-1}\sum_{t=1}^{T}(Y_{c,g,x,t} - \overline{Y}_{c,g,x,\bullet})^2;$$

- σ_2^2: $w^{(1)}_{c,g,x} = T$, $z^{(1)}_{c,g} = X \cdot T$, $\overline{B}^{(1)}_{c,g,\bullet} = \overline{Y}_{c,g,\bullet,\bullet}$, $c^{(1)}_{c,g} = 1$, and

$$\hat{T}^{(1)}_{c,g} = \frac{1}{X-1}\sum_{x=1}^{X}(\overline{Y}_{c,g,x,\bullet} - \overline{Y}_{c,g,\bullet,\bullet})^2 - \frac{\hat{\sigma}_1^2}{T};$$

- σ_3^2: $w^{(2)}_{c,g} = X \cdot \alpha^{(1)}$, $z^{(2)}_c = G \cdot X \cdot \alpha^{(1)}$, $B^{(2)}_{c,g} = \overline{Y}_{c,g,\bullet,\bullet}$, $\overline{B}^{(2)}_{c,\bullet} = \overline{Y}_{c,\bullet,\bullet,\bullet}$, $c^{(2)}_c = 1$, and

$$\hat{T}^{(2)}_c = \frac{1}{G-1}\sum_{g=1}^{G}(\overline{Y}_{c,g,\bullet,\bullet} - \overline{Y}_{c,\bullet,\bullet,\bullet})^2 - \left[\frac{\hat{\sigma}_2^2}{X} + \frac{\hat{\sigma}_1^2}{X \cdot T}\right];$$

- σ_4^2: $w^{(3)}_c = G \cdot \alpha^{(2)}$, $z^{(3)} = C \cdot G \cdot \alpha^{(2)}$, $B^{(3)}_c = \overline{Y}_{c,\bullet,\bullet,\bullet}$, $\overline{B}^{(3)}_{\bullet} = \overline{Y}_{\bullet,\bullet,\bullet,\bullet}$, $c^{(3)} = 1$, and

$$\hat{T}^{(3)} = \frac{1}{C-1}\sum_{c=1}^{C}(\overline{Y}_{c,\bullet,\bullet,\bullet} - \overline{Y}_{\bullet,\bullet,\bullet,\bullet})^2 - \left[\frac{\hat{\sigma}_3^2}{G} + \frac{\hat{\sigma}_2^2}{G \cdot X} + \frac{\hat{\sigma}_1^2}{G \cdot X \cdot T}\right].$$

The hierarchical credibility estimate of the decrement in the logarithm of central death rate for country c, gender g and age x over $[T,\, T+1]$ in (2.1) for this special case becomes

$$\begin{aligned} & \hat{Y}_{c,g,x,T+1} \\ = \ & \hat{\alpha}^{(1)} \cdot \overline{Y}_{c,g,x,\bullet} + [(1-\hat{\alpha}^{(1)}) \cdot \hat{\alpha}^{(2)}] \cdot \overline{Y}_{c,g,\bullet,\bullet} + [(1-\hat{\alpha}^{(1)}) \cdot (1-\hat{\alpha}^{(2)}) \cdot \hat{\alpha}^{(3)}] \cdot \overline{Y}_{c,\bullet,\bullet,\bullet} \\ & + [(1-\hat{\alpha}^{(1)}) \cdot (1-\hat{\alpha}^{(2)}) \cdot (1-\hat{\alpha}^{(3)})] \cdot \overline{Y}_{\bullet,\bullet,\bullet,\bullet}. \end{aligned} \tag{2.5}$$

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 $\{\ln(m_{c,g,x,t_L+t}) : t = 0, \cdots, T\}$, the **individual** time trend over $[0,\, T]$ for an individual of age x and gender g in country c);
- $\overline{Y}_{c,g,\bullet,\bullet} = \frac{1}{T}\sum_{t=1}^{T} \overline{Y}_{c,g,\bullet,t} = \frac{1}{T}\sum_{t=1}^{T}[\ln(m_{c,g,\bullet,t_L+t}) - \ln(m_{c,g,\bullet,t_L+t-1})]$ (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 **population** time trend over $[0,\, T]$ for a population of gender g and 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 trend 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 trend over $[0,\, T]$ for all C countries).

We use $(g,\,x,\,t)$ 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 $C=1$ and $\hat{\alpha}^{(3)}=1$, and we do not need to calculate $\hat{T}^{(3)}$ and $\hat{\sigma}_4^2$. Then the hierarchical credibility estimate of the decrement in the logarithm of central death rate for gender g and age x over $[T,\,T+1]$ under the special case is

$$\hat{Y}_{g,\,x,\,T+1}=\hat{\alpha}^{(1)}\cdot\overline{Y}_{g,\,x,\,\bullet}+[(1-\hat{\alpha}^{(1)})\cdot\hat{\alpha}^{(2)}]\cdot\overline{Y}_{g,\,\bullet,\,\bullet}+[(1-\hat{\alpha}^{(1)})\cdot(1-\hat{\alpha}^{(2)})]\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet},\tag{2.6}$$

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 the expressions for $\alpha^{(1)}$ and $\alpha^{(2)}$ in (2.6) are the same as (2.2) and (2.3), respectively, and $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ in $\alpha^{(1)}$ and $\alpha^{(2)}$ become

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

$\hat{\sigma}_2^2=\frac{1}{G}\sum_{g=1}^{G}\max[\hat{T}_g^{(1)},0]$ and $\hat{\sigma}_3^2=\max[\hat{T}^{(2)},0]$, where

$$\hat{T}_g^{(1)}=\frac{1}{X-1}\sum_{x=1}^{X}(\overline{Y}_{g,\,x,\,\bullet}-\overline{Y}_{g,\,\bullet,\,\bullet})^2-\frac{\hat{\sigma}_1^2}{T},$$

and

$$\hat{T}^{(2)}=\frac{1}{G-1}\sum_{g=1}^{G}(\overline{Y}_{g,\,\bullet,\,\bullet}-\overline{Y}_{\bullet,\,\bullet,\,\bullet})^2-\left[\frac{\hat{\sigma}_2^2}{X}+\frac{\hat{\sigma}_1^2}{X\cdot T}\right].$$

Similarly, the structural parameters for a three-level hierarchical tree can be estimated by setting $C=G=1$ and $\alpha^{(3)}=\alpha^{(2)}=1$. Then, we do not need to calculate $\hat{\sigma}_3^2$ and $\hat{\sigma}_4^2$. Therefore, the three-level hierarchical credibility mortality model is exactly the non-parametric Bühlmann credibility mortality model proposed by Tsai and Lin (2017b).

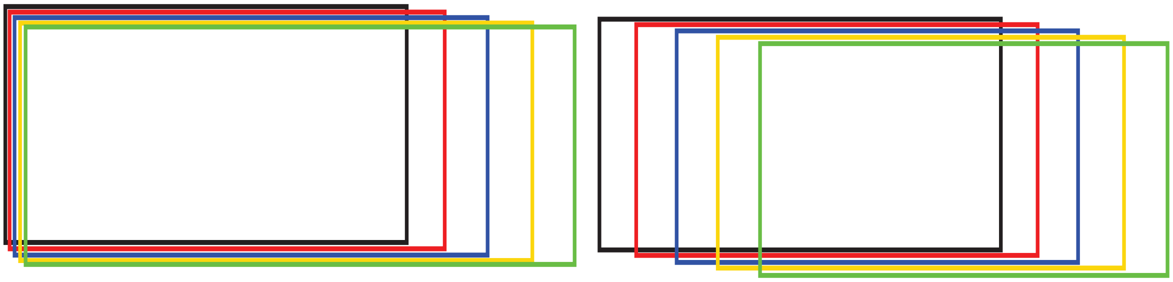
The expression in (2.5) gives the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+1}$ for country c, gender g and age x in year $T+1$. To obtain the hierarchical credibility estimate $\hat{Y}_{c,\,g,\,x,\,T+\tau}$ for year $T+\tau$ ($\tau\geq 2$), which is denoted by

$$\begin{aligned}\hat{Y}_{c,\,g,\,x,\,T+\tau} &= \hat{\alpha}_\tau^{(1)}\cdot\overline{Y}_{c,\,g,\,x,\,\bullet}^{T+\tau}+[(1-\hat{\alpha}_\tau^{(1)})\cdot\hat{\alpha}_\tau^{(2)}]\cdot\overline{Y}_{c,\,g,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad+[(1-\hat{\alpha}_\tau^{(1)})\cdot(1-\hat{\alpha}_\tau^{(2)})\cdot\hat{\alpha}_\tau^{(3)}]\cdot\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}^{T+\tau}\\ &\quad+[(1-\hat{\alpha}_\tau^{(1)})\cdot(1-\hat{\alpha}_\tau^{(2)})\cdot(1-\hat{\alpha}_\tau^{(3)})]\cdot\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}^{T+\tau},\end{aligned}\tag{2.7}$$

