A Multi-Dimensional Bühlmann Credibility Approach to Modeling Multi-Population Mortality Rates

by

Ying Zhang

B.B.A, Central University of Finance and Economics, 2015

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

in the
Department of Statistics and Actuarial Science
Faculty of Science

© Ying Zhang 2017
SIMON FRASER UNIVERSITY
Summer 2017

All rights reserved.
However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: **Ying Zhang**

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

Title: ***A Multi-Dimensional Bühlmann Credibility Approach to Modeling Multi-Population Mortality Rates***

Examining Committee: **Chair:** Dr. Tim Swartz
Professor

Dr. Cary Chi-Liang Tsai
Senior Supervisor
Associate Professor

Dr. Yi Lu
Supervisor
Associate Professor

Dr. Jean-François Bégin
Internal Examiner
Assistant Professor

Date Defended: 8 June 2017

Abstract

In this project, we first propose a multi-dimensional Bühlmann credibility approach to forecasting mortality rates for multiple populations, and then compare forecasting performances among the proposed approach and the joint-k/co-integrated/augmented common factor Lee-Carter models. The model is applied to mortality data of the Human Mortality Database for both genders of three well-developed countries with an age span and a wide range of fitting year spans. Empirical illustrations show that the proposed multi-dimensional Bühlmann credibility approach contributes to more accurate forecast results, measured by MAPE (mean absolute percentage error), than those based on the Lee-Carter model.

Keywords: Bühlmann Credibility Theory; Lee-Carter Model; Mortality Model

Dedication

To my beloved parents!

Acknowledgements

I would like to express my deepest gratitude to my senior supervisor Dr. Cary Tsai for his guidance, support and encouragement throughout my master's study. I am very lucky to receive professional guidance from a supervisor with rich industry experience, solid knowledge in mortality Modeling, and plenty of patience and kindness. I truly appreciate all his contributions to make my master studies productive and stimulating. The accomplishment of this project would never have been possible without him.

I am also extremely grateful to Dr. Yi Lu and Dr. Jean-François Bégin for their thorough reviews and constructive feedback, which was invaluable to the completion of this project. My thanks are further extended to Dr. Tim Swartz for taking the time out of his busy schedule to serve as the chair of the examining committee. I thank all faculty members and staff of my department who provided me with all kinds of help and support.

Special thanks are given to my managers and coworkers at Insurance Corporation of British Columbia for their generous mentorship and friendship. Further thanks are given to my friends for their encouragement and company during my journey through this degree at Simon Fraser University.

Finally, my deepest gratitude goes to my beloved parents for always being there to support me in all my pursuits. I would not be where I am today without their love, support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter first highlights the motivation for proposing a multi-dimensional credibility approach to Modeling the dynamics of mortality rates for multiple populations, and then gives a brief outline of this project.

1.1 Motivation

In the past decades, life expectancy has been observed to increase drastically and thus pose a severe challenge to pension plan sponsors, annuity providers and social security systems. To hedge against losses associated with this longevity risk, it is of crucial importance to find more accurate mortality projections, since it provides an important actuarial foundation for pricing annuities, life insurance and mortality-linked securities.

The convergence of demographic patterns around the world (see Wilson, 2001) challenged the traditional assumption of mortality independence, and thus finding more accurate mortality forecasts requires multi-population mortality projections. Multi-population mortality projections will eliminate potential long-term divergent behavior from single population mortality forecasts, and improve the model performance due to the increase in available mortality data. Moreover, the development of mortality-linked securities requires better estimate of the underlying mortality index, which is constructed as a weighted average of mortality rates over multiple populations. Therefore, to have a thorough assessment of multi-population mortality risk and the financial payoffs of mortality-linked securities, it is of great importance to take the dependence structure across different populations into account and project future mortality rates for multiple populations simultaneously.

We have seen numerous studies on the improving of forecast accuracy of multi-population mortality projection models, and we give a detailed literature review in Chapter 2. Nonetheless, a majority of research studies contributed to this subject by adding one or more terms to the classical Lee-Carter model so that they do not lead to divergent long-term projections.

On the other hand, Tsai and Lin (2017b) adopt the non-parametric Bühlmann credibility approach to forecasting mortality rates for single population; this approach has proved to generate more stable and satisfactory forecasts than the classical Lee-Carter model, and is convenient for practical implementation. Therefore, in this project, we concentrate on generalizing the single population Bühlmann credibility model to its multi-population counterparts; that is, applying the multi-dimensional Bühlmann credibility approach to Modeling mortality rates for multiple populations jointly. It provides a simple approach to Modeling the development of multi-population mortality rates under the multi-dimensional credibility framework, and eventually produces better forecast results compared to the multi-population versions of the Lee-Carter model.

1.2 Outline

This project is organized as follows. Chapter 2 gives a literature review on previous research on the development of mortality models, and the derivation of credibility theory. In Chapter 3, we first review the three multi-population Lee-Carter-based mortality projection models in details. Then we introduce the multi-dimensional Bühlmann credibility model, and develop formulas for estimating parameters and projecting future mortality rates under both non-parametric and semi-parametric frameworks. In Chapter 4, the models presented in Chapter 3 are applied to mortality data from the Human Mortality Database for both genders of three well-developed countries with an age span and a wide range of fitting year spans; numerical results are provided for illustrations, which show that the multi-dimensional Bühlmann credibility approach outperforms the multi-dimensional Lee-Carter models based on the measure of mean absolute percentage error (MAPE). Chapter 5 concludes this project. The classical Lee-Carter model and the one-dimensional Bühlmann credibility approach to Modeling mortality rates for single population can be referred to Appendices A and B, respectively.

Chapter 2

Literature Review

The two main parts of this project are mortality projection models and Bühlmann credibility theory. Section 2.1 goes through the main articles on mortality Modeling, and Section 2.2 reviews those about credibility models.

2.1 Mortality Forecasting Models

Lee and Carter (1992) develop the well-known Lee-Carter model for mortality fitting and forecasting. The model assumes that the dynamics of the natural logarithm of central death rates are modeled by an age-specific factors plus a bilinear term. The Cairns-Blake-Dowd (CBD) model by Cairns et al. (2006) introduce a novel approach to Modeling mortality dynamics for an elderly age group. It assumes that the dynamics of the logit function of one-year death probabilities are driven by an overall time trend and an age-specific time trend. Extensions of the Lee-Carter and CBD models are common in the literature: Renshaw and Haberman (2006) extend the Lee-Carter Modeling framework by including age-specific cohort effects; Plat (2009) proposes a model combining the good factors in the Lee–Carter and CBD models; Li et al. (2009) consider individual heterogeneity in each age-period cell in the Lee–Carter model; Mitchell et al. (2013) propose a model that accounts for the changes of mortality rate rather than mortality rate levels as in the classical Lee-Carter model; Lin et al. (2015) employ the copula method to capture the inter-age mortality dependence structure and AR-GARCH (autoregression-generalized autoregressive conditional heteroscedasticity) models to capture the marginal dynamics of mortality rates. Tsai and Yang (2015) propose an innovative linear regression approach to relating a target mortality sequence to the base mortality sequence. Different from all afore-mentioned models, Tsai and Lin (2017a, b) adopt a commonly used approach in property and casualty insurance, the Bühlmann credibility approach, to mortality fitting and Modeling: the former incorporates the parametric Bühlmann credibility into the existing Lee-Carter/CBD/liner relational models to improve their forecasting performances; the latter proposes the non-parametric Bühlmann credibility approach to forecasting mortality rates. Single population models and their extensions failed to take the dependence structure across populations into account, and thus may produce divergent forecast among populations.

Numerous multi-population extensions of the classical single population models were developed in the past decades to model the mortality rates for multiple populations simultaneously, and thus ensure the parallel mortality levels in the long-run. For example, Lee and Carter (1992) introduce the joint-k Lee-Carter model, which applies a common time-varying index to all populations. A modification is the the augmented common factor model suggested by Li and Lee (2005). They propose a two-step procedure to model the dynamics for multiple populations: a common factor is first used to capture the overall mortality level; then an augmented term is adopted to model the remaining residuals of a specific population. Another major extension is the co-integrated Lee-Carter model by Li and Hardy (2011), which assumes that there is a linear relationship between the time varying index of a base population and that of each of all other populations. Cairns et al. (2011) introduce a general framework for Modeling the mortality dynamics of two populations jointly: they adopted a mean-reverting process that permits different short-term trends in mortality improvements, but avoids long-term divergence problems. Dowd et al. (2011) give a gravity approach to Modeling mortality dynamics of two populations of different size. Chen et al. (2015) propose a factor copula approach to Modeling mortality dependence for a group of populations.

2.2 Bühlmann Credibility Theory

Credibility theory in property and casualty insurance is an effective and commonly used approach to determining the premium for a group of risks. Bühlmann (1967) proposes a general framework to compute Bühlmann credibility premium, which equals the weighted average of the collective premium and the sample mean of the past observations of a given risk. Bühlmann and Straub (1970) generalize the model by taking the exposure units of risks into consideration, which is called Bühlmann-Straub model.

However, the afore-mentioned research studies commonly ignore dependence among risks. It is not always the case in practice, and hence, Jewell (1975) introduces a hierarchical model that allows dependence among entities; Yeo and Valdez (2006) and Wen et al. (2009) extend the Bühlmann and Bühlmann–Straub credibility models to account for a special type of dependence across risks, induced by common stochastic effects; Dannenburg (1995) and Goulet (2001) propose crossed classification models and their generalizations; Wen and Wu (2011) re-build the credibility estimators for the Bühlmann and Bühlmann–Straub models assuming risks are generally dependent through risk parameters; Poon and Lu (2015) derive the credibility predictors for Bühlmann-type credibility models allowing for both a dependence structure among risk parameters and a conditional spatial cross-sectional dependence among losses.

Chapter 3

Multi-dimensional Mortality Models

In this chapter, we introduce the joint-k/co-integrated/augmented common factor Lee-Carter models and the multi-dimensional Bühlmann credibility approach in more details. The classical Lee-Carter model and the one-dimensional Bühlmann credibility approach to Modeling mortality rates for single population are explained in Appendices A and B, respectively. The study period $[T_1, T_2]$, for which mortality rates are available, is divided into two parts, $[t_L, t_U]$ and $[t_U + 1, T_2]$, where $t_L \geqslant T_1$ and $t_U < T_2$. Assuming that we currently stand at the end of year t_U, the in-sample data in the rectangle $[x_L, x_U] \times [t_L, t_U]$ are used in each model to get the estimated parameters, and then the out-of-sample data in the rectangle $[x_L, x_U] \times [t_U + 1, T_2]$ are compared to the projected mortality rates.

3.1 Concepts and notations

Let $q_{x,t,i}$ denote the probability that lives aged x in year t in the i-th population die between t and $t+1$. Denote $\mu_{x,t,i}$ the associated force of mortality, which represents the instantaneous rate of mortality. Under the assumption that the force of mortality $\mu_{x,t,i}$ is constant within each integer age x and year t, that is, $\mu_{x+r,t+s,i} = \mu_{x,t,i}$ for $r, s \in [0, 1)$, we have $\mu_{x,t,i} = -\ln(1 - q_{x,t,i})$.

The central death rate $m_{x,t,i}$, which is defined as the ratio of the number of deaths during year t at age x to the average number of surviving lives between age x in year t and age $x+1$ in year $t+1$, is another form of mortality rate that is frequently used in literature. Again, under the piecewise constant force of mortality assumption, it can be shown that $\mu_{x,t,i} = m_{x,t,i}$, and thus,

$$q_{x,t,i} = 1 - e^{-\mu_{x,t,i}} = 1 - e^{-m_{x,t,i}}. \tag{3.1}$$

The equations above provide mortality data conversion between $q_{x,t,i}$ and $m_{x,t,i}$ (or $\mu_{x,t,i}$).

3.2 Multi-population Lee-Carter Model

3.2.1 Joint-k Lee-Carter Model

To model the co-movements among the mortality rates for r different populations, Lee and Carter (1992) introduce the joint-k model where the time-varying index $k_{t,i}$, the general mortality level over time, is the same for all populations, that is, $k_{t,i} = K_t$, $i = 1, \ldots, r$.

The natural logarithm of central death rates, $\ln(m_{x,t,i})$, for lives aged x in year t and the i-th population can be expressed as

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

where

- $\alpha_{x,i}$ is the average age-specific mortality factor at age x for population i,
- K_t is the index of the mortality level in year t,
- $\beta_{x,i}$ is the age-specific reaction to K_t at age x for population i, and
- $\varepsilon_{x,t,i}$ is the model error, which is assumed to be independent and identically distributed (i.i.d.) normal for all t with mean 0 and variance $\sigma^2_{\varepsilon_x,i}$, that is, $\{\varepsilon_{x,t,i}\} \overset{i.i.d.}{\sim} N(0, \sigma^2_{\varepsilon_x,i})$.

For uniqueness of the model specification, the following two constraints are imposed:

$$\sum_{i=1}^{r} \sum_{x=x_L}^{x_U} \beta_{x,i} = 1 \text{ and } \sum_{t=t_L}^{t_U} K_t = 0.$$

Given the constrains, estimates of $\alpha_{x,i}$, K_t and $\beta_{x,i}$ can be obtained as follows:

- $\hat{\alpha}_{x,i}$ can be derived by averaging the sum of $\ln(m_{x,t,i})$ over the fitting year span $[t_L, t_U]$:

$$\begin{aligned} \sum_{t=t_L}^{t_U} \ln(m_{x,t,i}) &= n \times \alpha_{x,i} + \beta_{x,i} \times \sum_{t=t_L}^{t_U} K_t = n \times \alpha_{x,i} \\ \Rightarrow \hat{\alpha}_{x,i} &= \frac{\sum_{t=t_L}^{t_U} \ln(m_{x,t,i})}{n}, \quad x = x_L, \ldots, x_U, \end{aligned} \tag{3.2}$$

  where $n = t_U - t_L + 1$;

- $\hat{K}_t$ is equal to the sum of $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ over the fitting age span $[x_L, x_U]$ and the population index:

$$\sum_{i=1}^{r}\sum_{x=x_L}^{x_U}[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}] = K_t \times \sum_{i=1}^{r}\sum_{x=x_L}^{x_U}\hat{\beta}_{x,i} = K_t \times 1$$
$$\Rightarrow \hat{K}_t = \sum_{i=1}^{r}\sum_{x=x_L}^{x_U}[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}], \quad t = t_L, \ldots, t_U;$$

- $\hat{\beta}_{x,i}$ can be obtained by regressing $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ on $\hat{K}_t$ without the constant term for each age x:

$$\hat{\beta}_{x,i} = \sum_{t=t_L}^{t_U}\frac{[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}] \times \hat{K}_t}{(\hat{K}_t)^2}.$$

Assume that the general mortality level in year t, $\hat{K}_t$, follows ARIMA(0,1,0), a random walk with drift θ:

$$\hat{K}_t = \hat{K}_{t-1,} + \theta + \epsilon_t,$$

where

- the time trend error $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$, that is, $\hat{K}_t - \hat{K}_{t-1} \overset{i.i.d.}{\sim} N(\theta, \sigma_\epsilon^2)$, for all t, and
- the time trend errors, $\{\epsilon_t\}$, are assumed to be independent of the model errors, $\{\varepsilon_{x,t,i}\}$.

The unbiased estimator of the mean of the i.i.d. $(\hat{K}_t - \hat{K}_{t-1})$, $\hat{\theta}$, is given by

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

Thus, $K_{t_U+\tau}$ can be projected as $\hat{K}_{t_U+\tau} = \hat{K}_{t_U} + \tau \cdot \hat{\theta}$, where $\tau = t_U + 1, \ldots, T_2$.