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

Figure 4: Expanding Window (EW) and Moving Window (MW)



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

The EW strategy expands the original fitting year span by $\tau - 1$ years to $[1,\ T+\tau-1]$ from $[1,\ T]$ by appending $\{\hat{Y}_{c,\,g,\,x,\,T+1}, \cdots, \hat{Y}_{c,\,g,\,x,\,T+\tau-1}\}$ to the end of $\{Y_{c,\,g,\,x,\,1}, \cdots, Y_{c,\,g,\,x,\,T}\}$ for all c, g and x. See Figure 4 (a).

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

$$\overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} = \frac{1}{T+\tau-1}\left[\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t}\right]. \tag{2.8}$$

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

$$\overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} = \frac{1}{X}\sum_{x=1}^{X} \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet}, \tag{2.9}$$

$$\overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet} = \frac{1}{G}\sum_{g=1}^{G} \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet}, \tag{2.10}$$

and

$$\overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet} = \frac{1}{C}\sum_{c=1}^{C} \overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet}. \tag{2.11}$$

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

$$\hat{\alpha}^{(1)}_{\tau} = \frac{(T+\tau-1)\cdot\hat{\sigma}^2_2}{(T+\tau-1)\cdot\hat{\sigma}^2_2 + \hat{\sigma}^2_1}. \tag{2.12}$$

The formulas for the credibility factors $\hat{\alpha}^{(2)}_{\tau}$ and $\hat{\alpha}^{(3)}_{\tau}$ are the same as those for $\tau = 1$ given in (2.3) and (2.4), which are

$$\hat{\alpha}^{(2)}_{\tau} = \frac{X\cdot\hat{\alpha}^{(1)}_{\tau}\cdot\hat{\sigma}^2_3}{X\cdot\hat{\alpha}^{(1)}_{\tau}\cdot\hat{\sigma}^2_3 + \hat{\sigma}^2_2} = \frac{X\,(T+\tau-1)\cdot\hat{\sigma}^2_3}{X\,(T+\tau-1)\cdot\hat{\sigma}^2_3 + (T+\tau-1)\cdot\hat{\sigma}^2_2 + \hat{\sigma}^2_1}, \tag{2.13}$$

and

$$\begin{aligned}\hat{\alpha}_{\tau}^{(3)} &= \frac{G\cdot\hat{\alpha}_{\tau}^{(2)}\cdot\hat{\sigma}_4^2}{G\cdot\hat{\alpha}_{\tau}^{(2)}\cdot\hat{\sigma}_4^2+\hat{\sigma}_3^2}\\ &= \frac{G\,X\,(T+\tau-1)\cdot\hat{\sigma}_4^2}{G\,X\,(T+\tau-1)\cdot\hat{\sigma}_4^2+X\,(T+\tau-1)\cdot\hat{\sigma}_3^2+(T+\tau-1)\cdot\hat{\sigma}_2^2+\hat{\sigma}_1^2}. \end{aligned} \tag{2.14}$$

Also note that the values of $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$, $\hat{\sigma}_4^2$ 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 (2.7).

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 adding 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 4 (b).

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

$$\overline{Y}^{T+\tau}_{c,g,x,\bullet}=\frac{1}{T}\sum_{t=\tau}^{T+\tau-1}\hat{Y}_{c,g,x,t}, \tag{2.15}$$

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 (2.9), (2.10) and (2.11), respectively.

Next, the credibility factor assigned to $\overline{Y}^{T+\tau}_{c,g,x,\bullet}$ is achieved by $\hat{\alpha}_{\tau}^{(1)}=\dfrac{T\cdot\hat{\sigma}_2^2}{T\cdot\hat{\sigma}_2^2+\hat{\sigma}_1^2}$. As $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$, $\hat{\sigma}_4^2$ are unchanged for all τ, we have $\hat{\alpha}_{\tau}^{(1)}=\hat{\alpha}_1^{(1)}$, $\hat{\alpha}_{\tau}^{(2)}=\hat{\alpha}_1^{(2)}$ and $\hat{\alpha}_{\tau}^{(3)}=\hat{\alpha}_1^{(3)}$. Therefore, $\hat{\alpha}_{\tau}^{(1)}$, $\hat{\alpha}_{\tau}^{(2)}$ and $\hat{\alpha}_{\tau}^{(3)}$ are constant in τ under the MW strategy.

Finally, we can calculate $\hat{Y}_{c,g,x,T+\tau}$, the decrement of the logarithm of central death rate for country c, gender g and age x over $[T+\tau-1,\,T+\tau]$, using (2.7).

There are some properties for the EW and MW strategies, which are stated in the following three propositions. The proofs are given in Appendix B.

Proposition 1 *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 $\overline{Y}^{T+\tau}_{c,g,x,\bullet}$ over the same age, gender and country spans for year $T+\tau$. Specifically,*

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

Proposition 2 *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.,*

$$\overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}=\overline{Y}^{T+1}_{\bullet,\bullet,\bullet,\bullet},\qquad \tau=2,3,\cdots.$$

Proposition 3 *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 3, we have

$$\ln(\hat{m}_{c,\,g,\,x,\,t_U+\tau}) = \ln(m_{c,\,g,\,x,\,t_U}) + \sum_{t=1}^{\tau} \hat{Y}_{c,\,g,\,x,\,T+t} = \ln(m_{c,\,g,\,x,\,t_U}) + (\hat{Y}_{c,\,g,\,x,\,T+1}) \cdot \tau.$$

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

3. Numerical illustrations

This section applies the hierarchical models of three/four/five-level and the estimated structural parameters introduced in Section 2 to forecasting mortality rates for six populations (both genders of the US, the UK and Japan) for numerical illustrations. The mortality data are obtained from the Human Mortality Database (HMD). We fit five-level, four-level, and three-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, joint-k, cointegrated and augmented common factor Lee-Carter models 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.

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 three forecasting year spans. 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 model, 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).

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 years $[t_U + 1,\, T_2]$, and evaluate the forecast performances of the underlying mortality models. Below are detailed assumptions.

Table 2: Summary of the fitting and forecasting year spans

Fitting year spans	$[t_L,\, t_U]$	[1951, 2003] [1952, 2003] ⋮ [1999, 2003]	[1951, 1993] [1952, 1993] ⋮ [1989, 1993]	[1951, 1983] [1952, 1983] ⋮ [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

- For the age span $[x_L,\, x_U]$, we choose $x_L = 20$ 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, $[1951,\, t_U], \cdots, [t_U - 4,\, t_U]$, are selected where t_U takes three values of 1983, 1993 and 2003, and the shortest period 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).

Table 2 gives a summary of the fitting and forecasting year spans.

We compare the forecasting performances of the hierarchical credibility models with the classical, joint-k, cointegrated and augmented common factor Lee-Carter models by the measure of mean absolute percentage error (MAPE), which is a common measurement as used in Lin et al. (2015) and Tsai and Lin (2017a, b). The MAPE is calculated by dividing the absolute value of the difference between the forecast mortality rate $\hat{q}_{c,\, g,\, x,\, t_U+\tau}$ and the true mortality rate $q_{c,\, g,\, x,\, t_U+\tau}$ by the true mortality rate $q_{c,\, g,\, x,\, t_U+\tau,}$. Specifically, the $MAPE^{[t_L,\, t_U]}_{c,\, g,\, x,\, t_U+\tau}$ for age x, gender g and country c in the forecasting year $t_U + \tau$ based on the fitting year span $[t_L,\, t_U]$ is defined by

$$MAPE^{[t_L,\, t_U]}_{c,\, g,\, x,\, t_U+\tau} = \left| \frac{\hat{q}_{c,\, g,\, x,\, t_U+\tau} - q_{c,\, g,\, x,\, t_U+\tau}}{q_{c,\, g,\, x,\, t_U+\tau}} \right|, \tag{3.1}$$

where $\hat{q}_{c,\, g,\, x,\, t_U+\tau} = 1 - e^{-\hat{m}_{c,\, g,\, x,\, t_U+\tau}}$ is based on the assumption of constant force of mortality over $[x,\, x+1] \times [t_U+\tau,\, t_U+\tau+1]$. The $MAPE^{[t_L,\, t_U]}_{c,\, g,\, x,\, t_U+\tau}$ above is calculated for a single forecast. Here we introduce the $AMAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]}$ to evaluate the forecasting performance for mortality rates in the window $[x_L,\, x_U] \times [t_U + 1,\, T_2]$, which is obtained by averaging the $MAPE^{[t_L,\, t_U]}_{c,\, g,\, x,\, t_U+\tau}$ over the window as

$$AMAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, T_2]} = \frac{1}{T_2 - t_U} \cdot \frac{1}{m} \sum_{\tau=1}^{T_2 - t_U} \sum_{x=x_L}^{x_U} MAPE^{[t_L,\, t_U]}_{c,\, g,\, x,\, t_U+\tau}.$$

The value of $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ largely depends on the fitting year span $[t_L,\,t_U]$; therefore, we evaluate the overall forecasting performance of a mortality model by averaging the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ over the fitting year spans $[t_L,\,t_U]$ for $t_L = T_1,\,T_1+1,\ldots,t_U-4$ to get the $AAMAPE_{c,\,g,\,[t_U+1,\,T_2]}$, which is computed as

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

A smaller $AAMAPE_{c,\,g,\,[t_U+1,\,T_2]}$ 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 paper will be ranked based on the $AAMAPE_{c,\,g,\,[t_U+1,\,T_2]}$.

We produce seven figures and construct three tables for three forecasting year spans [2004, 2013], [1994, 2013] and [1984, 2013]. Figures 5–10 show the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L where $t_U = 2003, 1993, 1983$ for each of six populations (US male, US female, UK male, UK female, Japan male and Japan female), respectively, and Figure 11 exhibits the average of the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ over the six populations against t_L where $t_U = 2003, 1993, 1983$. Within each figure, the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ plots for three different forecasting year spans are shown in three rows, and those for 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L, the start year of the fitting year span, for the classical Lee-Carter and the three-level hierarchical credibility models; the second column exhibits the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L for the five-level hierarchical credibility model and the three Lee-Carter models for six populations. Note that the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ under the three Lee-Carter models for two populations and six populations are different. For the two-population graphs, the joint-k, cointegrated and 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ are given in (b), (e) and (h) of Figures 5 and 6, respectively. For the six-population graphs, the joint-k, cointegrated and augmented common factor Lee-Carter models are fitted to all of six populations, and the corresponding $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ for the US male and female are given in (c), (f) and (i) of Figures 5 and 6, respectively.

Observations from the figures are summarized below.

- The $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ value.
- The $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values depend on the length and location of the fitting year span. The pattern of $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curves largely depends on the data set. For example, the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curve for the MW strategy in Figure 6

Table 3: $AAMAPE_{c,\,g,\,[2004,2013]}$s (%)

AAMAPE	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 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 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

(h) increases for the first 20 t_L values, then decreases for the next five t_L values, and finally increases in the last few t_L values. However, Figure 7 (h) shows the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ 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 $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$, we calculate the $AAMAPE_{c,\,g,\,[t_U+1,\,T_2]}$ (the average of the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ over all values of t_L) and use it to rank the underlying mortality models.

- The $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ 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 5 (i) for the US male, Figure 6 (i) for the US female, and Figure 9 (h) for the Japan male; the corresponding $AAMAPE_{c,\,g,\,[t_U+1,\,T_2]}$ values are shown in Table 5. Moreover, the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ curves for the MW strategy look smoother in t_L than the EW one.

Below are observations from Tables 3–5, which exhibit the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$, the "Avg 2" (the average of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over both genders of a country) and the "Avg 6" (the average of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations) for $t_U = 2003$, 1993 and 1983, respectively.

- Observing the values in Tables 3–5, it is obvious that the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ 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 5–10 that the $AMAPE^{[t_L,\,t_U]}_{c,\,g,\,[t_U+1,\,T_2]}$ values increase in the width of the forecasting year span.
- Based on the average of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations ("Avg 6"), we observe that the hierarchical credibility model with more levels provides

Table 4: $AAMAPE_{c,\,g,\,[1994,\,2013]}$s (%)

AAMAPE	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 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 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 5: $AAMAPE_{c,\,g,\,[1984,\,2013]}$s (%)

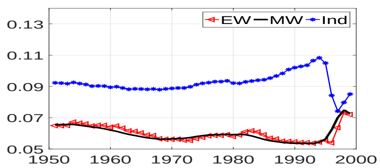
AAMAPE	Country	US			UK			Japan		
Model	**Avg 6**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**	**M**	**F**	**Avg 2**
	Hierarchical Credibility (HC) model									
EW-5	14.01	13.57	17.33	15.45	16.94	10.63	13.78	10.69	14.90	12.79
MW-5	14.02	13.48	18.26	15.87	16.44	10.60	13.52	10.85	14.50	12.67
EW-4	14.60	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 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 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

more accurate forecasting performance. However, this conclusion does not apply to "Avg 2" (the average of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ 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 5 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.

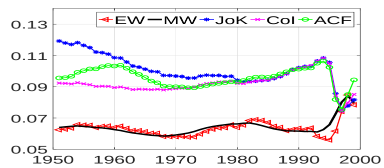
- The $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ values and their averages "Avg 6" and "Avg 2" under both of the EW and MW strategies, given the same level of the hierarchical credibility model, are close to each other.
- Among the Lee-Carter models applied to six populations, the averages of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations ("Avg 6") show 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 3–5, we observe that most of the $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ 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 $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over both genders of a country and over six populations for all of 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 $AAMAPE_{c,\,g,\,[t_U+1,\,2013]}$ over six populations for the joint-k, cointegrated and augmented common factor Lee-Carter models applied to six populations and two populations and for the classical Lee-Carter model are 19.57%, 17.97%, 17.26%, 18.40%, 17.92%, 18.41% and 18.25%, respectively, whereas those for the EW and MW strategies under the five/four/three-level hierarchical credibility models are 14.01%, 14.02%, 14.60%, 14.28%, 15.03% and 14.55%, respectively. Therefore, the numerical illustrations highly support the conclusion that the hierarchical credibility models outperform the Lee-Carter models.

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.