Then we can forecast the natural logarithm of the central death rates, $\ln(m_{x,t,i})$, for lives aged x in year t in the i-th population as

$$\begin{aligned}
\ln(\hat{m}_{x,t_U+\tau,i}) &= \hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \times (\hat{K}_{t_U} + \tau \times \hat{\theta}) \\
&= (\hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \times \hat{K}_{t_U}) + \hat{\beta}_{x,i} \times \tau \times \hat{\theta} \\
&= \ln(\hat{m}_{x,t_U,i}) + (\hat{\beta}_{x,i} \times \hat{\theta}) \times \tau, \quad \tau = 1, \ldots, T_2 - t_U.
\end{aligned}$$

It follows that

$$\hat{m}_{x,t_U+\tau,i} = \exp\left[\hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \times (\hat{K}_{t_U} + \tau \times \hat{\theta})\right].$$

From Equation (3.1), the predicted deterministic one-year death rate, $\hat{q}_{x,t_U+\tau,i}$, for age x in year $t_U + \tau$ for population i is given by

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

Two error terms, the model error $\varepsilon_{x,\, t,\, i}$ and the time trend error ϵ_t, can be added to the natural logarithm of the predicted central death rates, $\ln(\hat{m}_{x,\, t_U+\tau,\, i})$, to form the natural logarithm of the stochastic central death rates, $\ln(\tilde{m}_{x,\, t_U+\tau,\, i})$, for age x in year $t_U + \tau$ and population i. Specifically,

$$\begin{aligned}
\ln(\tilde{m}_{x,\, t_U+\tau,\, i}) &= \hat{\alpha}_{x,\, i} + \hat{\beta}_{x,\, i} \times (\hat{K}_{t_U} + \tau \times \hat{\theta} + \sqrt{\tau} \times \epsilon_{t_U+\tau}) + \varepsilon_{x,\, t_U+\tau,\, i} \\
&= \ln(\hat{m}_{x,\, t_U+\tau,\, i}) + \sqrt{\tau} \times \hat{\beta}_{x,\, i} \times \epsilon_{t_U+\tau} + \varepsilon_{x,\, t_U+\tau,\, i}.
\end{aligned}$$

Similarly, the predicted stochastic central death rate and one-year death rate, $\tilde{m}_{x,\, t_U+\tau,\, i}$ and $\tilde{q}_{x,\, t_U+\tau,\, i}$, for age x in year $t_U + \tau$ and population i are given respectively by

$$\tilde{m}_{x,\, t_U+\tau,\, i} = \exp\left[\ln(\hat{m}_{x,\, t_U+\tau,\, i}) + \sqrt{\tau} \times \hat{\beta}_{x,\, i} \times \epsilon_{t_U+\tau} + \varepsilon_{x,\, t_U+\tau,\, i}\right],$$

and

$$\tilde{q}_{x,\, t_U+\tau,\, i} = 1 - \exp\left[-\exp\left(\ln(\hat{m}_{x,\, t_U+\tau,\, i}) + \sqrt{\tau} \times \hat{\beta}_{x,\, i} \times \epsilon_{t_U+\tau} + \varepsilon_{x,\, t_U+\tau,\, i}\right)\right].$$

The estimate of the variance of the model error, $\hat{\sigma}^2_{\varepsilon_{x,\, i}}$, is obtained by

$$\hat{\sigma}^2_{\varepsilon_{x,\, i}} = \frac{1}{n-2} \sum_{t=t_L}^{t_U} (\varepsilon_{x,\, t,\, i})^2 = \frac{\sum_{t=t_L}^{t_U} \left[\ln(m_{x,\, t,\, i}) - \hat{\alpha}_{x,\, i} - \hat{\beta}_{x,\, i} \times \hat{K}_t\right]^2}{n-2},$$

and the estimate of the variance of the time trend error, $\hat{\sigma}^2_{\epsilon}$, is given by

$$\hat{\sigma}^2_{\epsilon} = \frac{1}{n-2} \sum_{t=t_L+1}^{t_U} (\epsilon_t)^2 = \frac{\sum_{t=t_L+1}^{t_U} \left(\hat{K}_t - \hat{K}_{t-1} - \hat{\theta}\right)^2}{n-2}.$$

Therefore, the estimate of the variance of the natural logarithm of the stochastic central death rate, $\hat{\sigma}^2\left(\ln(\tilde{m}_{x,\, t_U+\tau,\, i})\right)$, is given as

$$\hat{\sigma}^2\left(\ln(\tilde{m}_{x,\, t_U+\tau,\, i})\right) = \tau \times \hat{\beta}^2_{x,\, i} \times \hat{\sigma}^2_{\epsilon} + \hat{\sigma}^2_{\varepsilon_{x,\, i}}.$$

3.2.2 Co-integrated Lee-Carter Model

Unlike the joint-k model, which assumes that all populations have the same time-varying coefficient, the co-integrated model depicts the divergence of future forecasts in a different way; that is, assuming the time-varying index for all other populations is a linear transformation of that of a base population.

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

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

The identification of the afore-mentioned model is ensured by two constraints:

$$\sum_{x=x_L}^{x_U} \beta_{x,i} = 1 \text{ and } \sum_{t=t_L}^{t_U} k_{t,i} = 0, \; i = 1, \ldots, r.$$

Estimates of $\alpha_{x,i}$, $k_{t,i}$ and $\beta_{x,i}$ can be obtained as follows:

- estimates of $\alpha_{x,i}$ can be obtained by averaging $\ln(m_{x,t,i})$ over the fitting year span $[t_L, t_U]$:

$$\sum_{t=t_L}^{t_U} \ln(m_{x,t,i}) = n \times \alpha_{x,i} + \beta_{x,i} \times \sum_{t=t_L}^{t_U} k_{t,i} = n \times \alpha_{x,i}$$

$$\Rightarrow \hat{\alpha}_{x,i} = \frac{\sum_{t=t_L}^{t_U} \ln(m_{x,t,i})}{n}, \quad x = x_L, \ldots, x_U;$$

- the initial estimates of $k_{t,i}$ can be obtained by summing $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ over the fitting age span $[x_L, x_U]$:

$$\sum_{x=x_L}^{x_U} [\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}] = k_{t,i} \times \sum_{x=x_L}^{x_U} \beta_{x,i} = k_{t,i} \times 1$$

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

- estimates of $\beta_{x,i}$ can be obtained by regressing $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ on $\hat{k}_{t,i}$ without the constant term involved for each age x:

$$\hat{\beta}_{x,i} = \sum_{t=t_L}^{t_U} \frac{[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}] \times \hat{k}_{t,i}}{(\hat{k}_{t,i})^2}.$$

The time-varying coefficient in year t, $\hat{k}_{t,i}$, is modelled by an ARIMA(0,1,0) process, a random walk with drift θ:

$$\hat{k}_{t,i} = \hat{k}_{t-1,i} + \theta_i + \epsilon_{t,i},$$

where

- $\epsilon_{t,i} \overset{i.i.d.}{\sim} N(0, \sigma^2_{\epsilon_i})$, that is, $k_{t,i} - k_{t-1,i} \overset{i.i.d.}{\sim} N(\theta_i, \sigma^2_{\epsilon_i})$ for all t, and

- the time trend errors, $\{\epsilon_{t,i}\}$, are independent of the model errors, $\{\varepsilon_{x,t,i}\}$.

In this sense, the drift parameter θ_i for population i can be estimated by

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

Hence, $k_{t_U+\tau,\,i}$ can be projected as $\hat{k}_{t_U+\tau,\,i} = \hat{k}_{t_U,i} + \tau \times \hat{\theta}_i$.

In the co-integrated Lee-Carter model, we assume there is a linear relationship plus an error term $e_{t,i}$ between $\hat{k}_{t,\,1}$ (the time-varying index for the base population) and $\hat{k}_{t,\,i}$ for $i = 2, \ldots, r$, that is,

$$\hat{k}_{t,\,i} = a_i + b_i \times \hat{k}_{t,\,1} + e_{t,\,i}, \quad i = 2, .., r.$$

Then $k_{t,\,i}$ is re-estimated as

$$\hat{\hat{k}}_{t,\,i} = \begin{cases} \hat{k}_{t,\,1}, \quad i = 1, \\ \\ \hat{a}_i + \hat{b}_i \times \hat{k}_{t,1}, \quad i = 2, \ldots, r, \end{cases}$$

where $\hat{a}_i$ and $\hat{b}_i$ are obtained by the simple linear regression. Therefore, the re-estimated drift of the time-varying index for population i, $\hat{\hat{\theta}}_i$, is given by

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

Similarly, we can forecast the natural logarithm of the central death rates, $\ln(m_{x,\,t,\,i})$, for lives aged x in year t and the i-th population as

$$\begin{aligned} \ln(\hat{m}_{x,\,t_U+\tau,\,i}) &= \hat{\alpha}_{x,\,i} + \hat{\beta}_{x,\,i} \times (\hat{k}_{t_U,\,i} + \tau \times \hat{\hat{\theta}}_i) \\ &= (\hat{\alpha}_{x,\,i} + \hat{\beta}_{x,\,i} \times \hat{k}_{t_U,\,i}) + \hat{\beta}_{x,\,i} \times \tau \times \hat{\hat{\theta}}_i \\ &= \ln(\hat{m}_{x,t_U,i}) + (\hat{\beta}_{x,\,i} \times \hat{\hat{\theta}}_i) \times \tau, \quad \tau = 1, \ldots, T_2 - t_U. \end{aligned}$$

The logarithm of the stochastic central death rate for age x in year $t_U + \tau$ and population i, denoted by $\ln(\tilde{m}_{x,\,t_U+\tau,\,i})$, is

$$\begin{aligned} \ln(\tilde{m}_{x,\,t_U+\tau,\,i}) &= \hat{\alpha}_{x,\,i} + \hat{\beta}_{x,\,i} \times (\hat{k}_{t_U,\,i} + \tau \times \hat{\hat{\theta}}_i + \sqrt{\tau} \times \epsilon_{t_U+\tau,\,i}) + \varepsilon_{x,\,t_U+\tau,\,i} \\ &= \ln(\hat{m}_{x,\,t_U+\tau,\,i}) + \sqrt{\tau} \times \hat{\beta}_{x,\,i} \times \epsilon_{t_U+\tau,\,i} + \varepsilon_{x,\,t_U+\tau,\,i}. \end{aligned}$$

The estimate of the variance of the model error, $\hat{\sigma}^2_{\varepsilon_{x,\,i}}$, is obtained by

$$\hat{\sigma}^2_{\varepsilon_{x,\,i}} = \frac{1}{n-2}\sum_{t=t_L}^{t_U}(\varepsilon_{x,\,t,\,i})^2 = \frac{\sum\limits_{t=t_L}^{t_U}\left[\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i} - \hat{\beta}_{x,\,i}\times\hat{k}_{t,\,i}\right]^2}{n-2},$$

and the estimate of the variance of the time trend error, $\hat{\sigma}^2_{\epsilon,\,i}$, is given by

$$\hat{\sigma}^2_{\epsilon,\,i} = \begin{cases} \dfrac{\sum\limits_{t=t_L+1}^{t_U}\left(\hat{k}_{t,1}-\hat{k}_{t-1,1}-\hat{\theta}_1\right)^2}{n-2}, & i=1,\\[2ex] \hat{b}_i^2\times\sum\limits_{t=t_L+1}^{t_U}\dfrac{(\hat{k}_{t,1}-\hat{k}_{t-1,1}-\hat{\theta}_1)^2}{n-2} = \hat{b}_i^2\times\hat{\sigma}^2_{\epsilon,1}, & i=2,\dots,r. \end{cases}$$

Therefore, the estimate of the variance of the natural logarithm of the stochastic central death rate, $\sigma^2\left(\ln(\tilde{m}_{x,\,t_U+\tau,\,i})\right)$, is given as

$$\hat{\sigma}^2\left(\ln(\tilde{m}_{x,\,t_U+\tau,\,i})\right) = \tau\times\hat{\beta}^2_{x,\,i}\times\hat{\sigma}^2_{\epsilon,\,i} + \hat{\sigma}^2_{\varepsilon_{x,\,i}}.$$

3.2.3 Augmented Common Factor Model

Lee and Li (2005) propose another approach, adding a common factor term to deal with multiple population data without producing divergent future forecasts in the long-run.

First, the natural logarithm of central death rates, $\ln(m_{x,\,t,\,i})$, for lives aged x in year t and the i-th population is represented as

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

The following features are applied to the common term

$$\begin{aligned} \beta_{x,\,i} &= B_x, \quad x = x_L,\dots,x_U,\\ k_{t,\,i} &= K_t, \quad t = t_L,\dots,t_U. \end{aligned}$$

Thus, the classical Lee-Carter model becomes the so-called common factor model as follows:

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

Two similar constraints are applied to determine a unique solution, i.e.,

$$\sum_{i=1}^{r}\sum_{x=x_L}^{x_U} w_i B_x = 1 \text{ and } \sum_{t=t_L}^{t_U} K_t = 0,$$

where w_i, set to be $\frac{1}{r}$ in this project, is the weight for population i and $\sum_{i=1}^{r} w_i = 1$.

The expression of the estimate of $\alpha_{x,\,i}$ is the same as that in Equation (3.2) for the joint-k Lee-Carter model. That is,

$$\hat{\alpha}_{x,\,i} = \frac{\sum\limits_{t=t_L}^{t_U} \ln(m_{x,\,t,\,i})}{n}, \quad x = x_L, \ldots, x_U.$$

The estimate of K_t is obtained by

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

To get $\hat{B}_x$, we can regress $\sum_{i=1}^{r} w_i \times [\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i}]$ on $\hat{K}_t$ without the constant term involved for each age x, since $\sum_{i=1}^{r} w_i \times [\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i}] = B_x \times \hat{K}_t \times \sum_{i=1}^{r} w_i + \sum_{i=1}^{r} \varepsilon_{x,\,t,\,i} = B_x \times \hat{K}_t + \sum_{i=1}^{r} \varepsilon_{x,\,t,\,i}$.

To further improve the forecasting accuracy of the common factor model, Li and Lee (2005) add a factor $\beta'_{x,\,i} \times k'_{k,i}$ for each population to the common factor model, and thus the augmented common factor model is formed as

$$\ln(m_{x,\,t,\,i}) = \alpha_{x,\,i} + B_x \times K_t + \beta'_{x,\,i} \times k'_{k,i} + \epsilon_{x,\,t,\,i}, \quad i = 1, \ldots, r,$$

with an extra constraint $\sum\limits_{x=x_L}^{x_U} \beta'_{x,\,i} = 1$, which implies

$$\hat{k}'_{t,\,i} = \sum_{x=x_L}^{x_U} \left[\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i} - \hat{B}_x \times \hat{K}_t\right].$$

Finally, $\hat{\beta}'_{x,\,i}$ can be derived by regressing $\left[\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i} - \hat{B}_x \times \hat{K}_t\right]$ on $\hat{k}'_{t,\,i}$ without the constant term involved for each age x.

Moreover, $\hat{K}_t$ and $\hat{k}'_{t,\,i}$ are assumed to follow a random walk with drifts θ and θ'_i, respectively, that is, $\hat{K}_t = \hat{K}_{t-1} + \theta + \epsilon_t$, and $\hat{k}'_{t,\,i} = \hat{k}'_{t-1,i} + \theta'_i + \epsilon_{t,\,i}$, where

- $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$ and $\epsilon_{t,\,i} \overset{i.i.d.}{\sim} N(0, \sigma_{\epsilon,i}^2)$, and thus, $\hat{K}_t - \hat{K}_{t-1} \overset{i.i.d.}{\sim} N(\theta,\, \sigma_\epsilon^2)$ and $\hat{k}'_{t,\,i} - \hat{k}'_{t-1,\,i} \overset{i.i.d.}{\sim} N(\theta'_i,\, \sigma_{\epsilon,i}^2)$, and
- all of the three error terms, $\{\varepsilon_{x,\,t,\,i}\}$, $\{\epsilon_t\}$ and $\{\epsilon_{t,\,i}\}$ are assumed to be independent.

Similarly, the parameters θ and θ'_i can be estimated by

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

and

$$\hat{\theta}'_i = \frac{1}{n-1} \sum_{t=t_L+1}^{t_U} (\hat{k}'_{t,\,i} - \hat{k}'_{t-1,i}) = \frac{\hat{k}'_{t_U,i} - \hat{k}'_{t_L,i}}{n-1}.$$

Hence, $K_{t_U+\tau}$ and $k'_{t_U+\tau,\,i}$ can be projected as

$$\hat{K}_{t_U+\tau} = \hat{K}_{t_U} + \tau \times \hat{\theta} \text{ and } \hat{k}'_{t_U+\tau,\,i} = \hat{k}'_{t_U,i} + \tau \times \hat{\theta}'_i.$$

Therefore, the natural logarithm of the predicted central death rates, $\ln(\hat{m}_{x,\,t,\,i})$, for lives aged x in year t and the i-th population can be expressed as

$$\begin{aligned}
\ln(\hat{m}_{x,\,t_U+\tau,\,i}) &= \hat{\alpha}_{x,\,i} + \hat{B}_x \times (\hat{K}_{t_U} + \tau \times \hat{\theta}) + \hat{\beta}'_{x,\,i} \times (\hat{k}'_{t_U,\,i} + \tau \times \hat{\theta}'_i) \\
&= (\hat{\alpha}_{x,\,i} + \hat{B}_x \times \hat{K}_{t_U} + \hat{\beta}'_{x,\,i} \times \hat{k}'_{t_U,\,i}) + \hat{B}_x \times \tau \times \hat{\theta} + \hat{\beta}'_{x,\,i} \times \tau \times \hat{\theta}'_i \\
&= \ln(\hat{m}_{x,\,t_U,\,i}) + (\hat{B}_x \times \hat{\theta} + \hat{\beta}'_{x,\,i} \times \hat{\theta}'_i) \times \tau, \quad \tau = 1, \ldots, T_2 - t_U.
\end{aligned}$$

Moreover, the logarithm of the stochastic central death rate for age x year $t_U + \tau$ and population i denoted by $\ln(\tilde{m}_{x,\,t_U+\tau,\,i})$ is

$$\begin{aligned}
&\ln(\tilde{m}_{x,\,t_U+\tau,\,i}) \\
&= \hat{\alpha}_{x,\,i} + \hat{B}_x \times (\hat{K}_{t_U} + \tau \times \hat{\theta} + \sqrt{\tau} \times \epsilon_{t_U+\tau}) + \hat{\beta}'_{x,\,i} \times \left(\hat{k}'_{t_U,i} + \tau \times \hat{\theta}_i + \sqrt{\tau} \times \epsilon_{t_U+\tau,\,i}\right) + \varepsilon_{x,\,t_U+\tau,\,i} \\
&= \ln(\hat{m}_{x,\,t_U+\tau,\,i}) + \sqrt{\tau} \times (\hat{B}_x \times \epsilon_{t_U+\tau} + \hat{\beta}'_{x,\,i} \times \epsilon_{t_U+\tau,\,i}) + \varepsilon_{x,\,t_U+\tau,\,i}.
\end{aligned}$$

Note that the variance of $\ln(\tilde{m}_{x,\,t_U+\tau,\,i})$ is estimated by

$$\hat{\sigma}^2 \left(\ln(\tilde{m}_{x,\,t_U+\tau,\,i})\right) = \tau \times (\hat{B}_x^2 \times \hat{\sigma}_\epsilon^2 + \hat{\beta}'^2_{x,\,i} \times \hat{\sigma}^2_{\epsilon,i}) + \hat{\sigma}^2_{\varepsilon_{x,\,i}},$$

where

- the estimate of the variance of the model error $\varepsilon_{x,\,t,\,i}$ is

$$\hat{\sigma}^2_{\varepsilon_{x,\,i}} = \frac{\sum_{t=t_L}^{t_U} \left[\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i} - \hat{B}_x \times \hat{K}_t - \hat{\beta}'_{x,\,i} \times \hat{k}'_{t,\,i}\right]^2}{n-3}, \quad i = 1, \ldots, r;$$

- the estimate of the variance of the time trend error ϵ_t is

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{t=t_L}^{t_U} \left(\hat{K}_t - \hat{K}_{t-1} - \hat{\theta}\right)^2}{n-2};$$

- the estimate of the variance of the time trend error $\epsilon_{t,\,i}$ for population i is

$$\hat{\sigma}_{\epsilon,i}^2 = \frac{\sum_{t=t_L}^{t_U} \left(\hat{k}'_{t,\,i} - \hat{k}'_{t-1,i} - \hat{\theta}'_i\right)^2}{n-2}, \quad i = 1, \ldots, r.$$

3.3 Multi-dimensional Bühlmann Credibility Approach

The afore-mentioned Lee-Carter-based models govern $\ln(m_{x,\,t,\,i})$. Empirical mortality data of both genders of the U.S.A., the U.K. and Japan show that $\ln(m_{x,\,t})$s display a downward trend over t (see Figure 3.1), where the "Avg" curve is the average of $\ln(m_{x,t})$s over $x = [25, 84]$). To eliminate the downward trend, and thus apply the Bühlmann credibility approach, we choose to model $Y_{x,\,t,\,i} = \ln(m_{x,\,t,\,i}) - \ln(m_{x,\,t-1,\,i})$ for $x \in [x_L, x_U]$ and $t \in [t_L + 1, t_U]$ (see Figure 3.2).

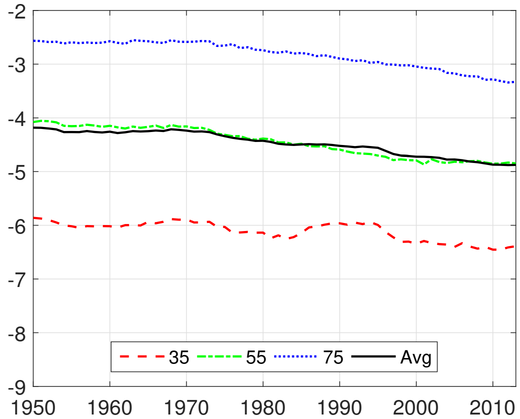
3.3.1 Credibility Estimation

Assume we have $(n-1)$ column vectors of past observed values, $\boldsymbol{Y}_{x,\,t_L+1}, \ldots, \boldsymbol{Y}_{x,\,t_U}$, and we would like to get the credibility estimator $\hat{\boldsymbol{Y}}_{x,\,t_U+1}$ for ages $x = x_L, \ldots, x_U$ and the next year $t_U + 1$, where $t_U - t_L = n - 1$ and $\boldsymbol{Y}_{x,\,t} = (Y_{x,\,t,\,1}, \ldots, Y_{x,\,t,\,r})'$ is a column vector of length r. To implement the Bühlmann credibility approach, we further assume that $\boldsymbol{Y}_{x,\,t}$, where $t = t_L + 1, \ldots, t_U$, is characterized by an $r \times 1$ column vector of risk parameters, $\boldsymbol{\Theta}_x$, associated with age x.

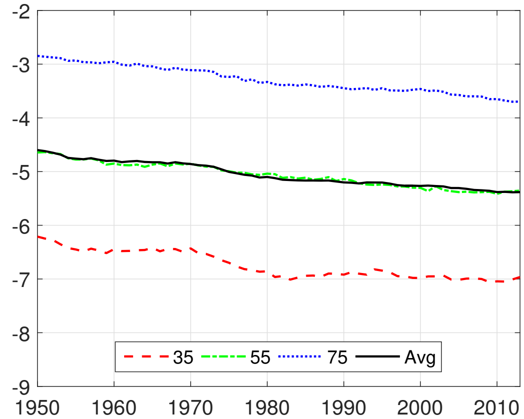
Among the various possible predictors of $\boldsymbol{Y}_{x,\,t_U+1}$, we choose to forecast $\boldsymbol{Y}_{x,\,t_U+1}$ with a linear function of the past data $\boldsymbol{Y}_x(i)$, $i = 1, \ldots, r$, i.e., $\boldsymbol{c}_{x,\,0} + \sum_{i=1}^{r} \boldsymbol{C}'_{x,\,i}\, \boldsymbol{Y}_x(i)$, where $\boldsymbol{c}_{x,\,0}$ is an $r \times 1$ vector with each element taking values in $\mathbb{R}$, $\boldsymbol{C}_{x,\,i}$ is the $(n-1) \times r$ coefficient matrix with each element taking values in $\mathbb{R}$, and $\boldsymbol{Y}_x(i) = (Y_{x,\,t_L+1,\,i}, \ldots, Y_{x,\,t_U,\,i})'$. Specifically, we would like to choose $\boldsymbol{c}_{x,\,0}, \boldsymbol{C}_{x,\,1} \ldots, \boldsymbol{C}_{x,\,r}$ to minimize the quadratic loss function Q, where

$$\mathrm{Q} = \mathrm{E}\left\{\left[\boldsymbol{Y}_{x,t_U+1} - \boldsymbol{c}_{x,\,0} - \sum_{i=1}^{r} \boldsymbol{C}'_{x,\,i}\, \boldsymbol{Y}_x(i)\right]^2\right\}. \tag{3.3}$$

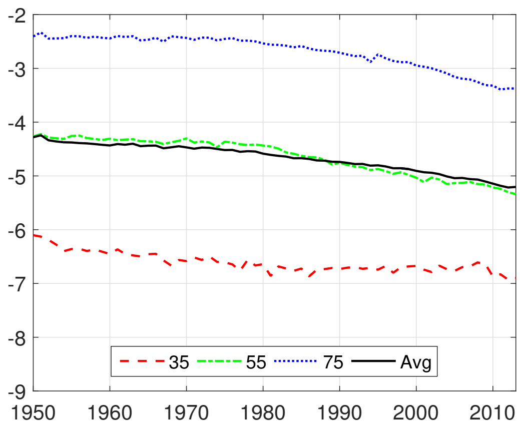
Figure 3.1: $\ln(m_{x,t,i})$ against t



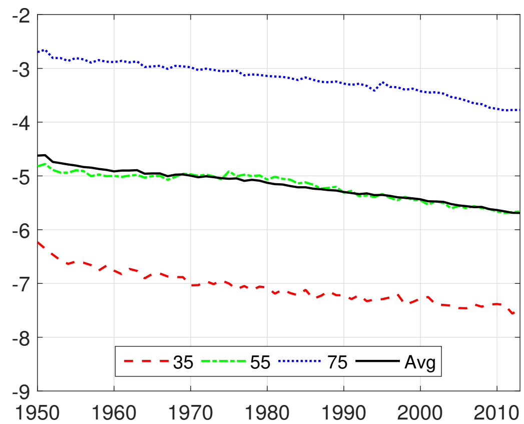
(a) U.S.A. Males



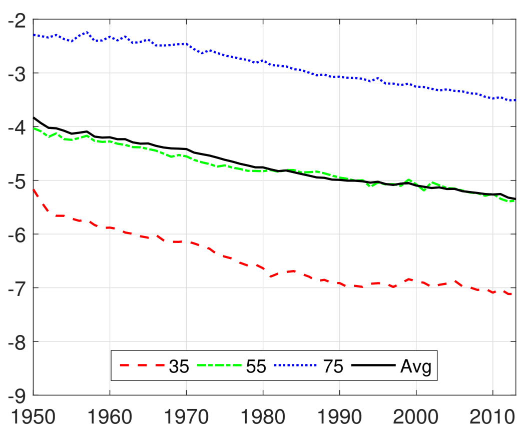
(b) U.S.A. Females



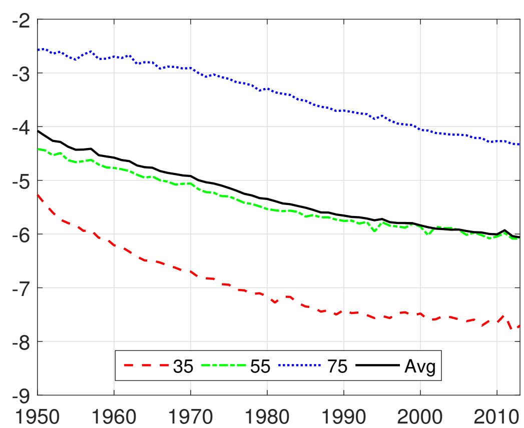
(c) U.K. Males



(d) U.K. Females

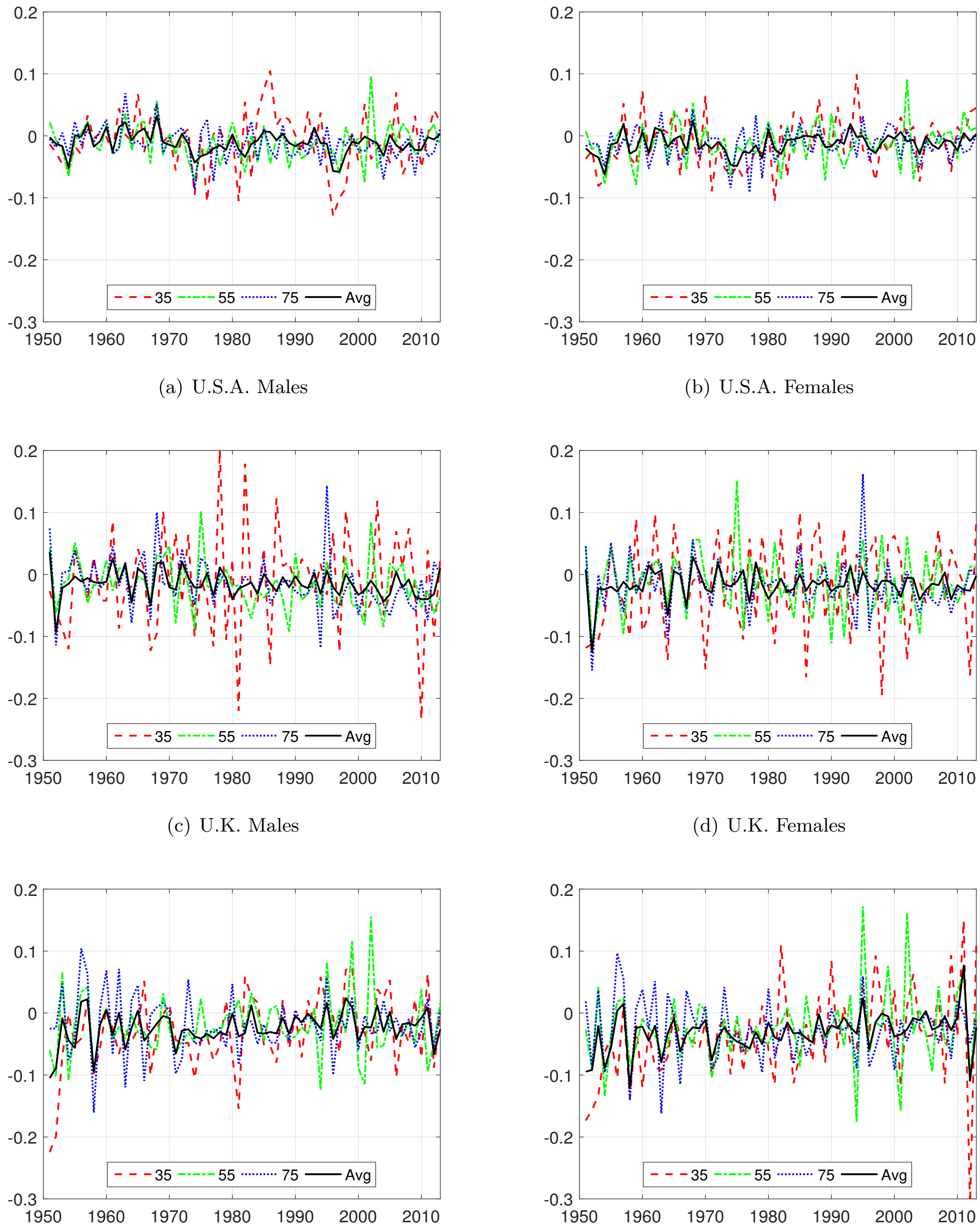


(e) Japan Males



(f) Japan Females

Figure 3.2: $Y_{x,t,i} = \ln(m_{x,t,i}) - \ln(m_{x,t-1,i})$ against t



(a) U.S.A. Males

(b) U.S.A. Females

(c) U.K. Males

(d) U.K. Females

(e) Japan Males

(f) Japan Females

It was proven in Wen et al. (2009) that the inhomogeneous (that is, $\boldsymbol{c}_{x,0}$ is included in Q) linear credibility estimator which minimizes the quadratic loss function in (3.3) is given by

$$\hat{\boldsymbol{Y}}_{x,t_U+1} = \boldsymbol{\mu}_{\boldsymbol{Y}_{x,\, t_U+1}} + (\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\, \boldsymbol{Y}_x})(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}), \tag{3.4}$$

where

- $\boldsymbol{Y}_x = (\boldsymbol{Y}_{x,\, t_L+1}, \ldots, \boldsymbol{Y}_{x,\, t_U})'$,
- $\mathrm{E}(\boldsymbol{Y}_x) = \boldsymbol{\mu}_{\boldsymbol{Y}_x}$, $\mathrm{E}(\boldsymbol{Y}_{x,\, t_U+1}) = \boldsymbol{\mu}_{\boldsymbol{Y}_{x,\, t_U+1}}$, and
- $\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x} = \mathrm{Cov}[\boldsymbol{Y}_x,\, \boldsymbol{Y}_x']$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\boldsymbol{Y}_x} = \mathrm{Cov}[\boldsymbol{Y}_{x,t_U+1},\, \boldsymbol{Y}_x]$ are invertible covariance matrices.

3.3.2 Parametric Bühlmann Model

This section first introduces the notations and the assumptions used in this project. Then, the construction of a multi-dimensional parametric Bühlmann credibility model is presented.