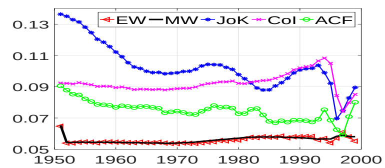
Figure 5: $AMAPE^{[t_L,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L for US Male with age span 20–84



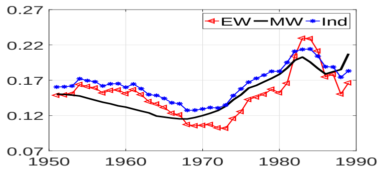
(a) $t_U = 2003$, single population



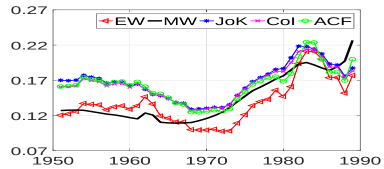
(b) $t_U = 2003$, two populations



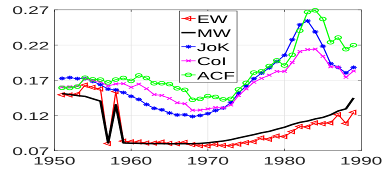
(c) $t_U = 2003$, six populations



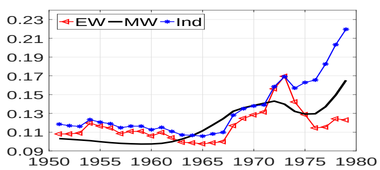
(d) $t_U = 1993$, single population



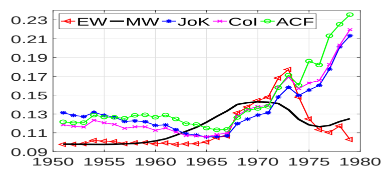
(e) $t_U = 1993$, two populations



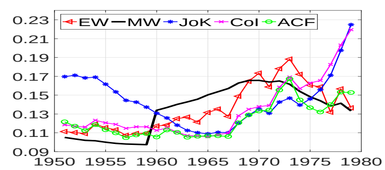
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

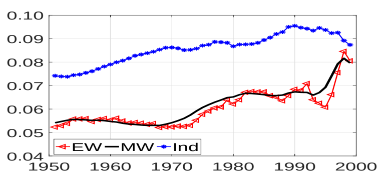


(h) $t_U = 1983$, two populations

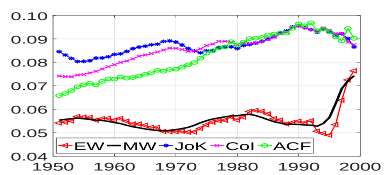


(i) $t_U = 1983$, six populations

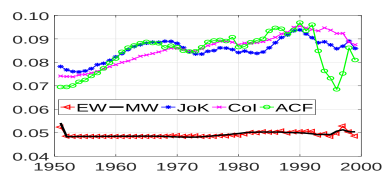
Figure 6: $AMAPE^{[t_L,t_U]}_{c,\,g,\,[t_U+1,\,2013]}$ against t_L for US Female with age span 20–84



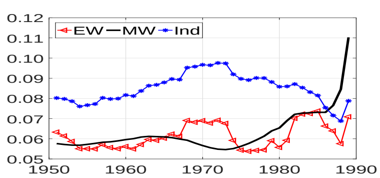
(a) $t_U = 2003$, single population



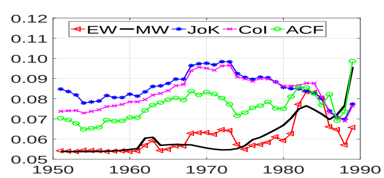
(b) $t_U = 2003$, two populations



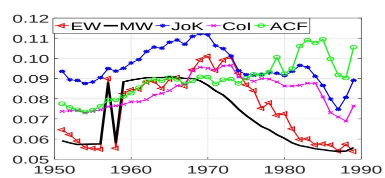
(c) $t_U = 2003$, six populations



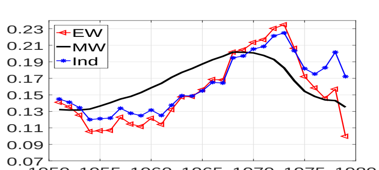
(d) $t_U = 1993$, single population



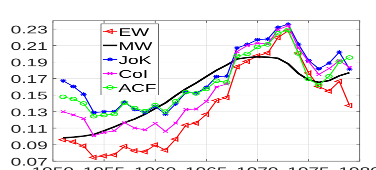
(e) $t_U = 1993$, two populations



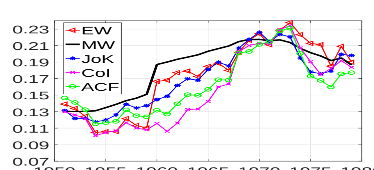
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

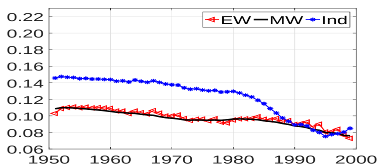


(h) $t_U = 1983$, two populations

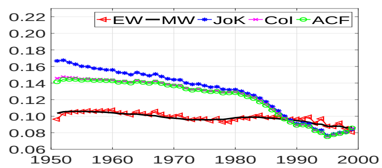


(i) $t_U = 1983$, six populations

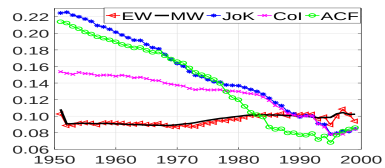
Figure 7: $AMAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for UK Male with age span 20–84



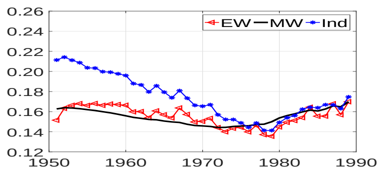
(a) $t_U = 2003$, single population



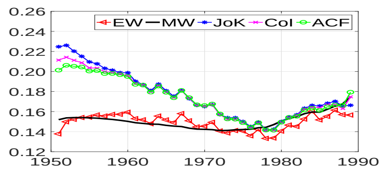
(b) $t_U = 2003$, two populations



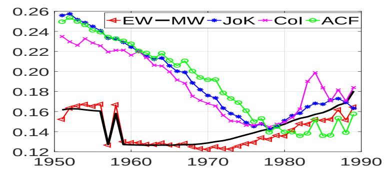
(c) $t_U = 2003$, six populations



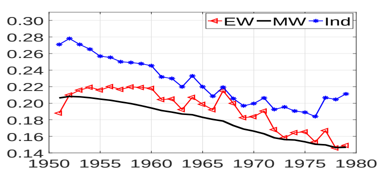
(d) $t_U = 1993$, single population



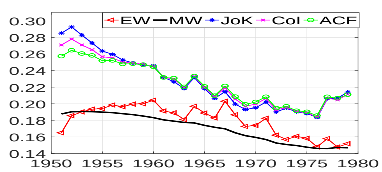
(e) $t_U = 1993$, two populations



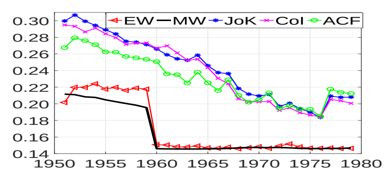
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

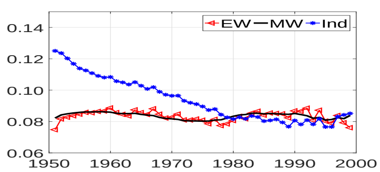


(h) $t_U = 1983$, two populations

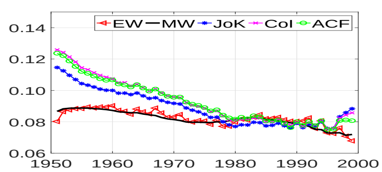


(i) $t_U = 1983$, six populations

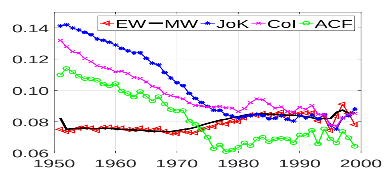
Figure 8: $AMAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for UK Female with age span 20–84