The following specifies the additional assumptions of the distributions of $\boldsymbol{Y}_{x,t}|\boldsymbol{\Theta}_x$ and $\boldsymbol{\Theta}_x$ to construct the multi-dimensional parametric Bühlmann credibility model,

Assumption 1. *Conditional on the $r \times 1$ vector of risk parameters $\boldsymbol{\Theta}_x = (\Theta_{x,1}, \ldots, \Theta_{x,r})'$, $\boldsymbol{Y}_{x,t} = (Y_{x,t,1}, \ldots, Y_{x,t,r})'$ are independent and identically distributed for $t = t_L, \ldots, t_U$ with*

$$\begin{cases} E\left[\boldsymbol{Y_{x,t}}|\boldsymbol{\Theta}_x\right] & = \mu(\boldsymbol{\Theta}_x) & = \left(\mu\left(\Theta_{x,1}\right), \ldots, \mu\left(\Theta_{x,r}\right)\right)', \\ Cov\left[\boldsymbol{Y_{x,t}},\, \boldsymbol{Y_{x,t}'}|\boldsymbol{\Theta}_x\right] & = \boldsymbol{\Sigma}(\boldsymbol{\Theta}_x) & = \left[\sigma^2\left(\Theta_{x,i}, \Theta_{x,j}\right)\right]_{i,\, j=1,\ldots,r}. \end{cases}$$

Assumption 2. *$\boldsymbol{\Theta}_x = (\Theta_{x,1}, \ldots, \Theta_{x,r})'$ are independent and identically distributed as $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_r)'$ for $x = x_L, \ldots, x_U$ with*

$$\begin{cases} \mu\left(\boldsymbol{\Theta_x}\right) & = \mu(\boldsymbol{\Theta}) & = \left(\mu\left(\Theta_1\right), \ldots, \mu\left(\Theta_r\right)\right)', \\ \boldsymbol{\Sigma}\left(\boldsymbol{\Theta}_x\right) & = \boldsymbol{\Sigma}(\boldsymbol{\Theta}) & = \left[\sigma^2\left(\Theta_i, \Theta_j\right)\right]_{i,\, j=1,\ldots,r}. \end{cases}$$

Assumption 3. *The distribution of $\boldsymbol{\Theta}$ is such that*

$$\begin{cases} \boldsymbol{\mu} & = E[\mu\left(\boldsymbol{\Theta}\right)] & = (\mu(1), \ldots, \mu(r))', \\ \boldsymbol{V} & = E[\boldsymbol{\Sigma}\left(\boldsymbol{\Theta}\right)] & = [v(i,\, j)]_{i,\, j=1,\ldots,r}, \\ \boldsymbol{A} & = Cov[\mu\left(\boldsymbol{\Theta}\right)] & = [a(i,\, j)]_{i,\, j=1,\ldots,r}. \end{cases}$$

We denote the structural parameters of the multi-dimensional credibility model as follows:

- the hypothetical mean, $\boldsymbol{\mu}(\boldsymbol{\Theta}_x) = \mathrm{E}[\boldsymbol{Y}_{x,\,t}|\boldsymbol{\Theta}_x]$;
- the process covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\Theta}_x) = \mathrm{Cov}[\boldsymbol{Y}_{x,\,t},\, \boldsymbol{Y}'_{x,\,t}|\boldsymbol{\Theta}_x]$;
- the expected value of the hypothetical mean, $\boldsymbol{\mu} = \mathrm{E}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right] = \mathrm{E}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}\right)\right] = (\mu(1), \ldots, \mu(r))'$;
- the expected process covariance matrix, $\boldsymbol{V} = \mathrm{E}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Theta}_x\right)\right] = \mathrm{E}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Theta}\right)\right] = \left[v(i,\, j)\right]_{i,\,j=1,\ldots,r}$;
- the covariance matrix of hypothetical mean, $\boldsymbol{A} = \mathrm{Cov}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right] = \mathrm{Cov}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}\right)\right] = \left[a(i,\, j)\right]_{i,\,j=1,\ldots,r}$.

Lemma 1. *Under Assumptions* 1 *to* 3 *and the notations above,*

1. *The means of* $\boldsymbol{Y}_x$ *and* $\boldsymbol{Y}_{x,\,t_U+1}$ *are given by*

$$\boldsymbol{\mu}_{\boldsymbol{Y}_x} = E(\boldsymbol{Y}_x) = \boldsymbol{\mu} \otimes \mathbf{1}_{n-1} \text{ and } \boldsymbol{\mu}_{\boldsymbol{Y}_{x,t_U+1}} = E(\boldsymbol{Y}_{x,t_U+1}) = \boldsymbol{\mu},$$

where $\otimes$ *is the Kronecker product operator (see Appendix C for its definition and properties) and* $\mathbf{1}_{n-1} = (1, \ldots, 1)'$*, a column vector of length* $(n-1)$*.*

2. *The* $r(n-1) \times r(n-1)$ *covariance matrix of* $\boldsymbol{Y}_x$ *is given by*

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x} = \boldsymbol{V} \otimes \boldsymbol{I}_{n-1} + \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}', \tag{3.5}$$

where

$$\boldsymbol{I}_{n-1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}_{(n-1)\times(n-1)},$$

and

$$\boldsymbol{U} = \begin{bmatrix} \mathbf{1}_{n-1} & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{1}_{n-1} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \mathbf{1}_{n-1} \end{bmatrix}_{r(n-1)\times r}.$$

3. *The* $r \times r(n-1)$ *covariance matrix between* $\boldsymbol{Y}_x$ *and* $\boldsymbol{Y}_{x,t_U+1}$ *is given by*

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,\,t_U+1},\boldsymbol{Y}_x} = \boldsymbol{A}\boldsymbol{U}'. \tag{3.6}$$

4. The inverse of the $r(n-1) \times r(n-1)$ covariance matrix of $\mathbf{Y}_x$ is given by

$$(\mathbf{\Sigma}_{\mathbf{Y}_x,\mathbf{Y}_x})^{-1} = (\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_{n-1}) - (\boldsymbol{V}^{-1} \otimes \mathbf{1}_{n-1})\boldsymbol{U}\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}\boldsymbol{U}'(\boldsymbol{V}^{-1} \otimes \mathbf{1}'_{n-1}).$$

Proof. Please refer to Appendix D. □

Theorem 1. *Under Assumptions 1 to 3, the parametric Bühlmann estimate of $\mathbf{Y}_{x,\,t_U+1}$, $\hat{\mathbf{Y}}_{x,\,t_U+1} = \left(\hat{Y}_{x,\,t_U+1,\,1}, \ldots, \hat{Y}_{x,\,t_U+1,\,r}\right)'$, for age x in year t_U+1, which is obtained by minimizing the quadratic loss function in Equation (3.3) is given by*

$$\hat{\mathbf{Y}}_{x,\,t_U+1} = \boldsymbol{Z}\overline{\boldsymbol{Y}}_{x,\bullet} + (\boldsymbol{I}_r - \boldsymbol{Z})\,\boldsymbol{\mu},$$

where

- $\boldsymbol{Z} = \boldsymbol{A}\left(\frac{1}{n-1}\boldsymbol{V} + \boldsymbol{A}\right)^{-1}$, *and*
- $\overline{\boldsymbol{Y}}_{x,\bullet} = \left(\overline{Y}_{x,\bullet,1}, \ldots, \overline{Y}_{x,\bullet,r}\right)' = \frac{1}{n-1}\left(\sum_{t=t_L+1}^{t_U} Y_{x,\,t,\,1}, \ldots, \sum_{t=t_L+1}^{t_U} Y_{x,\,t,\,r}\right)'.$

Proof. Please refer to Appendix D. □

3.3.3 Non-parametric Bühlmann Model

To get the non-parametric estimators of $\boldsymbol{\mu}$, $\boldsymbol{V}$ and $\boldsymbol{A}$, we further assume that

Assumption 4. *The pairs $\{(\mathbf{\Theta}_x, \mathbf{Y}_x),\ x = x_L, \ldots, x_U\}$, where $\mathbf{Y}_{\boldsymbol{x}} = (\mathbf{Y}_{x,\,t_L+1}, \ldots, \mathbf{Y}_{x,\,t_U})$, are independent.*

The preliminary unbiased estimators of $\boldsymbol{\mu}$, $\boldsymbol{V}$ and $\boldsymbol{A}$ are presented in Tables 3.1 and 3.2. Here, $\hat{\boldsymbol{A}} = [\hat{a}(i,\,j)]_{i,\,j=1,\ldots,r}$ is the estimator of the variance-covariance matrix of the hypothetical mean vector; the diagonal elements, $\hat{a}(i,\,i)$'s, are the estimators of the variances of the hypothetical means, and the off-diagonal elements, $\hat{a}(i,\,j)$'s for $i \neq j$, are the estimators of the covariances between all possible pairs of the hypothetical means. Note that it is possible that $\hat{a}(i,i) < 0$ or $\hat{a}(i,j) > \sqrt{\hat{a}(i,i) \times \hat{a}(j,j)}$. It is customary to follow the steps below suggested by Bühlmann and Gisler (2005)

- Step 1: set $\hat{a}(i,\,i) = 0$ if $\hat{a}(i,\,i) < 0$ for $i = 1, \ldots, r$;
- Step 2: set $\hat{a}(i,\,j) = \text{sign}[\hat{a}(i,\,j)] \times \min\left[|\hat{a}(i,\,j)|, \sqrt{\hat{a}(i,\,i) \times \hat{a}(j,\,j)}\right]$ for $i,\,j = 1, \ldots, r$ and $i \neq j$.

Table 3.1: Non-parametric Bühlmann estimation for $\mu(i)$, $v(i,\, j)$ and $a(i,\, j)$

$\boldsymbol{Y}_{x_L}(i) = (Y_{x_L,\,t_L+1,\,i}, \ldots, Y_{x_L,\,t_U,\,i})'$	$\hat{\mu}_{x_L}(i) = \overline{Y}_{x_L,\,\bullet,\,i} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U} Y_{x_L,\,t,\,i}, \quad i = 1, \ldots, r$ $\hat{v}_{x_L}(i,\, j) = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}\left[Y_{x_L,\,t,\,i} - \overline{Y}_{x_L,\,\bullet,\,i}\right]\left[Y_{x_L,\,t,\,j} - \overline{Y}_{x_L,\,\bullet,\,j}\right]$
$\vdots$	$\vdots$
$\boldsymbol{Y}_{x_U}(i) = (Y_{x_U,\,t_L+1,\,i}, \ldots, Y_{x_U,\,t_U,\,i})'$	$\hat{\mu}_{x_U}(i) = \overline{Y}_{x_U,\,\bullet,\,i} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U} Y_{x_U,\,t,\,i}, \quad i = 1, \ldots, r$ $\hat{v}_{x_U}(i,\, j) = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}\left[Y_{x_U,\,t,\,i} - \overline{Y}_{x_U,\,\bullet,\,i}\right]\left[Y_{x_U,\,t,\,j} - \overline{Y}_{x_U,\,\bullet,\,j}\right]$
$\hat{a}(i,\, j) = \frac{1}{m-1}\sum_{x=x_L}^{x_U}\left[\overline{Y}_{x,\,\bullet,\,i} - \overline{Y}_{\bullet,\,\bullet,\,i}\right]\left[\overline{Y}_{x,\,\bullet,\,j} - \overline{Y}_{\bullet,\,\bullet,\,i}\right] - \frac{\hat{v}(i,j)}{n-1}$, where $m = x_U - x_L + 1$	$\hat{\mu}(i) = \overline{Y}_{\bullet,\,\bullet,\,i} = \frac{1}{m}\sum_{x=x_L}^{x_U} \overline{Y}_{x,\,\bullet,\,i}, \quad i = 1, \ldots, r$ $\hat{v}(i,\, j) = \frac{1}{m}\sum_{x=x_L}^{x_U} \hat{v}_x(i,\, j), \quad i,\, j = 1, \ldots, r$

Table 3.2: Non-parametric Bühlmann estimation for $\boldsymbol{\mu}$, $\boldsymbol{V}$ and $\boldsymbol{A}$

$\boldsymbol{Y}_{x_L} = (\boldsymbol{Y}_{x_L}(1), \ldots, \boldsymbol{Y}_{x_L}(r))$	$\hat{\boldsymbol{\mu}}_{x_L} = \overline{\boldsymbol{Y}}_{x_L} = \left(\overline{Y}_{x_L,\,\bullet,\,1}, \ldots, \overline{Y}_{x_L,\,\bullet,\,r}\right)'$ $\hat{\boldsymbol{V}}_{x_L} = [\hat{v}_{x_L}(i,\, j)]_{i,\,j=1,\ldots,r}$
$\vdots$	$\vdots$
$\boldsymbol{Y}_{x_U} = (\boldsymbol{Y}_{x_U}(1), \ldots, \boldsymbol{Y}_{x_U}(r))$	$\hat{\boldsymbol{\mu}}_{x_U} = \overline{\boldsymbol{Y}}_{x_U} = \left(\overline{Y}_{x_U,\,\bullet,\,1}, \ldots, \overline{Y}_{x_U,\,\bullet,\,r}\right)'$ $\hat{\boldsymbol{V}}_{x_U} = [\hat{v}_{x_U}(i,\, j)]_{i,\,j=1,\ldots,r}$
$\hat{\boldsymbol{A}} = [\hat{a}(i,\, j)]_{i,\,j=1,\ldots,r}$	$\hat{\boldsymbol{\mu}} = (\hat{\mu}(1), \ldots, \hat{\mu}(r))' = \frac{1}{m}\sum_{x=x_L}^{x_U} \hat{\boldsymbol{\mu}}_x$ $\hat{\boldsymbol{V}} = [\hat{v}(i,\, j)]_{i,\,j=1,\ldots,r} = \frac{1}{m}\sum_{x=x_L}^{x_U} \hat{\boldsymbol{V}}_x$

Therefore, the non-parametric Bühlmann estimate, $\hat{\boldsymbol{Y}}_{x,\,t_U+1} = \left(\hat{Y}_{x,\,t_U+1,\,1}, \ldots, \hat{Y}_{x,\,t_U+1,\,r}\right)'$, for age x in year $t_U + 1$, which is obtained by minimizing the quadratic loss function in Equation (3.3) is given by

$$\hat{\boldsymbol{Y}}_{x,t_U+1} = \hat{\boldsymbol{Z}}\overline{\boldsymbol{Y}}_{x,\,\bullet} + \left(\boldsymbol{I}_r - \hat{\boldsymbol{Z}}\right)\hat{\boldsymbol{\mu}},$$

where

- $\hat{\boldsymbol{Z}} = \hat{\boldsymbol{A}}\left(\frac{1}{n-1}\hat{\boldsymbol{V}} + \hat{\boldsymbol{A}}\right)^{-1}$,
- $\overline{\boldsymbol{Y}}_{x,\,\bullet} = \left(\overline{Y}_{x,\,\bullet,\,1}, \ldots, \overline{Y}_{x,\,\bullet,\,r}\right)' = \frac{1}{n-1}\left(\sum_{t=t_L+1}^{t_U} Y_{x,\,t,\,1}, \ldots, \sum_{t=t_L+1}^{t_U} Y_{x,\,t,\,r}\right)'$, and
- $\hat{\boldsymbol{\mu}} = (\hat{\mu}(1), \ldots, \hat{\mu}(r))' = (\overline{Y}_{\bullet\bullet,\,1}, \ldots, \overline{Y}_{\bullet\bullet,\,r})' = \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}$.

Remark 1. *The non-parametric Bühlmann credibility estimate* $\hat{\boldsymbol{Y}}_{x,\,t_U+1}$ *is the weighted average of* $\overline{\boldsymbol{Y}}_{x,\bullet}$ *and* $\overline{\boldsymbol{Y}}_{\bullet,\bullet}$*, where*

- $\overline{\boldsymbol{Y}}_{x,\bullet} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,t} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U}(Y_{x,t,1},\ldots,Y_{x,t,r})'$ *is the average decrement of the* r*-dimensional individual time trend,* $(\{\ln(m_{x,t,1})\},\ldots,\{\ln(m_{x,t,r})\})$*, for age* x *and* r *populations per year over the period* $[t_L+1,\, t_U]$*, and*

- $\overline{\boldsymbol{Y}}_{\bullet,\bullet} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U}\overline{\boldsymbol{Y}}_{\bullet,t} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U}(\overline{Y}_{\bullet,t,1},\ldots,\overline{Y}_{\bullet,t,r})$*, with*

$$\overline{Y}_{\bullet,t,i} = \frac{1}{m}\sum_{x=x_L}^{x_U}\ln(m_{x,t,i}) - \frac{1}{m}\sum_{x=x_L}^{x_U}\ln(m_{x,t-1,i})$$

for $i=1,\ldots,r$*, is the average decrement of the* r*-dimensional group time trend,* $\left(\left\{\frac{1}{m}\sum_{x=x_L}^{x_U}\ln(m_{x,t,1})\right\},\ldots,\left\{\frac{1}{m}\sum_{x=x_L}^{x_U}\ln(m_{x,t,r})\right\}\right)$*, for* r *populations per year over the period* $[t_L+1,\, t_U]$*.*

Finally, the r-dimensional non-parametric Bühlmann estimate of $\ln(\boldsymbol{m}_{x,\,t_U+1})$ for age x in year t_U+1 is $\ln(\hat{\boldsymbol{m}}_{x,\,t_U+1}) = \ln(\boldsymbol{m}_{x,\,t_U}) + \hat{\boldsymbol{Y}}_{x,\,t_U+1}$.

To obtain the non-parametric Bühlmann credibility estimate $\hat{Y}_{x,\,t_U+\tau}$ for year $t_U+\tau$ ($\tau \geqslant 2$), which is

$$\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} = \boldsymbol{Z}(t_U+\tau)\cdot\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) + [\boldsymbol{I}_r - \boldsymbol{Z}(t_U+\tau)]\cdot\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau), \tag{3.7}$$

where $(t_U+\tau)$ is attached to each of $\boldsymbol{Z}$, $\overline{\boldsymbol{Y}}_{x,\bullet}$, and $\overline{\boldsymbol{Y}}_{\bullet,\bullet}$ to indicate those quantities are for year $t_U+\tau$, two strategies, the expanding window (EW) strategy and the moving window (MW) strategy, are proposed (see Tsai and Lin, 2017a, b) in the following section.

Strategy EW: Expanding window by one year.

Under the expanding window (EW) strategy, the following steps are adopted to get the Bühlmann credibility estimate $\hat{\boldsymbol{Y}}_{x,\,t_U+\tau}$ for $\tau \geqslant 2$:

- first, add the credibility estimates $\{\hat{\boldsymbol{Y}}_{x,\,t_U+1},\ldots,\hat{\boldsymbol{Y}}_{x,\,t_U+\tau-1}\}$ to $\{\boldsymbol{Y}_{x,\,t_L+1},\ldots,\boldsymbol{Y}_{x,\,t_U}\}$ such that the fitting year span is expanded by τ years to $[t_L, t_U+\tau-1]$;

- next, obtain

$$\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) = \frac{1}{n+\tau-2}\left[\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,t} + \sum_{t=t_U+1}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,t}\right], \tag{3.8}$$

$$\hat{\boldsymbol{\mu}}(t_U+\tau) = \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) = \frac{1}{m}\sum_{x=x_L}^{x_U}\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau), \tag{3.9}$$

and

$$\hat{\boldsymbol{Z}}(t_U+\tau) = \hat{\boldsymbol{A}}\left(\frac{1}{n+\tau-2}\hat{\boldsymbol{V}} + \hat{\boldsymbol{A}}\right)^{-1} \tag{3.10}$$

using the data in the year span $[t_L, t_U+\tau-1]$, where $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{A}}$ in $\boldsymbol{Z}(t_U+\tau)$ are the same as those in $\hat{\boldsymbol{Z}}(t_U+1)$;

- finally, obtain the Bühlmann credibility estimates $\hat{\boldsymbol{Y}}_{x,\,t_U+\tau}$ and $\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau})$ for year $t_U+\tau$ by

$$\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} = \hat{\boldsymbol{Z}}(t_U+\tau)\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) + \left[\boldsymbol{I}_r - \hat{\boldsymbol{Z}}(t_U+\tau)\right]\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau), \tag{3.11}$$

and

$$\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau}) = \ln(\boldsymbol{m}_{x,\,t_U}) + \sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}_{x,\,t_U+t} \tag{3.12}$$

for $\tau = 1, \ldots, T_2 - t_U$.

It is worth noting that $\boldsymbol{Z}(t_U+\tau)$ in Equation (3.10) is increasing in τ for the EW strategy.

Strategy MW: Moving window by one year.

Under the moving window (MW) strategy, the following steps are adopted to get the Bühlmann credibility estimate $\hat{\boldsymbol{Y}}_{x,\,t_U+\tau}$ for $\tau \geqslant 2$:

- first, add the Bühlmann credibility estimates $\{\hat{\boldsymbol{Y}}_{x,\,t_U+1}, \ldots, \hat{\boldsymbol{Y}}_{x,\,t_U+\tau-1}\}$ to and remove $\{\hat{\boldsymbol{Y}}_{x,\,t_L+1}, \ldots, \hat{\boldsymbol{Y}}_{x,\,t_L+\tau-1}\}$ from $\{\boldsymbol{Y}_{x,\,t_L+1}, \ldots, \boldsymbol{Y}_{x,\,t_U}\}$, with $\hat{\boldsymbol{Y}}_{x,\,t} = \boldsymbol{Y}_{x,\,t}$ for $t \le t_U$, such that the fitting year span is moved by one year to $[t_L+\tau-1, t_U+\tau-1]$;

- next, obtain

$$\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) = \frac{1}{n-1}\sum_{t=t_L+\tau}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,\,t}, \tag{3.13}$$

$$\hat{\boldsymbol{\mu}}(t_U+\tau) = \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) = \frac{1}{m}\sum_{x=x_L}^{x_U}\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau), \tag{3.14}$$

and

$$\hat{\boldsymbol{Z}}(t_U+\tau) = \hat{\boldsymbol{A}}\left(\frac{1}{n-1}\hat{\boldsymbol{V}} + \hat{\boldsymbol{A}}\right)^{-1} \tag{3.15}$$

using the data in the year span $[t_L+\tau-1, t_U+\tau-1]$, where $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{A}}$ in $\hat{\boldsymbol{Z}}(t_U+\tau)$ are the same as those in $\hat{\boldsymbol{Z}}(t_U+1)$;

- finally, obtain the Bühlmann credibility estimates of $\hat{\boldsymbol{Y}}_{x,\,t_U+\tau}$ and $\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau})$ for year $t_U+\tau$ by

$$\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} = \hat{\boldsymbol{Z}}(t_U+\tau)\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) + \left[\boldsymbol{I}_r - \hat{\boldsymbol{Z}}(t_U+\tau)\right]\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau), \tag{3.16}$$

and

$$\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau}) = \ln(\boldsymbol{m}_{x,\,t_U}) + \sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}_{x,\,t_U+t} \tag{3.17}$$

for $\tau = 1, \ldots, T_2 - t_U$

One noteworthy property of the MW strategy is that $\boldsymbol{Z}(t_U + \tau)$ in Equation (3.16) is constant in τ, i.e., $\boldsymbol{Z}(t_U + \tau) = \boldsymbol{Z}(t_U + 1)$.

3.3.4 Some Properties

In this section, we prove $\hat{Y}_{x,\, t_U+\tau}$ is constant for $\tau = 1, 2, \ldots$ under the EW strategy with Propositions 1 to 3. Therefore, the non-parametric Bühlmann credibility estimate of $\ln(\boldsymbol{m}_{x,t_U+\tau})$ is a linear function of τ with slope parameter $\hat{\boldsymbol{Y}}_{x,t_U+1}$ and intercept parameter $\ln(\boldsymbol{m}_{x,\, t_U})$.

Proposition 1. *Under both EW and MW strategies, we have*

$$\frac{1}{m}\sum_{x=x_L}^{x_U} \hat{\boldsymbol{Y}}_{x,\, t_U+\tau} = \frac{1}{m}\sum_{x=x_L}^{x_U} \overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) = \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau), \quad \tau = 1,\, 2, \ldots. \tag{3.18}$$

Proof. From Equation (3.7), we get

$$\begin{aligned}\sum_{x=x_L}^{x_U} \hat{\boldsymbol{Y}}_{x,\, t_U+\tau} &= \boldsymbol{Z}(t_U+\tau)\sum_{x=x_L}^{x_U} \overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau) + [\boldsymbol{I}_r - \boldsymbol{Z}(t_U+\tau)]\sum_{x=x_L}^{x_U} \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) \\ &= \boldsymbol{Z}(t_U+\tau)\cdot m \cdot \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) + [\boldsymbol{I}_r - \boldsymbol{Z}(t_U+\tau)]\cdot m \cdot \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau).\end{aligned}$$

Dividing by m on both sides of the equation above gives

$$\frac{1}{m}\sum_{x=x_L}^{x_U} \hat{\boldsymbol{Y}}_{x,\, t_U+\tau} = \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) = \frac{1}{m}\sum_{x=x_L}^{x_U} \overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau), \quad \tau = 1,\, 2, \ldots.$$

□

Remark 2. *Proposition* 1 *implies that under both EW and MW strategies, the average of the non-parametric Bühlmann credibility estimates* $\hat{Y}_{x,\, t_U+\tau}$ *over the fitting age span* $[x_L, x_U]$ *equals the average of the* $\overline{Y}_{x,\bullet}(t_U+\tau)$ *over the same age span.*

Proposition 2. *Under the EW strategy, we have*

$$\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau) = \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1), \quad \tau = 2,\, 3, \ldots \tag{3.19}$$

Proof. We prove Equation (3.19) by induction on τ.

Base Case: When $\tau = 2$, by definition, and from Proposition 1 and Equation (3.8)

$$\begin{aligned}\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+2) &= \frac{1}{m}\sum_{x=x_L}^{x_U}\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+2)\\ &= \frac{1}{m\cdot n}\sum_{x=x_L}^{x_U}\left[\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,\,t}+\hat{\boldsymbol{Y}}_{x,\,t_U+1}\right]\\ &= \frac{n-1}{n}\cdot\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1)+\frac{1}{n}\cdot\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1)\\ &= \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1).\end{aligned}$$

Thus, Equation (3.19) holds for $\tau = 2$.

Induction Step: Suppose Equation (3.19) holds for $t = \tau$; then by definition, and from Proposition 2 and Equation (3.8) again,

$$\begin{aligned}\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau+1) &= \frac{1}{m}\sum_{x=x_L}^{x_U}\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+\tau+1)\\ &= \frac{1}{m\cdot(n+\tau-1)}\sum_{x=x_L}^{x_U}\left[\left(\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,\,t}+\sum_{t=t_U+1}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,\,t}\right)+\hat{\boldsymbol{Y}}_{x,\,t_U+\tau}\right]\\ &= \frac{n+\tau-2}{n+\tau-1}\cdot\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau)+\frac{1}{n+\tau-1}\cdot\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau)\\ &= \overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+\tau).\end{aligned}$$

Hence, Equation (3.19) holds for $t = \tau + 1$, and the induction step is complete.

Conclusion: By the principle of mathematical induction, it follows that Equation (3.19) holds for all $t = 2, 3, \ldots,$ under the EW strategy. $\square$

Proposition 3. *Under the EW strategy, we have*

$$\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} = \hat{\boldsymbol{Y}}_{x,\,t_U+1}, \quad \tau = 2, 3, \ldots, \tag{3.20}$$

and thus $\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau})$ *is a linear function of* τ *with slope parameter* $\hat{\boldsymbol{Y}}_{x,\,t_U+1}$ *and intercept parameter* $\ln(\boldsymbol{m}_{x,\,t_U})$*.*

Proof. For the EW strategy, from Equation (3.19), multiplying

$$\begin{aligned}(n+\tau-2)\hat{\boldsymbol{Z}}^{-1}(t_U+\tau) &= (n+\tau-2)\left(\frac{1}{n+\tau-2}\hat{\boldsymbol{V}}+\hat{\boldsymbol{A}}\right)\hat{\boldsymbol{A}}^{-1}\\ &= (n+\tau-2)\boldsymbol{I}_r+\hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\end{aligned}$$

on both sides of Equation (3.11) and using Equation (3.8) yields

$$
\begin{aligned}
&\left[(n+\tau-2)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} \\
= &\left[\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,\,t} + \sum_{t=t_U+1}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,\,t}\right] + (n+\tau-2)\left[\boldsymbol{Z}^{-1}(t_U+\tau) - \boldsymbol{I}_r\right]\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1) \\
= &\left[\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,\,t} + \sum_{t=t_U+1}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,\,t}\right] + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1).
\end{aligned}
$$

Similarly, the equation above with τ being replaced by $\tau+1$ gives

$$
\left[(n+\tau-1)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau+1} = \left[\sum_{t=t_L+1}^{t_U}\boldsymbol{Y}_{x,\,t} + \sum_{t=t_U+1}^{t_U+\tau-1}\hat{\boldsymbol{Y}}_{x,\,t} + \hat{\boldsymbol{Y}}_{x,\,t_U+\tau}\right] + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+1).
$$

The difference between the preceding two equations produces

$$
\begin{aligned}
&\left[(n+\tau-1)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau+1} - \left[(n+\tau-2)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} = \hat{\boldsymbol{Y}}_{x,\,t_U+\tau} \\
\Leftrightarrow\ &\left[(n+\tau-1)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau+1} = \left[(n+\tau-1)\boldsymbol{I}_r + \hat{\boldsymbol{V}}\hat{\boldsymbol{A}}^{-1}\right]\hat{\boldsymbol{Y}}_{x,\,t_U+\tau} \\
\Leftrightarrow\ &\hat{\boldsymbol{Y}}_{x,\,t_U+\tau+1} = \hat{\boldsymbol{Y}}_{x,\,t_U+\tau} \text{ for } \tau = 1,\, 2, \ldots.
\end{aligned}
$$

Therefore, the non-parametric Bühlmann credibility estimates of $\ln(\boldsymbol{m}_{x,\,t_U+\tau})$ for age x in year $t_U+\tau$ under the EW strategy is

$$
\ln(\hat{\boldsymbol{m}}_{x,\,t_U+\tau}) = \ln(\boldsymbol{m}_{x,\,t_U}) + \sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}_{x,\,t_U+t} = \ln(\boldsymbol{m}_{x,\,t_U}) + (\hat{\boldsymbol{Y}}_{x,\,t_U+1})\cdot\tau, \tag{3.21}
$$

which is a linear function of τ with slope parameter $\hat{\boldsymbol{Y}}_{x,t_U+1}$ and intercept parameter $\ln(\boldsymbol{m}_{x,\,t_U})$. □

3.3.5 Stochastic Mortality Rates and Semi-parametric Bühlmann Model

In this section, we first derive the formulas for the stochastic non-parametric Bühlmann estimate, $\ln(\boldsymbol{m}^s_{x,\,t_U+\tau})$. Next, we propose an alternative semi-parametric approach for getting the estimators of the structural parameters and the corresponding $\ln(\boldsymbol{m}^s_{x,\,t_U+\tau})$.

To get the stochastic non-parametric Bühlmann estimate, $\ln(\boldsymbol{m}^s_{x,\,t_U+\tau})$, for age x in year $t_U+\tau$, we need to add error terms to the corresponding deterministic non-parametric Bühlmann estimate, $\ln(\boldsymbol{m}_{x,\,t_U+\tau})$.

• For the EW strategy, from Equation (3.11) and Proposition 3 that $\hat{\boldsymbol{Y}}_{x,\,t_U+t} = \hat{\boldsymbol{Y}}_{x,\,t_U+1}$, we have

$$\begin{aligned}\hat{\boldsymbol{Y}}^s_{x,\,t_U+t} &= \hat{\boldsymbol{Z}}(t_U+t)\cdot[\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+t)+\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}]+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+t)]\cdot[\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+t)+\boldsymbol{\varepsilon}^{BC}_{t_U+t}]\\ &= \hat{\boldsymbol{Y}}_{x,\,t_U+1}+\left\{\hat{\boldsymbol{Z}}(t_U+t)\cdot\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+t)]\cdot\boldsymbol{\varepsilon}^{BC}_{t_U+t}\right\};\end{aligned}$$

and then

$$\sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}^s_{x,\,t_U+1}=(\hat{\boldsymbol{Y}}_{x,\,t_U+1})\cdot\tau+\sum_{t=1}^{\tau}\left\{\hat{\boldsymbol{Z}}(t_U+t)\cdot\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+t)]\cdot\boldsymbol{\varepsilon}^{BC}_{t_U+t}\right\}. \quad (3.22)$$

• For the MW strategy, from the fact that $\boldsymbol{Z}(t_U+t)=\boldsymbol{Z}(t_U+1)$ for $t=1,\,2,\ldots$, we have

$$\begin{aligned}\hat{\boldsymbol{Y}}^s_{x,\,t_U+t} &= \hat{\boldsymbol{Z}}(t_U+t)\cdot[\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+t)+\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}]+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+t)]\cdot[\overline{\boldsymbol{Y}}_{\bullet,\bullet}(t_U+t)+\boldsymbol{\varepsilon}^{BC}_{t_U+t}]\\ &= \hat{\boldsymbol{Y}}_{x,\,t_U+t}+\left\{\hat{\boldsymbol{Z}}(t_U+1)\cdot\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+1)]\cdot\boldsymbol{\varepsilon}^{BC}_{t_U+t}\right\};\end{aligned}$$

and then

$$\sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}^s_{x,\,t_U+t}=\sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}_{x,\,t_U+t}+\hat{\boldsymbol{Z}}(t_U+1)\sum_{t=1}^{\tau}\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}+[\boldsymbol{I}_r-\hat{\boldsymbol{Z}}(t_U+1)]\sum_{t=1}^{\tau}\boldsymbol{\varepsilon}^{BC}_{t_U+t}; \quad (3.23)$$

where

- $\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}$ follows an i.i.d. multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x}$ for $t=1,\ldots,\tau$, i.e., $\{\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}\}\overset{i.i.d.}{\sim}\mathcal{N}(\mathbf{0},\,\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x})$ for each fixed age x;
- $\boldsymbol{\varepsilon}^{BC}_{t_U+t}$ follows an i.i.d. multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}}$ for $t=1,\ldots,\tau$, i.e., $\{\boldsymbol{\varepsilon}^{BC}_{t_U+t}\}\overset{i.i.d.}{\sim}\mathcal{N}(\mathbf{0},\,\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}})$; and
- $\begin{bmatrix}\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}\\ \boldsymbol{\varepsilon}^{BC}_{t_U+t}\end{bmatrix}$ follows a multivariate normal distribution with mean $\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and covariance matrix $\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x} & \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x,\boldsymbol{\varepsilon}^{BC}}\\ \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x,\boldsymbol{\varepsilon}^{BC}} & \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}}\end{bmatrix}$, i.e.,

$$\left\{\begin{bmatrix}\boldsymbol{\varepsilon}^{BC}_{x,\,t_U+t}\\ \boldsymbol{\varepsilon}^{BC}_{t_U+t}\end{bmatrix}\right\}\overset{i.i.d.}{\sim}\mathcal{N}\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x} & \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x,\boldsymbol{\varepsilon}^{BC}}\\ \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x,\boldsymbol{\varepsilon}^{BC}} & \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}}\end{bmatrix}\right).$$

.