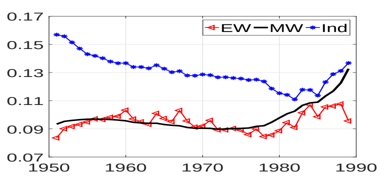
(a) $t_U = 2003$, single population



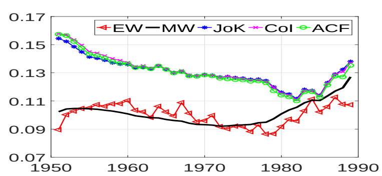
(b) $t_U = 2003$, two populations



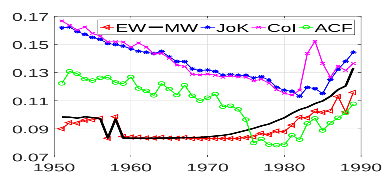
(c) $t_U = 2003$, six populations



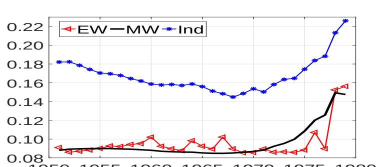
(d) $t_U = 1993$, single population



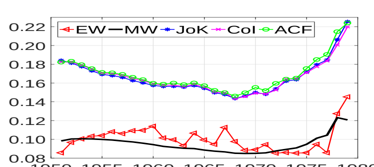
(e) $t_U = 1993$, two populations



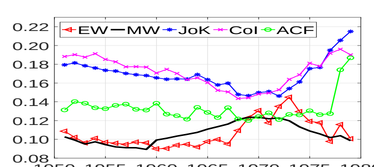
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

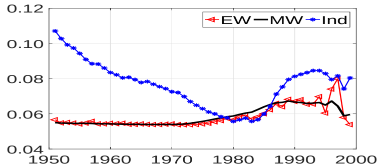


(h) $t_U = 1983$, two populations

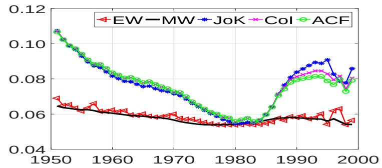


(i) $t_U = 1983$, six populations

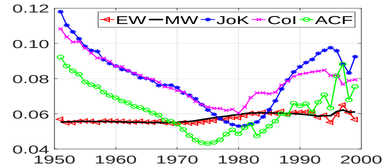
Figure 9: $AMAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for Japan Male with age span 20–84



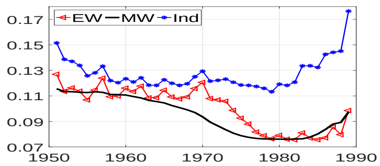
(a) $t_U = 2003$, single population



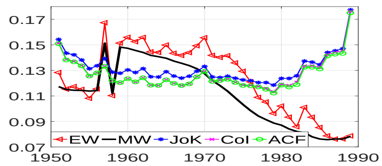
(b) $t_U = 2003$, two populations



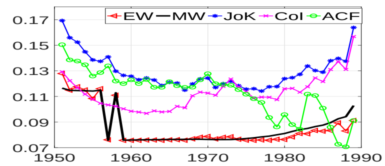
(c) $t_U = 2003$, six populations



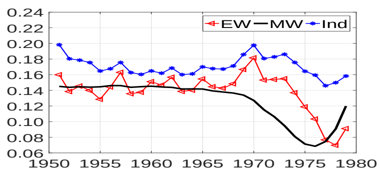
(d) $t_U = 1993$, single population



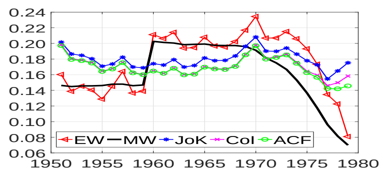
(e) $t_U = 1993$, two populations



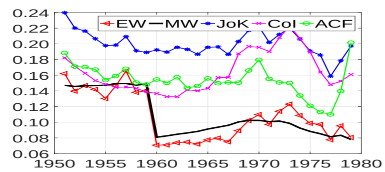
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

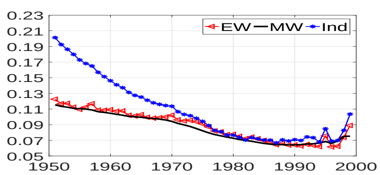


(h) $t_U = 1983$, two populations

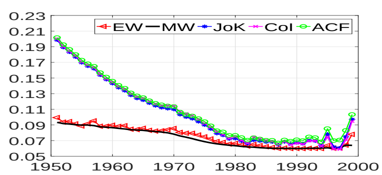


(i) $t_U = 1983$, six populations

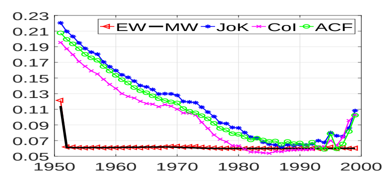
Figure 10: $AMAPE^{[t_L, t_U]}_{c, g, [t_U+1, 2013]}$ against t_L for Japan Female with age span 20–84