Note that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x}$, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}^{BC}_x,\boldsymbol{\varepsilon}^{BC}}$ are estimated by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}^{BC}_x}=\frac{1}{n-2}\sum_{t=t_L+1}^{t_U}[\boldsymbol{Y}_{x,\,t}-\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+1)][\boldsymbol{Y}_{x,\,t}-\overline{\boldsymbol{Y}}_{x,\bullet}(t_U+1)]',$$

$$\hat{\boldsymbol{\Sigma}}_{\varepsilon^{BC}} = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}[\overline{\boldsymbol{Y}}_{\bullet,\,t} - \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)][\overline{\boldsymbol{Y}}_{\bullet,\,t} - \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)]',$$

and

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}_x^{BC},\,\boldsymbol{\varepsilon}^{BC}} = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}[\boldsymbol{Y}_{x,\,t} - \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1)][\overline{\boldsymbol{Y}}_{\bullet,\,t} - \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)]',$$

where $\overline{\boldsymbol{Y}}_{\bullet,\,t} = \frac{1}{m}\sum\limits_{x=x_L}^{x_U}\boldsymbol{Y}_{x,\,t}$.

The stochastic semi-parametric Bühlmann estimate of $\ln(\boldsymbol{m}_{x,\,t_U+\tau})$ is therefore given by

$$\ln(\hat{\boldsymbol{m}}^s_{x,\,t_U+\tau}) \quad = \ln(\hat{\boldsymbol{m}}_{x,\,t_U}) + \sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}^s_{x,\,t_U+t}, \tag{3.24}$$

where $\sum_{t=1}^{\tau}\hat{\boldsymbol{Y}}^s_{x,\,t_U+t}$ is given in Equations (3.22) and (3.23) for the EW and MW strategies, respectively.

If $\boldsymbol{Y}_{x,\,t} = ([\ln(m_{x,\,t,\,1}) - \ln(m_{x,\,t-1,\,1})], \ldots, [\ln(m_{x,\,t,\,r}) - \ln(m_{x,\,t-1,\,r})])$, given $\boldsymbol{\Theta}_x$ for $t = t_L+1, \ldots, t_U$, follows an independent and identically multivariate normal distribution (see Figure 3.2) with mean $\boldsymbol{\Theta}_x^{BC}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_x^{BC}}$, then a semi-parametric Bühlmann credibility approach can be applied to estimate the structural parameters. The normality assumption of $Y_{x,\,t,\,i}$ for $x = 35$, 55, and 75 in six populations is supported by the Q-Q plots displayed in Figures 3.3 to 3.5. The procedure is exactly analogous to that under the non-parametric framework:

- estimate the hypothetical mean, $\boldsymbol{\mu}(\boldsymbol{\Theta}_x) = \mathrm{E}[\boldsymbol{Y}_{x,\,t}|\boldsymbol{\Theta}_x] = \boldsymbol{\Theta}_x^{BC}$, by

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\Theta}_x) = \hat{\boldsymbol{\Theta}}_x^{BC} = \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1) = \hat{\boldsymbol{\mu}}_x;$$

- estimate the process covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\Theta}_x) = \mathrm{Cov}[\boldsymbol{Y}_{x,\,t},\, \boldsymbol{Y}'_{x,\,t}|\boldsymbol{\Theta}_x] = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}_x^{BC}}$, by

$$\hat{\boldsymbol{\Sigma}}(\Theta_x) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}_x^{BC}} = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}[\boldsymbol{Y}_{x,\,t} - \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1)][\boldsymbol{Y}_{x,\,t} - \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1)]' = \hat{\boldsymbol{V}}_{\boldsymbol{x}};$$

- estimate the expected value of the hypothetical mean, $\boldsymbol{\mu} = \mathrm{E}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right]$, by

$$\hat{\boldsymbol{\mu}} = \frac{1}{m}\sum_{x=x_L}^{x_U}\hat{\boldsymbol{\Theta}}_x^{BC} = \frac{1}{m}\sum_{x=x_L}^{x_U}\hat{\boldsymbol{\mu}}_x = \frac{1}{m}\sum_{x=x_L}^{x_U}\overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1) = \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1);$$

- estimate the expected process covariance matrix, $\boldsymbol{V} = \mathrm{E}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Theta}_x\right)\right]$, by

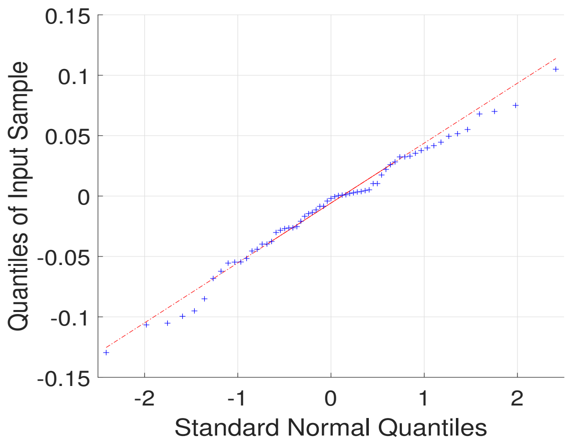
$$
\begin{aligned}
\hat{\boldsymbol{V}} &= \frac{1}{m} \sum_{x=x_L}^{x_U} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}_x^{BC}} \\
&= \frac{1}{m} \sum_{x=x_L}^{x_U} \left\{ \frac{1}{n-2} \sum_{t=t_L+1}^{t_U} [\boldsymbol{Y}_{x,\,t} - \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1)][\boldsymbol{Y}_{x,\,t} - \overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1)]' \right\} \\
&= \frac{1}{m} \sum_{x=x_L}^{x_U} \hat{\boldsymbol{V}}_x;
\end{aligned}
$$

- estimate the covariance matrix of hypothetical mean, $\boldsymbol{A} = \mathrm{Cov}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right] = \mathrm{E}\left\{\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right]^2\right\} - \left\{\mathrm{E}\left[\boldsymbol{\mu}\left(\boldsymbol{\Theta}_x\right)\right]\right\}^2$, by

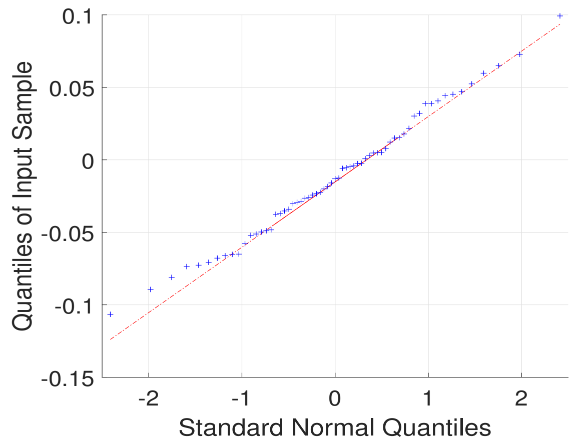
$$
\begin{aligned}
\hat{\boldsymbol{A}} &= \frac{1}{m} \sum_{x=x_L}^{x_U} [\hat{\boldsymbol{\Theta}}_x^{BC}][\hat{\boldsymbol{\Theta}}_x^{BC}]' - [\overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)][\overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)]' \\
&= \frac{1}{m} \sum_{x=x_L}^{x_U} [\overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1) - \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)][\overline{\boldsymbol{Y}}_{x,\,\bullet}(t_U+1) - \overline{\boldsymbol{Y}}_{\bullet,\,\bullet}(t_U+1)]'.
\end{aligned}
$$

It is worthwhile mentioning that the semi-parametric estimators for $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{V}}$ are the same as those under the non-parametric credibility approach, whereas the estimators for $\hat{\boldsymbol{A}}$ under the two approaches are different. The stochastic estimate $\ln(\boldsymbol{m}^s_{x,\,t_U+\tau})$ given in Equations (3.22) and (3.24) for the non-parametric Bühlmann approach still applies to the semi-parametric one.

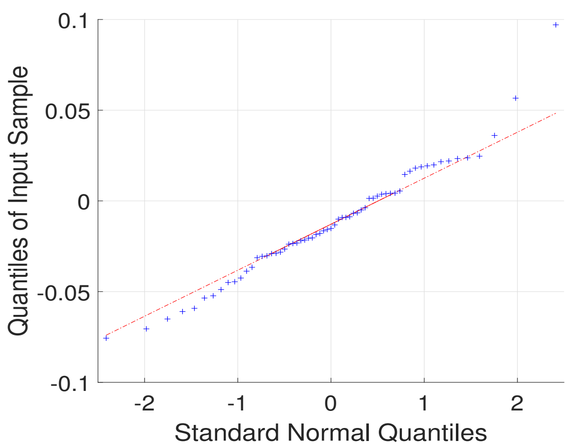
Figure 3.3: Q-Q plots of $Y_{x,t,i}$ for U.S.A. males and females



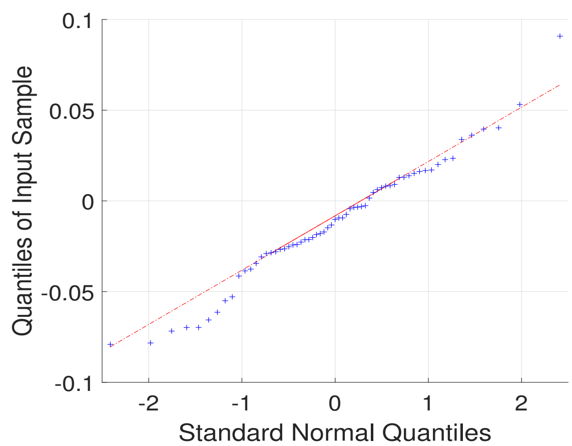
(a) U.S.A. Males; $x = 35$



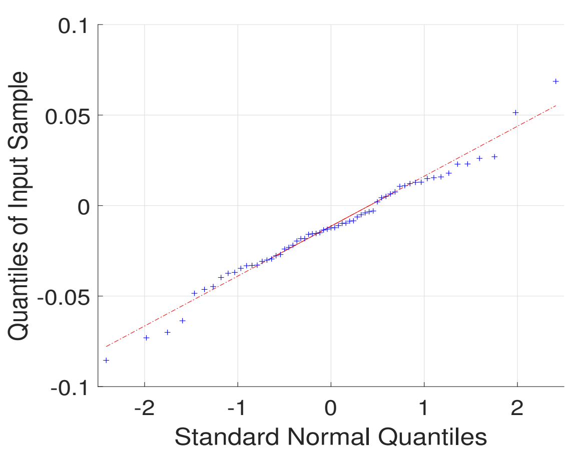
(b) U.S.A. Females; $x = 35$



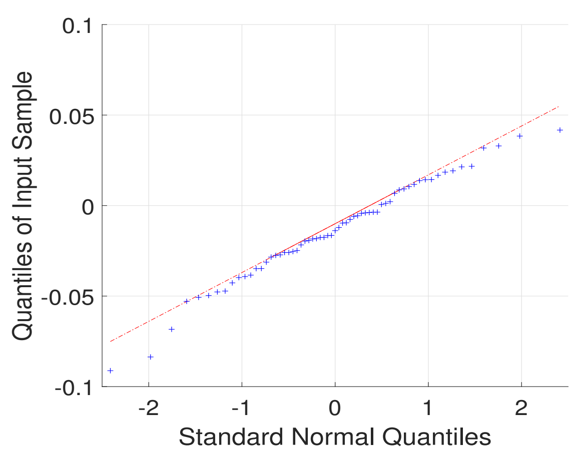
(c) U.S.A. Males; $x = 55$



(d) U.S.A. Females; $x = 55$

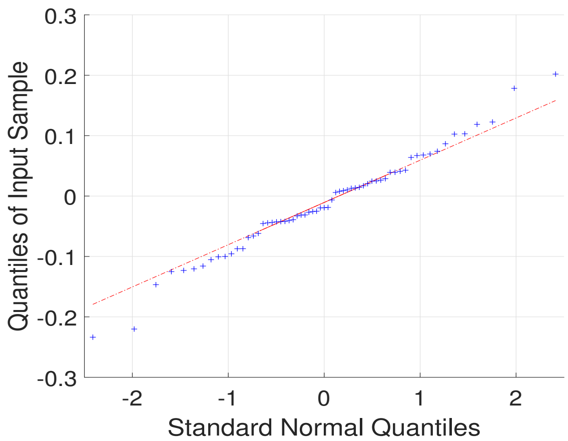


(e) U.S.A. Males; $x = 75$

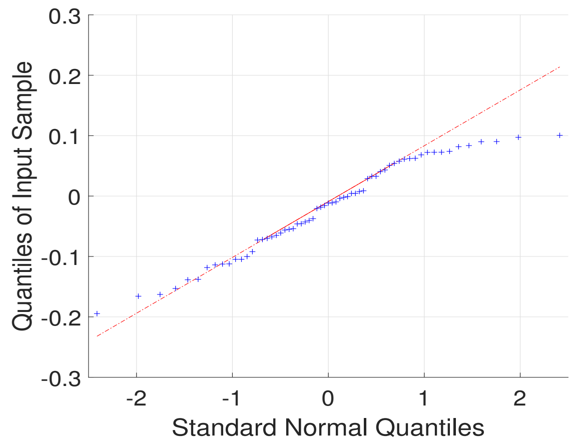


(f) U.S.A. Females; $x = 75$

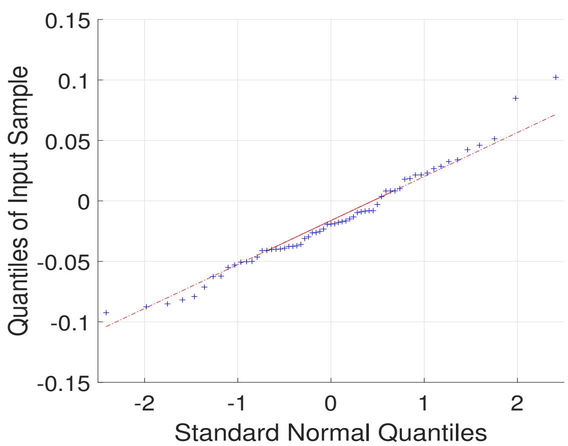
Figure 3.4: Q-Q plots of $Y_{x,t,i}$ for U.K. males and females



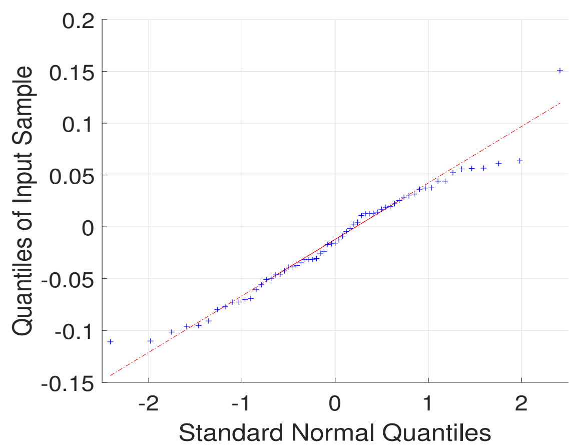
(a) U.K. Males; $x = 35$



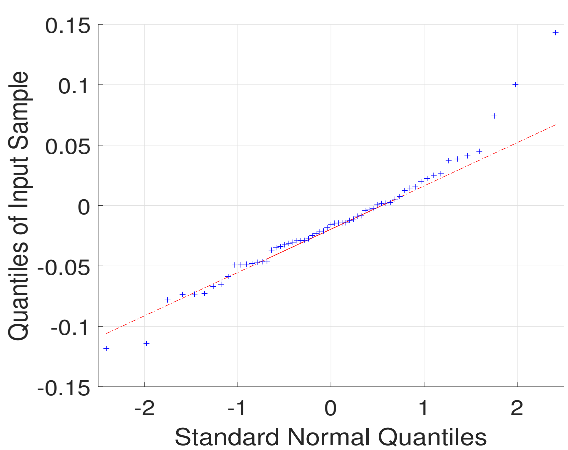
(b) U.K. Females; $x = 35$



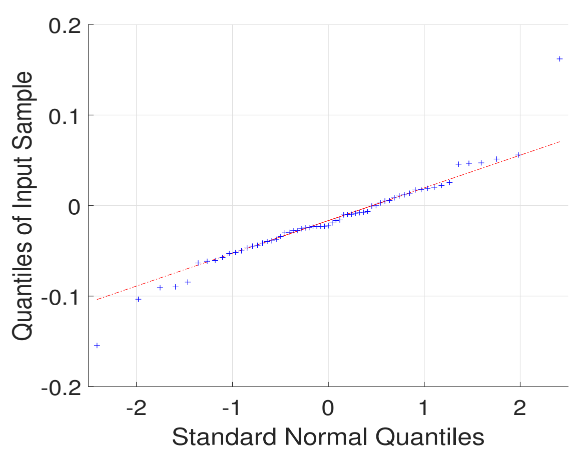
(c) U.K. Males; $x = 55$



(d) U.K. Females; $x = 55$

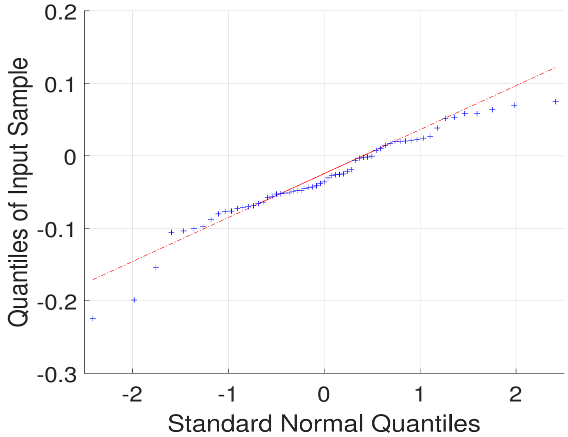


(e) U.K. Males; $x = 75$

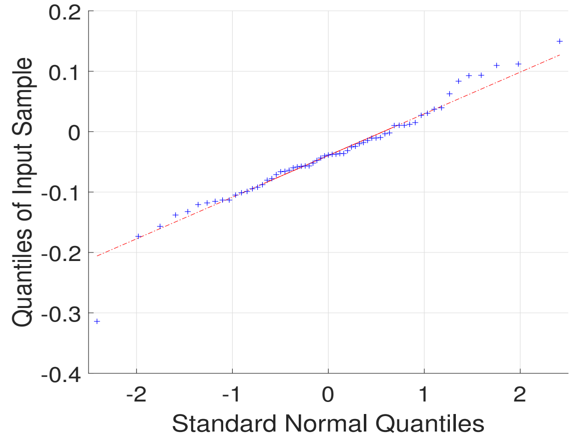


(f) U.K. Females; $x = 75$

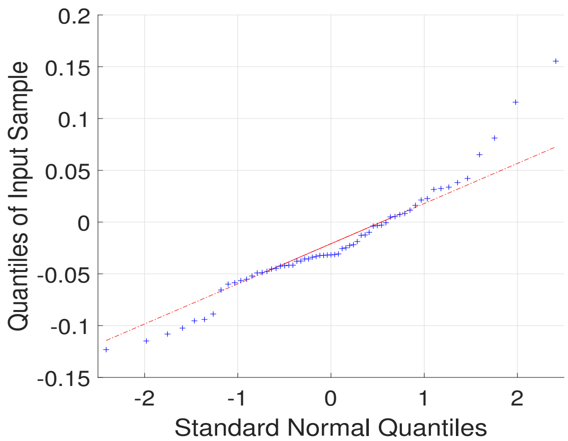
Figure 3.5: Q-Q plots of $Y_{x,t,i}$ for Japan males and females



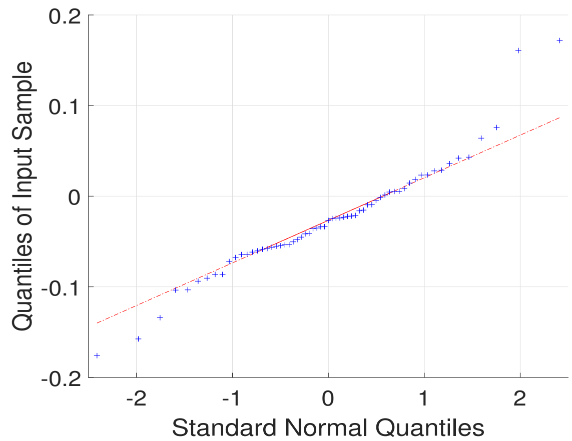
(a) Japan Males; $x = 35$



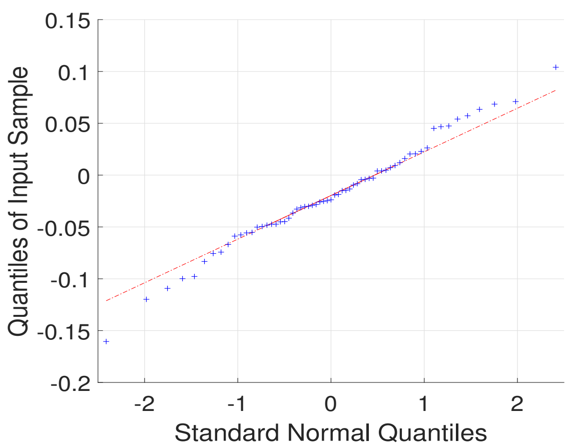
(b) Japan Females; $x = 35$



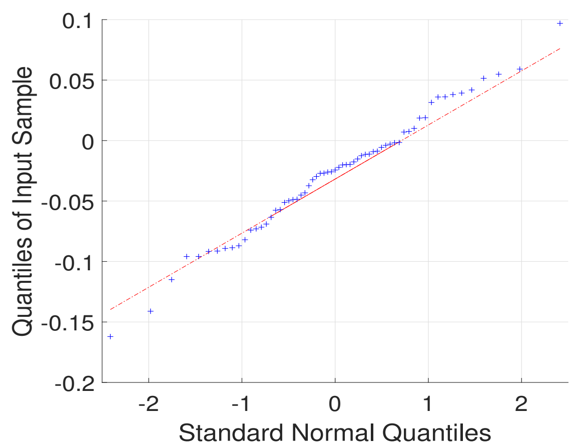
(c) Japan Males; $x = 55$



(d) Japan Females; $x = 55$



(e) Japan Males; $x = 75$



(f) Japan Females; $x = 75$

Chapter 4

Numerical Illustrations

In this chapter, the models presented in Chapter 3 are applied to mortality data of the Human Mortality Database for both genders of three well-developed countries; numerical results are provided for illustrations, which show that the multi-dimensional Bühlmann credibility approach outperforms the multi-dimensional Lee-Carter models based on the measure of mean absolute percentage error (MAPE).

This chapter is organized as follows. In Section 4.1, the parameters used for numerical illustrations are presented. Section 4.2 gives a detailed illustration of the MAPE (mean absolute percentage error), the statistical quantity for comparing the forecasting performances among models. In the last section, we present the numerical results of the models constructed in Chapter 3 with both visualized plots and summarized tables, followed by comparisons and analyses.

4.1 Model Specification

Given a study period $[T_1, T_2]$ for which mortality rates are available, we assume that we are currently at the end of year t_U, where $t_U < T_2$, and would like to forecast future mortality rates and evaluate the forecast performance of each mortality model for the forecasting period, $[t_U + 1, T_2]$.

The mortality rates of each population for the age-year rectangle $[x_L, x_U] \times [t_L, T_2]$, where $t_L \geqslant T_1$, are divided into two parts, the in-sample data, $[x_L, x_U] \times [t_L, t_U]$, and the out-of-sample data, $[x_L, x_U] \times [t_U + 1, T_2]$. The in-sample data, which consists of the mortality data in the first rectangle $[x_L, x_U] \times [t_L, t_U]$, are used in each model to estimate the parameters, and then the out-of-sample data, comprising the mortality data in the second rectangle $[x_L, x_U] \times [t_U + 1, T_2]$, are used for comparing with the predicted mortality rates. Specifically,

- for the study period, we adopt a 63-year period, $1951 - 2013$, that is, $T_1 = 1951$ and $T_2 = 2013$;
- for the fitting age span, we choose $25 - 84$, i,e., $x_L = 25$, an age of adults, $x_U = 84$, a life expectancy for some well-developed countries and $m = x_U - x_L + 1 = 60$;
- for the fitting year spans, we adopt a series of periods, $[t_L, t_U] = [1951, t_U], \ldots, [t_U - 4, t_U]$, with the shortest fitting year span being 5 years, to extensively evaluate the forecast performances of the underlying mortality models;
- for the forecasting year span, we choose three periods, $[t_U + 1, T_2] = [2004, 2013]$ (10 years window with $t_U = 2003$), $[1994, 2013]$ (20 years window with $t_U = 1993$) and $[1984, 2013]$ (30 years window with $t_U = 1983$).

Table 4.1: Summary of three forecasting year spans

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

We illustrate the proposed models with mortality rates from the Human Mortality database for four groups of populations:

- males and females of the U.S.A. ($r = 2$, with males of the U.S.A. as the base population for the co-integrated Lee-Carter model);
- males and females of the U.K. ($r = 2$, with males of the U.K. as the base population for the co-integrated Lee-Carter model);
- males and females of Japan ($r = 2$, with males of Japan as the base population for the co-integrated Lee-Carter model); and
- males and females of the U.S.A., the U.K. and Japan ($r = 6$, with males of the U.S.A. as the base population for the co-integrated Lee-Carter model).

4.2 Forecasting Errors

The forecasting performances of the proposed multi-dimensional non-parametric and semi-parametric Bühlmann credibility models are compared with those of existing joint-k/co-integrated/augmented common factor Lee-Carter models. The MAPE (mean absolute percentage error), a widely adopted forecasting performance measure as used in D'Amato et al. (2012) and Lin et al. (2015), is selected to measure the forecasting errors between the true (observed) out-of-sample mortality rate q and the estimated/forecasted one $\hat{q}$.

Specifically, with the out-of-sample data, the forecast accuracy of a single one-year death probability for age x in year $t_U + \tau$ and population i , $\hat{q}_{x,\, t_U+\tau,\, i}$, measured by the MAPE based on the fitting year span $[t_L, t_U]$ is given by

$$MAPE^{[t_L,t_U]}_{x,\, t_U+\tau,\, i} = \left| \frac{\hat{q}_{x,\, t_U+\tau,\, i} - q_{x,\, t_U+\tau,\, i}}{q_{x,\, t_U+\tau,\, i}} \right|. \tag{4.1}$$

For any given fitting year span $[t_L, t_U]$, to further examine the forecasting accuracy of a model for the entire rectangle $[x_L, x_U] \times [t_U + 1, T_2]$, the average of $MAPE^{[t_L,t_U]}_{x,\, t_U+\tau,\, i}$ over the rectangle is calculated as

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

where $m = x_U - x_L + 1 = 60$.

Then the values of all $AMAPE^{[t_L,t_U]}_{[t_U+1,T_2],\, i}$ over $t_L = T_1, \ldots, t_U - 4$ are summed up and the average, $AAMAPE_{[t_U+1,T_2],\, i}$, which is employed as a measurement of the overall performance of the underlying mortality model, is computed as

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

By comparing the $AAMAPE$, we are able to rank the prediction accuracy of different mortality models: the model which achieves a smaller $AAMAPE$ implies better forecasting performance.

4.3 Numerical Results

This section summarizes, with six figures and two tables, the forecast errors of one-year death probabilities for all models we discussed in Chapter 3 for four groups of populations.

- Table 4.2 presents $AAMAPE_{[t_U+1,\, 2013],\, i}$, the average of $AMAPE^{[t_L,\, t_U]}_{[t_U+1,\, 2013],\, i}$ over all fitting year spans for t_U = 2003, 1993 and 1983, for the first three groups of populations, i.e., males and females of the U.S.A. ($r = 2$), males and females of the U.K. ($r = 2$) and males and females of Japan ($r = 2$).
- Figures 4.1–4.3 display $AMAPE^{[t_L,\, t_U]}_{[t_U+1,\, 2013],\, i}$ against $t_L = 1951, \ldots, t_U - 4$ for t_U = 2003, 1993 and 1983, respectively, for the first three groups of populations.

- Table 4.3 gives $AAMAPE_{[t_U+1,\,2013],\,i}$, the average of $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013],\,i}$ over all fitting year spans for $t_U = 2003$, 1993 and 1983, for the last group of populations, i.e., males and females of the U.S.A., the U.K. and Japan ($r = 6$).

- Figures 4.4−4.6 exhibit $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013],\,i}$ against $t_L = 1951, \ldots, t_U - 4$ for $t_U =$ 2003, 1993 and 1983, respectively, for the last group of populations.

Informative observations are summarized as follows:

- The proposed semi-parametric and non-parametric Bühlmann approaches contribute to similar forecasting performances, with the non-parametric Bühlmann approach slightly outperforms the semi-parametric one. Due to this, we only plot $AMAPE$ for the EW and MW strategies under the non-parametric approach in Figures 4.1−4.6. As exhibited in Tables 4.2 and 4.3, the $AAMAPE_{[t_U+1,\,2013]}$ values under the semi-parametric and non-parametric Bühlmann approaches are quite similar for two strategies (EW and MW) and three forecast periods over four groups of populations. That is to say, the Bühlmann credibility approach under semi-parametric framework still achieves better forecasting performances than the multi-dimensional Lee-Carter models.

- In Figures 4.1 to 4.6, $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013],\,i}$ is neither monotonically increasing nor decreasing in t_L for all of the multi-dimensional Lee-Carter models and the EW and MW strategies. Stated differently, a shorter or longer fitting year span does not guarantee a lower AMAPE. As a matter of fact, the forecasting performance in terms of AMAPE values totally depends on the dataset. For example, Figure 4.2 (c) demonstrates that under the EW strategy, the smallest AMAPE for the U.K. males occurs at $t_L = 1979$ (the fitting year span $[1979, 1993]$), and after that AMAPE increases in t_L until the fitting year span narrows to the shortest one, $[1989, 1993]$. However, in Figure 4.2 (f) for Japan females, the AMAPE curve displays a decreasing pattern where the minimum AMAPE occurs at the shortest fitting year span $[1989, 1993]$.

- $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013],\,i}$ values under all of the models and strategies are generally decreasing in t_U, that is, the wider the forecasting period $[t_U + 1, 2013]$, the higher the $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013],\,i}$ value. From Tables 4.2 and 4.3, we observe that the AAMAPE values for all of the models and strategies for a wider forecasting period are on average higher than those for a narrower forecasting period. For example, from Table 4.3, for the 10-year forecasting period ($t_U = 2003$), the average of the $AMAPE^{[t_L,\,t_U]}_{[t_U+1,\,2013]}$ values over both genders of the three countries for the EW and MW strategies under the semi-parametric approach are 7.06% and 6.99%, whereas for the other two forecasting periods, the corresponding AAMAPE values are 12.05% and 11.84% for 20-year forecasting period ($t_U = 1993$) and 14.95% and 14.10% for 30-year forecasting period ($t_U = 1983$).

- Most of the $AMAPE_{[t_U+1,\,2013]}^{[t_L,\,t_U]}$ values for both of the MW and EW strategies under the non-parametric approach are lower than those for the joint-k, co-integrated and augmented common factor Lee-Carter models, except for a few cases. The $AMAPE$ EW or MW plot is almost located lowest in Figures 4.1 to 4.6. Moreover, in Table 4.3, for the 10-year forecasting period ($t_U = 2003$), the averages of the AMAPE values over both genders of the three countries for the joint-k, co-integrated and augmented common factor Lee-Carter models are 10.17% 9.32% and 8.85%, respectively, whereas those for the non-parametric EW and MW strategies are 7.10% and 7.05%. For the other two forecasting periods, the corresponding AAMAPE values for the non-parametric EW and MW strategies are 11.77% and 11.66% (14.58% and 13.98%) for $t_U = 1993$ ($t_U = 1983$), and 14.57%, 13.85% and 14.13% (19.05% 17.68% and 17.04%) with respect to the joint-k, co-integrated and augmented common factor Lee-Carter models. The numerical illustrations above give strong evidence that the non-parametric Bühlmann credibility approach outperforms the multi-dimensional Lee-Carter models in forecasting mortality rates.

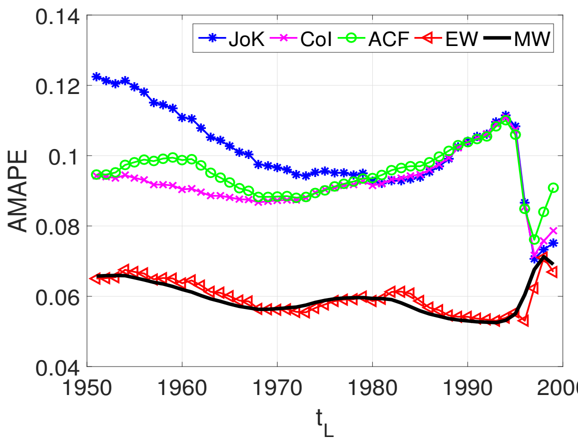
- In Figures 4.1 to 4.6, $AMAPE_{[t_U+1,\,2013]}^{[t_L,\,t_U]}$ curves for the MW strategy under non-parametric approach are smoother in t_L and generally lower than those for the EW strategy under same approach. We also observe in Table 4.2 and Table 4.3 that the MW strategy generally achieves lower AAMAPEs than the EW one. From this, we see that mortality models with different strategies under the same credibility approach can produce distinct responses in terms of their AAMAPE values.