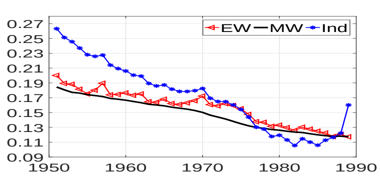
(a) $t_U = 2003$, single population



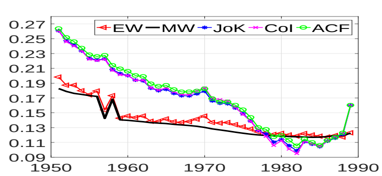
(b) $t_U = 2003$, two populations



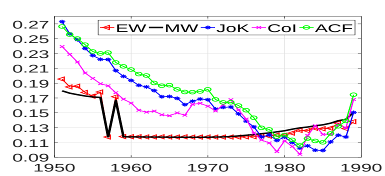
(c) $t_U = 2003$, six populations



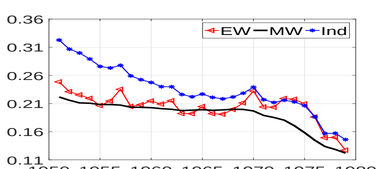
(d) $t_U = 1993$, single population



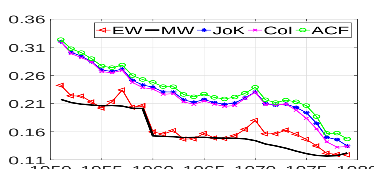
(e) $t_U = 1993$, two populations



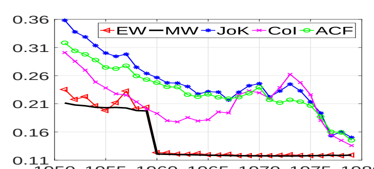
(f) $t_U = 1993$, six populations



(g) $t_U = 1983$, single population

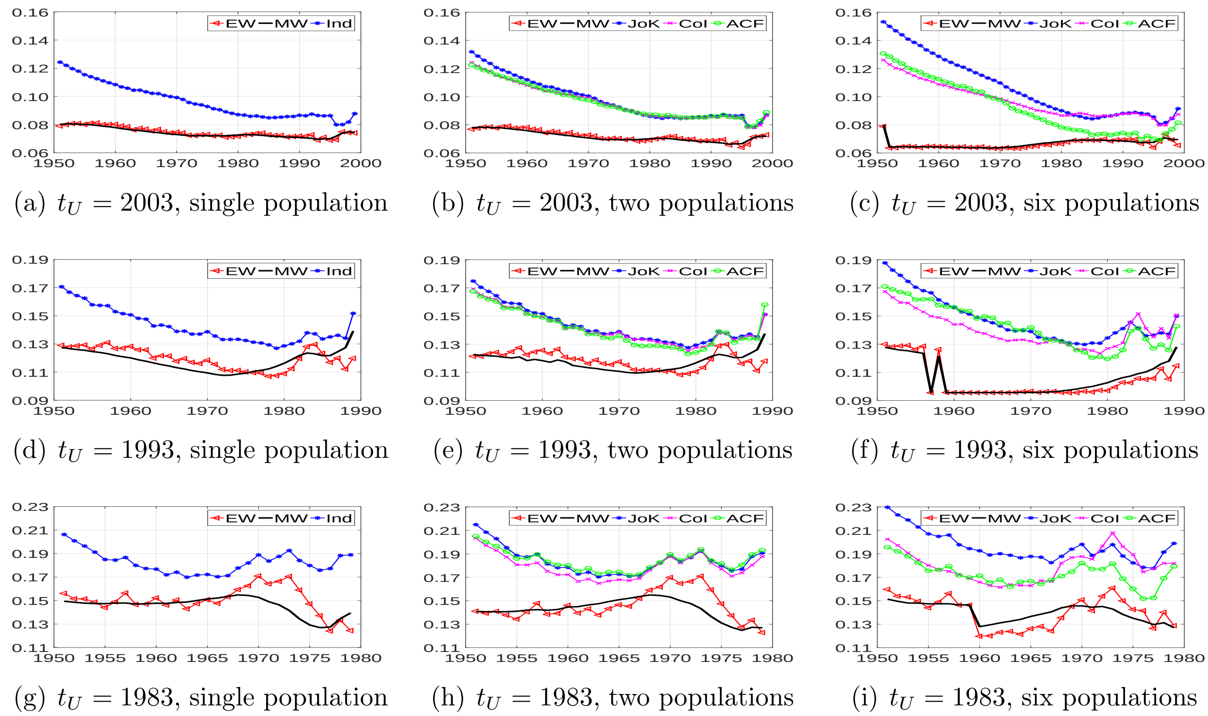


(h) $t_U = 1983$, two populations



(i) $t_U = 1983$, six populations

Figure 11: The average of the $AAMAPE^{[t_L,\, t_U]}_{c,\, g,\, [t_U+1,\, 2013]}$ over 6 populations against t_L 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

4. Conclusions

Longevity risk is a critical issue for annuity providers, social security systems and defined benefit pension plans. This paper 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 in the logarithm of central death rate for country c, gender g and age x over $[T,\, T+1]$, 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 the credibility-factor-weighted average of the average annual decrements of the four time trends $\overline{Y}_{c,\,g,\,x,\,\bullet}$, $\overline{Y}_{c,\,g,\,\bullet,\,\bullet}$, $\overline{Y}_{c,\,\bullet,\,\bullet,\,\bullet}$ and $\overline{Y}_{\bullet,\,\bullet,\,\bullet,\,\bullet}$. 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 $\hat{Y}_{c,g,x,T+\tau}$ is constant for $\tau = 1, 2, \cdots$. Thus, $\ln(\hat{m}_{c,g,x,t_U+\tau})$ is a linear function of τ with slope $\hat{Y}_{c,g,x,T+1}$ and intercept $\ln(m_{c,g,x,t_U})$.

The forecasting performance is measured by the AMAPE for a single fitting year span. Based on the figures displayed in Section 3, we conclude that models have larger AMAPEs as the forecasting year span gets wider. Since we do not know which fitting year span will produce the lowest AMAPE, and actually the AMAPE varies largely in the fitting year span, we rank the models according to AAMAPE, the average of AMAPEs over all of the fitting year spans. From Tables 3–5, we conclude that the hierarchical credibility model with more levels overall yields better prediction, and 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 the dependency of 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 or reduced to a tree structure of four or three levels. 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 with dependency can be further used to construct a mortality index for more accurately pricing mortality-indexed securities.

Appendix A: Multi-population Lee-Carter models

A.1 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 $k_{t,i} = K_t$, $i = 1, \cdots, r$. The logarithm of central death rates, $\ln(m_{x,t,i})$, for lives aged x in year t and population i can be expressed as

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

where $\alpha_{x,i}$ is the average age-specific mortality factor at age x for population i, $\beta_{x,i}$ 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 $\varepsilon_{x,t,i}$, $t = t_L, \cdots, t_U$, 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 $\sum_{i=1}^{r} \sum_{x=x_L}^{x_U} \beta_{x,i} = 1$ 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,

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

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

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

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

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 the time trend errors, $\{\epsilon_t\}$, are assumed to be independent of the model errors, $\{\varepsilon_{x,t,i}\}$. Then we can estimate the drift parameter θ with

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

where $n = t_U - t_L + 1$. The logarithm of the projected central death rate for age x in year $t_U + \tau$ and population i is given by

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

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

A.2 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 classical Lee-Carter model as follows:

$$\ln(m_{x,t,i}) = \alpha_{x,i} + \beta_{x,i} \times k_{t,i} + \varepsilon_{x,t,i}, \; x = x_L, \cdots, x_U, t = t_L, \cdots, t_U, \; i = 1, \cdots, r. \quad \text{(A.3)}$$

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

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

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

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

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

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

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

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

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

$$\ln(\hat{m}_{x,t_U+\tau,i}) = \hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \times (\hat{k}_{t_U,i} + \tau \times \hat{\hat{\theta}}_i) = \ln(\hat{m}_{x,t_U,i}) + (\hat{\beta}_{x,i} \times \hat{\hat{\theta}}_i) \times \tau, \ \tau = 1, \cdots, \quad \text{(A.4)}$$

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

A.3 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 B_x and a uniform time-varying index K_t for all populations as follows:

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

subject to two constraints, $\sum_{t=t_L}^{t_U} K_t = 0$ and $\sum_{i=1}^{r} \sum_{x=x_L}^{x_U} w_i \, B_x = 1$, 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 α_x by

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

and K_t as

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

Then $\hat{B}_x$ can be similarly obtained by regressing $\sum_{i=1}^{r} w_i \times [\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ 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 $\beta'_{x,i} \times k'_{t,i}$ to the common factor model to get

$$\ln(m_{x,t,i}) = \alpha_{x,i} + B_x \times K_t + \beta'_{x,i} \times k'_{t,i} + \varepsilon_{x,t,i}, \tag{A.5}$$

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

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

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

$$\begin{aligned} \ln(\hat{m}_{x,t_U+\tau,i}) &= \hat{\alpha}_{x,i} + \hat{B}_x \times (\hat{K}_{t_U} + \tau \times \hat{\theta}) + \hat{\beta}'_{x,i} \times (\hat{k}'_{t_U,i} + \tau \times \hat{\theta}'_i) \\ &= \ln(\hat{m}_{x,t_U,i}) + (\hat{B}_x \times \hat{\theta} + \hat{\beta}'_{x,i} \times \hat{\theta}'_i) \times \tau, \quad \tau = 1, \cdots, \end{aligned} \tag{A.6}$$

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

Appendix B: Proofs of Propositions 1–3

B.1 Proof of Proposition 1

Proof: By (2.7),

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

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

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

□

B.2 Proof of Proposition 2

Proof: This proposition is proven by mathematical induction on τ. First, for $\tau = 2$, by definition, (2.8) 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 (2.8) and Proposition 1 lead to

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

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. □

B.3 Proof of Proposition 3

Proof: Let $f_\tau = (T+\tau) \cdot \hat{\sigma}_2^2 + \hat{\sigma}_1^2$, $g_\tau = X \cdot (T+\tau) \cdot \hat{\sigma}_3^2 + f_\tau$ and $h_\tau = G \cdot X \cdot (T+\tau) \cdot \hat{\sigma}_4^2 + g_\tau$. Then by (2.12), (2.13) and (2.14), we have

$$\begin{aligned}
&\hat{\alpha}_\tau^{(1)} = \tfrac{f_{\tau-1} - \hat{\sigma}_1^2}{f_{\tau-1}}, \quad 1 - \hat{\alpha}_\tau^{(1)} = \tfrac{\hat{\sigma}_1^2}{f_{\tau-1}}, \quad (1 - \hat{\alpha}_\tau^{(1)})\, \hat{\alpha}_\tau^{(2)} = \tfrac{\hat{\sigma}_1^2 (g_{\tau-1} - f_{\tau-1})}{f_{\tau-1} \cdot g_{\tau-1}}, \\
&\hat{\alpha}_\tau^{(2)} = \tfrac{g_{\tau-1} - f_{\tau-1}}{g_{\tau-1}}, \quad 1 - \hat{\alpha}_\tau^{(2)} = \tfrac{f_{\tau-1}}{g_{\tau-1}}, \quad (1 - \hat{\alpha}_\tau^{(1)})\,(1 - \hat{\alpha}_\tau^{(2)})\, \hat{\alpha}_2^{(3)} = \tfrac{\hat{\sigma}_1^2 (h_{\tau-1} - g_{\tau-1})}{g_{\tau-1} \cdot h_{\tau-1}}, \\
&\hat{\alpha}_\tau^{(3)} = \tfrac{h_{\tau-1} - g_{\tau-1}}{h_{\tau-1}}, \quad 1 - \hat{\alpha}_\tau^{(3)} = \tfrac{g_{\tau-1}}{h_{\tau-1}}, \quad (1 - \hat{\alpha}_\tau^{(1)})\,(1 - \hat{\alpha}_\tau^{(2)})\,(1 - \hat{\alpha}_2^{(3)}) = \tfrac{\hat{\sigma}_1^2}{h_{\tau-1}}.
\end{aligned}$$

From (2.7), we can express $\hat{Y}_{c,\,g,\,x,\,T+\tau}$ as

$$
\begin{aligned}
&\hat{Y}_{c,\,g,\,x,\,T+\tau} \\
= \;& \frac{1}{f_{\tau-1}\, g_{\tau-1}\, h_{\tau-1}} \Big[g_{\tau-1}\, h_{\tau-1}\, (f_{\tau-1} - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} + \hat{\sigma}_1^2\, (g_{\tau-1} - f_{\tau-1})\, h_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} \\
& + \hat{\sigma}_1^2\, (h_{\tau-1} - g_{\tau-1})\, f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\,\bullet,\,\bullet,\,\bullet} + \hat{\sigma}_1^2\, f_{\tau-1}\, g_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\,\bullet,\,\bullet,\,\bullet} \Big]. \qquad \text{(B.1)}
\end{aligned}
$$

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

$$
\begin{aligned}
& g_\tau\, h_\tau\, (f_\tau - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau+1}_{c,\,g,\,x,\,\bullet} - g_{\tau-1}\, h_{\tau-1}\, (f_{\tau-1} - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} \\
&= \hat{\sigma}_2^2 \Bigg[g_\tau h_\tau \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) - g_{\tau-1} h_{\tau-1} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) \Bigg] \\
&= \hat{\sigma}_2^2 \Bigg[(g_\tau\, h_\tau - g_{\tau-1}\, h_{\tau-1}) \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) + g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau} \Bigg] \\
&= (g_\tau\, h_\tau - g_{\tau-1}\, h_{\tau-1})\, (f_{\tau-1} - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,x,\,\bullet} + \hat{\sigma}_2^2\, g_\tau\, h_\tau \cdot \hat{Y}_{c,\,g,\,x,\,T+\tau}, \qquad \text{(B.2)}
\end{aligned}
$$

$$
\begin{aligned}
& \hat{\sigma}_1^2 (g_\tau - f_\tau)\, h_\tau \cdot \overline{Y}^{T+\tau+1}_{c,\,g,\,\bullet,\,\bullet} - \hat{\sigma}_1^2\, (g_{\tau-1} - f_{\tau-1})\, h_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} \\
&= \hat{\sigma}_1^2\, \hat{\sigma}_3^2 \Bigg[h_\tau \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) - h_{\tau-1} \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) \Bigg] \\
&= \hat{\sigma}_1^2 \hat{\sigma}_3^2 \Bigg[(h_\tau - h_{\tau-1}) \sum_{x=1}^{X} \Bigg(\sum_{t=1}^{T} Y_{c,\,g,\,x,\,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,\,g,\,x,\,t} \Bigg) + h_\tau \sum_{x=1}^{X} \hat{Y}_{c,\,g,\,x,\,T+\tau} \Bigg] \\
&= \hat{\sigma}_1^2\, (h_\tau - h_{\tau-1})\, (g_{\tau-1} - f_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,\,g,\,\bullet,\,\bullet} + \hat{\sigma}_1^2\, \hat{\sigma}_3^2\, h_\tau \sum_{x=1}^{X} \hat{Y}_{c,\,g,\,x,\,T+\tau}, \qquad \text{(B.3)}
\end{aligned}
$$

$$
\begin{aligned}
&\hat{\sigma}_1^2\,(h_\tau - g_\tau)\,f_\tau \cdot \overline{Y}^{T+\tau+1}_{c,\bullet,\bullet,\bullet} - \hat{\sigma}_1^2\,(h_{\tau-1} - g_{\tau-1})\,f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} \\
&= \hat{\sigma}_1^2\,\hat{\sigma}_4^2 \Bigg[f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau} \hat{Y}_{c,g,x,t}\Bigg) \\
&\quad - f_{\tau-1} \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,g,x,t}\Bigg)\Bigg] \\
&= \hat{\sigma}_1^2\,\hat{\sigma}_4^2 \Bigg[(f_\tau - f_{\tau-1}) \sum_{g=1}^{G}\sum_{x=1}^{X}\Bigg(\sum_{t=1}^{T} Y_{c,g,x,t} + \sum_{t=T+1}^{T+\tau-1} \hat{Y}_{c,g,x,t}\Bigg) + f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X} \hat{Y}_{c,g,x,T+\tau}\Bigg] \\
&= \hat{\sigma}_1^2\,(f_\tau - f_{\tau-1})\,(h_{\tau-1} - g_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} + \hat{\sigma}_1^2\,\hat{\sigma}_4^2 f_\tau \sum_{g=1}^{G}\sum_{x=1}^{X} \hat{Y}_{c,g,x,T+\tau},
\end{aligned} \tag{B.4}
$$

$$
\hat{\sigma}_1^2\, f_\tau g_\tau \cdot \overline{Y}^{T+\tau+1}_{\bullet,\bullet,\bullet,\bullet} - \hat{\sigma}_1^2\, f_{\tau-1}\, g_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet} = \hat{\sigma}_1^2\,(f_\tau\, g_\tau - f_{\tau-1}\, g_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}. \tag{B.5}
$$

Next,

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

and

$$
(f_\tau g_\tau h_\tau - f_{\tau-1} g_{\tau-1} h_{\tau-1}) \cdot \hat{Y}_{c,g,x,T+\tau} = \hat{\sigma}_2^2 g_\tau h_\tau \cdot \hat{Y}_{c,g,x,T+\tau} + f_{\tau-1}(g_\tau h_\tau - g_{\tau-1} h_{\tau-1}) \cdot \hat{Y}_{c,g,x,T+\tau}. \tag{B.6}
$$

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

$$
\begin{aligned}
&(g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(f_{\tau-1} - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau}_{c,g,x,\bullet} + \frac{\hat{\sigma}_1^2}{g_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(g_{\tau-1} - f_{\tau-1}) \cdot \overline{Y}^{T+\tau}_{c,g,\bullet,\bullet} \\
&+ \frac{\hat{\sigma}_1^2}{g_{\tau-1} h_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1})(h_{\tau-1} - g_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} \\
&+ \frac{\hat{\sigma}_1^2}{h_{\tau-1}} (g_\tau h_\tau - g_{\tau-1} h_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}.
\end{aligned} \tag{B.7}
$$

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

$$
\begin{aligned}
\sum_{x=1}^{X} \hat{Y}_{c,g,x,T+\tau} =& \frac{X}{f_{\tau-1} g_{\tau-1} h_{\tau-1}} \Big\{ \Big[g_{\tau-1} h_{\tau-1} (f_{\tau-1} - \hat{\sigma}_1^2) + \hat{\sigma}_1^2 (g_{\tau-1} - f_{\tau-1}) h_{\tau-1} \Big] \cdot \overline{Y}^{T+\tau}_{c,g,\bullet,\bullet} \\
&\quad + \hat{\sigma}_1^2 (h_{\tau-1} - g_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} + \hat{\sigma}_1^2 f_{\tau-1} g_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet} \Big\} \\
=& \frac{X}{f_{\tau-1} g_{\tau-1} h_{\tau-1}} \Big[f_{\tau-1} h_{\tau-1} (g_{\tau-1} - \hat{\sigma}_1^2) \cdot \overline{Y}^{T+\tau}_{c,g,\bullet,\bullet} \\
&\quad + \hat{\sigma}_1^2 (h_{\tau-1} - g_{\tau-1}) f_{\tau-1} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} + \hat{\sigma}_1^2 f_{\tau-1} g_{\tau-1} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet} \Big] \\
=& \frac{X (g_{\tau-1} - \hat{\sigma}_1^2)}{g_{\tau-1}} \cdot \overline{Y}^{T+\tau}_{c,g,\bullet,\bullet} + \frac{X \hat{\sigma}_1^2 (h_{\tau-1} - g_{\tau-1})}{g_{\tau-1} h_{\tau-1}} \cdot \overline{Y}^{T+\tau}_{c,\bullet,\bullet,\bullet} + \frac{X \hat{\sigma}_1^2}{h_{\tau-1}} \cdot \overline{Y}^{T+\tau}_{\bullet,\bullet,\bullet,\bullet}
\end{aligned} \tag{B.8}
$$

and

$$\begin{aligned}
\sum_{g=1}^{G}\sum_{x=1}^{X}\hat{Y}_{c,g,x,T+\tau} =& \frac{GX}{f_{\tau-1}g_{\tau-1}h_{\tau-1}}\Big\{\Big[f_{\tau-1}h_{\tau-1}(g_{\tau-1}-\hat{\sigma}_1^2)+\hat{\sigma}_1^2(h_{\tau-1}-g_{\tau-1})f_{\tau-1}\Big]\cdot\overline{Y}_{c,\bullet,\bullet,\bullet}^{T+\tau} \\
&+\hat{\sigma}_1^2 f_{\tau-1}g_{\tau-1}\cdot\overline{Y}_{\bullet,\bullet,\bullet,\bullet}^{T+\tau}\Big\} \\
=& \frac{GX(h_{\tau-1}-\hat{\sigma}_1^2)}{h_{\tau-1}}\cdot\overline{Y}_{c,\bullet,\bullet,\bullet}^{T+\tau}+\frac{GX\hat{\sigma}_1^2}{h_{\tau-1}}\cdot\overline{Y}_{\bullet,\bullet,\bullet,\bullet}^{T+\tau},
\end{aligned} \tag{B.9}$$

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

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

$$\frac{\hat{\sigma}_1^2}{g_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})(g_{\tau-1}-f_{\tau-1})=\hat{\sigma}_1^2(h_\tau-h_{\tau-1})(g_{\tau-1}-f_{\tau-1})+\frac{X\hat{\sigma}_1^2\hat{\sigma}_3^2h_\tau(g_{\tau-1}-\hat{\sigma}_1^2)}{g_{\tau-1}}; \tag{B.10}$$

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

$$\begin{aligned}
&\frac{\hat{\sigma}_1^2}{g_{\tau-1}\,h_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})(h_{\tau-1}-g_{\tau-1})f_{\tau-1} \\
=\;&\frac{X\,\hat{\sigma}_1^4\,\hat{\sigma}_3^2\,h_\tau(h_{\tau-1}-g_{\tau-1})}{g_{\tau-1}\,h_{\tau-1}}+\hat{\sigma}_1^2\,(f_\tau-f_{\tau-1})(h_{\tau-1}-g_{\tau-1})+\frac{G\,X\,\hat{\sigma}_1^2\,\hat{\sigma}_4^2\,f_\tau(h_{\tau-1}-\hat{\sigma}_1^2)}{h_{\tau-1}};
\end{aligned} \tag{B.11}$$

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

$$\frac{\hat{\sigma}_1^2}{h_{\tau-1}}(g_\tau h_\tau-g_{\tau-1}h_{\tau-1})f_{\tau-1}=\frac{X\,\hat{\sigma}_1^4\,\hat{\sigma}_3^2\,h_\tau}{h_{\tau-1}}+\frac{G\,X\,\hat{\sigma}_1^4\,\hat{\sigma}_4^2\,f_\tau}{h_{\tau-1}}+\hat{\sigma}_1^2\,(f_\tau\,g_\tau-f_{\tau-1}\,g_{\tau-1}). \tag{B.12}$$

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

$$(g_{\tau-1}-f_{\tau-1})(g_\tau-g_{\tau-1})=X\,\hat{\sigma}_3^2\,(g_{\tau-1}-\hat{\sigma}_1^2),$$

$$(h_{\tau-1}-g_{\tau-1})(f_{\tau-1}\,g_\tau\,h_\tau-f_\tau\,g_{\tau-1}\,h_{\tau-1}-X\,\hat{\sigma}_1^2\,\hat{\sigma}_3^2\,h_\tau)=G\,X\,\hat{\sigma}_4^2\,f_\tau\,g_{\tau-1}\,(h_{\tau-1}-\hat{\sigma}_1^2),$$

$$g_\tau\,(f_{\tau-1}\,h_\tau-f_\tau\,h_{\tau-1})=X\,\hat{\sigma}_1^2\,(\hat{\sigma}_3^2\,h_\tau+G\,\hat{\sigma}_4^2\,f_\tau),$$

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

References

Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin* 4, 199–207.

Bühlmann, H. and Gisler, A. (2005). *A Course in credibility theory and its applications.* Springer, Netherlands.

Bühlmann, H. and Straub, E. (1970). Glaubwrdigkeit frschadenstze. *Bulletin of the Swiss Association of Actuaries* 70, 111–133.

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

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* 13 (1), 1–35.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D. and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *Astin Bulletin* 41 (1), 29–59.

Carter, L.R. and Lee, R.D. (1992). Forecasting demographic components: Modeling and forecasting U.S. sex differentials in mortality. *International Journal of Forecasting* 8 (3), 393–411.

Chen, H., MacMinn, R. and Sun, T. (2015). Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics* 63, 135–146.

Enchev, V., Kleinow, T. and Cairns, A.J.G. (2016). Multi-population mortality models: Fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, 1–24.

Lee, R.D. and Carter, L.R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87, 659–671.

Li, J.S.H., Chan, W.S. and Zhou, R. (2017). Semicoherent multipopulation mortality modelling: The impact on longevity risk securitization. *Journal of Risk and Insurance* 84, 1025–1065.

Li, J.S.H. and Hardy, M.R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal* 15 (2), 177–200.

Li, J.S.H., Hardy, M.R. and Tan, K.S. (2009). Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. *Astin Bulletin* 39 (1), 137–164.

Li, J.S.H., Zhou, R. and Hardy, M. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics* 63, 121–134.

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

Lin, T., Wang, C.W. and Tsai, C.C.L. (2015). Age-specific copula-AR-GARCH mortality models. *Insurance: Mathematics and Economics* 61, 110–124.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R. and Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics* 52 (2), 275–285.

Plat, R. (2009). On stochastic mortality modelling. *Insurance: Mathematics and Economics* 45, 393–404.

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

Tsai, C.C.L. and Lin, T. (2017a). Incorporating Bühlmann credibility approach to improving mortality forecasting. *Scandinavian Actuarial Journal* 2017, 419–440.

Tsai, C.C.L. and Lin, T. (2017b). A Bühlmann credibility approach to modelling mortality rates. *North American Actuarial Journal* 21 (2), 204–227.

Tsai, C.C.L. and Yang, S. (2015). A linear regression approach to modelling mortality rates of different forms. *North American Actuarial Journal* 19 (1), 1–23.

Wang, C.W., Yang, S.S., and Huang, H.C. (2015). Modeling multi-country mortality dependence and its application in pricing survivor index swaps - A dynamic copula approach. *Insurance: Mathematics and Economics* 63, 30–39.