To sum up, by comparing the forecasting performances across different mortality models, countries and genders, we can see that the EW or MW strategy generally produce the best forecasting performance for the wide age span 25 to 84, regardless of longer or shorter fitting year span and forecasting period. This indicates that the multi-dimensional Bühlmann credibility approach is an effective way to reduce forecasting errors. Out of the two proposed strategies, the MW strategy is slightly favourable than the EW one for the non-parametric and semi-parametric Bühlmann approaches.

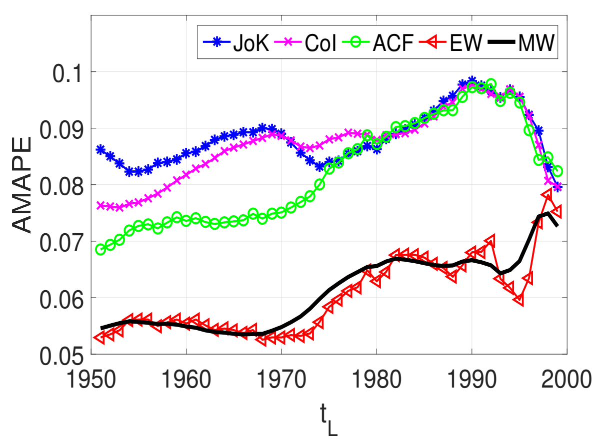
Table 4.2: $AAMAPE_{[t_U+1,\,2013],\,i}$s; two populations ($r = 2$)

		Credibility models				Lee-Carter models		
		Non		Semi				
Country	Gender	EW	MW	EW	MW	JoK	CoI	ACF
Panel A: $t_U = 2003$; forecasting $2004 - 2013$ (%)								
U.S.A.	Male	5.98	5.93	6.00	5.90	10.12	9.24	9.52
	Female	6.03	6.10	6.14	6.17	8.85	8.71	8.22
Average		6.00	6.01	6.07	6.03	9.48	8.97	8.87
U.K.	Male	9.08	8.93	8.43	8.31	11.74	11.17	11.07
	Female	7.92	7.94	7.82	7.85	8.51	8.90	8.85
Average		8.50	8.44	8.13	8.08	10.13	10.04	9.96
Japan	Male	5.55	5.64	5.83	5.86	7.32	7.50	7.44
	Female	8.03	7.75	8.03	7.72	9.42	9.58	9.92
Average		6.79	6.69	6.93	6.79	8.37	8.54	8.68
Panel B: $t_U = 1993$; forecasting $1994 - 2013$ (%)								
U.S.A.	Male	15.53	15.62	16.89	16.76	17.34	16.97	16.76
	Female	6.18	6.44	6.60	6.72	8.68	8.43	7.55
Average		10.86	11.03	11.75	11.74	13.01	12.70	12.16
U.K.	Male	15.31	15.21	13.67	13.31	17.19	16.95	16.88
	Female	9.29	9.58	10.21	10.25	12.95	13.05	12.93
Average		12.30	12.39	11.94	11.78	15.07	15.00	14.90
Japan	Male	9.93	9.48	10.75	10.27	12.93	12.47	12.40
	Female	14.36	13.62	14.28	13.51	15.54	15.55	15.96
Average		12.14	11.55	12.52	11.89	14.24	14.01	14.18
Panel C: $t_U = 1983$; forecasting $1984 - 2013$ (%)								
U.S.A.	Male	11.25	10.77	11.55	10.74	12.93	13.05	13.49
	Female	15.47	16.15	15.42	16.02	16.93	15.52	16.14
Average		13.36	13.46	13.49	13.38	14.93	14.28	14.81
U.K.	Male	19.59	18.25	20.32	18.65	23.07	22.92	22.77
	Female	9.08	8.63	11.86	10.44	16.24	16.34	16.47
Average		14.33	13.44	16.09	14.54	19.66	19.63	19.62
Japan	Male	13.20	12.19	13.58	12.32	17.44	16.45	16.29
	Female	18.86	17.87	19.06	17.90	21.31	20.92	22.23
Average		16.03	15.03	16.32	15.11	19.38	18.69	19.26

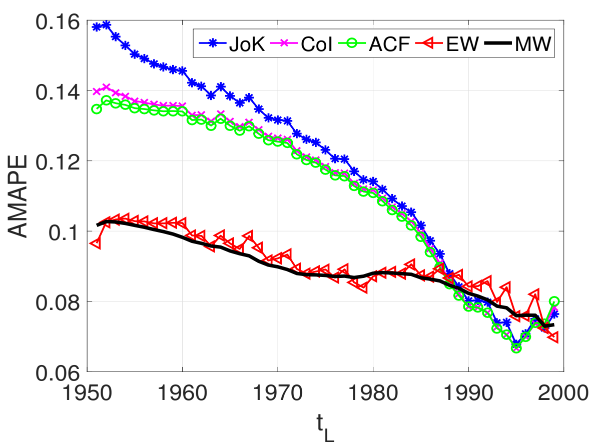
Figure 4.1: $AMAPE^{[t_L, 2003]}_{[2004, 2013], i}$ against t_L with age span $25 - 84$ (2 populations)



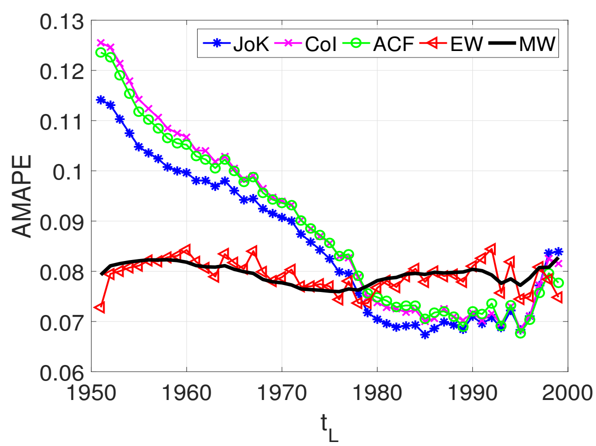
(a) U.S.A. Males



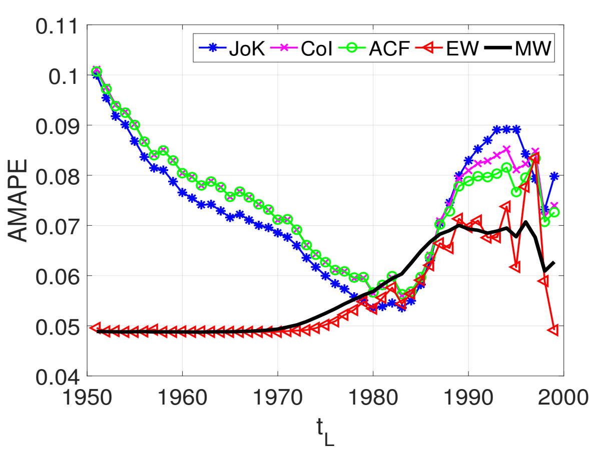
(b) U.S.A. Females



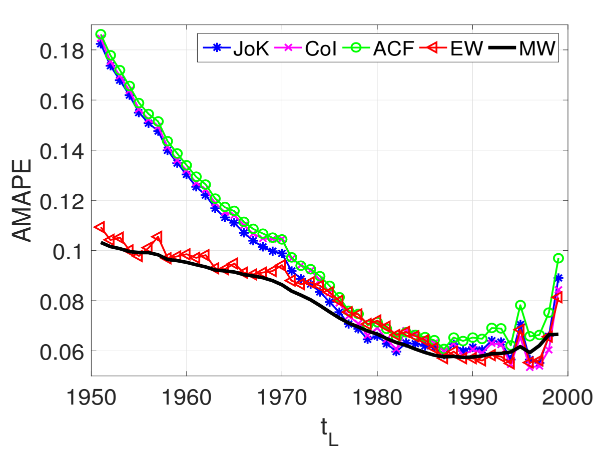
(c) U.K. Males



(d) U.K. Females

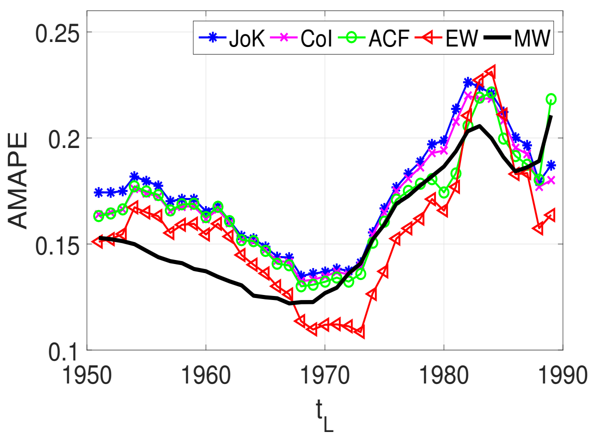


(e) Japan Males

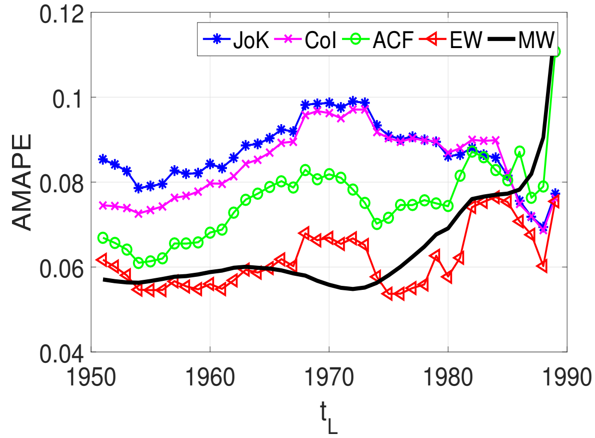


(f) Japan Females

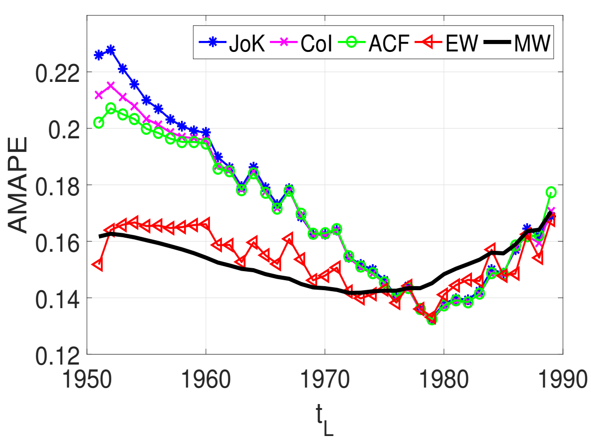
Figure 4.2: $AMAPE^{[t_L,\,1993]}_{[1994,\,2013],\,i}$ against t_L with age span $25-84$ (2 populations)



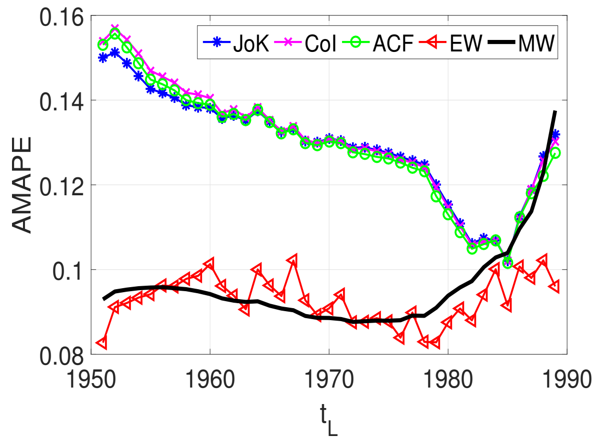
(a) U.S.A. Males



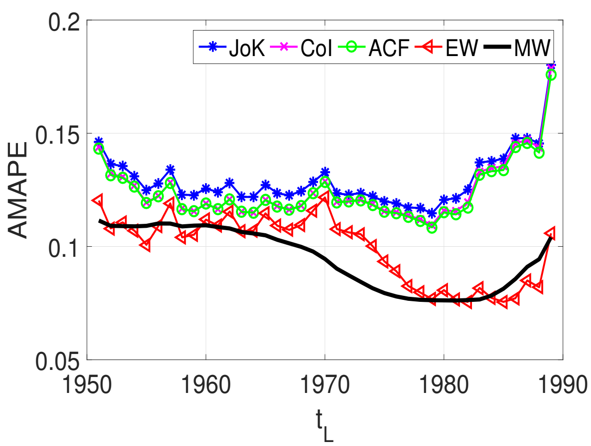
(b) U.S.A. Females



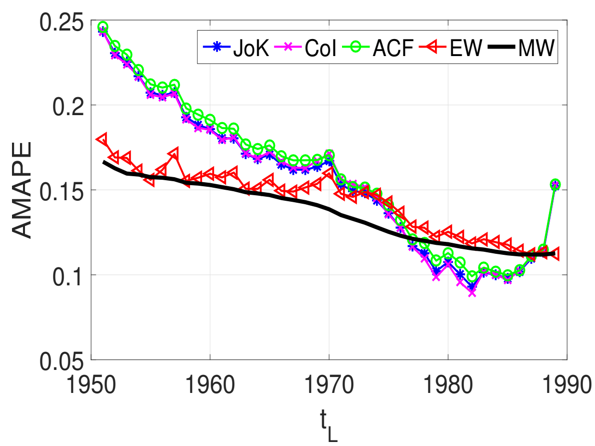
(c) U.K. Males



(d) U.K. Females

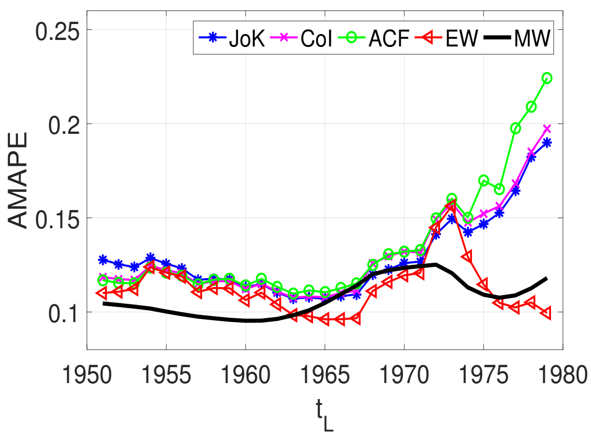


(e) Japan Males

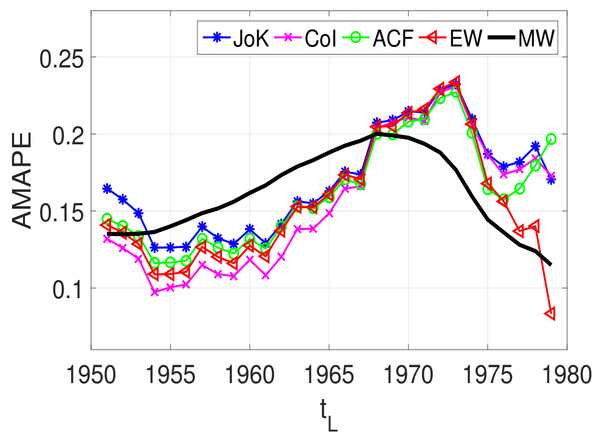


(f) Japan Females

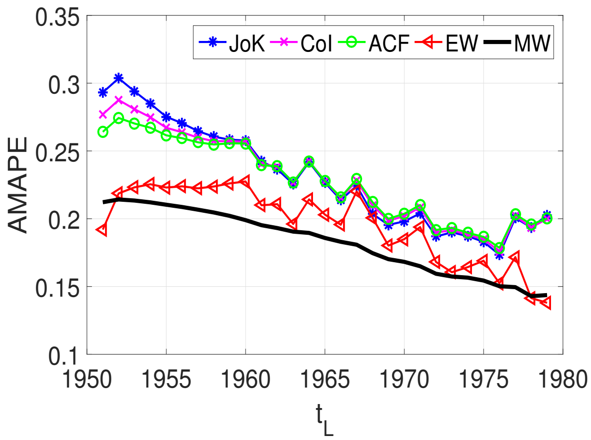
Figure 4.3: $AMAPE_{[1984,\,2013],\,i}^{[t_L,\,1983]}$ against t_L with age span $25-84$ (2 populations)



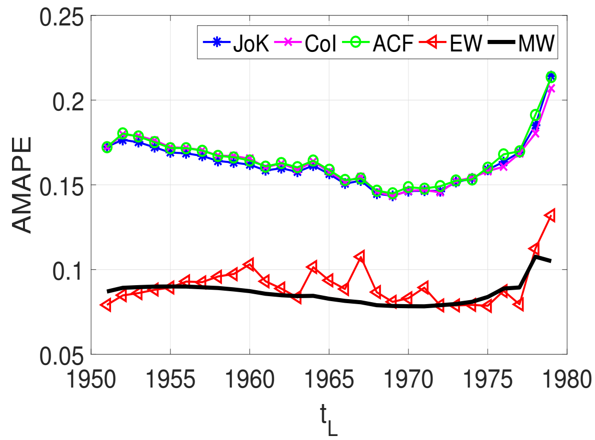
(a) U.S.A. Males



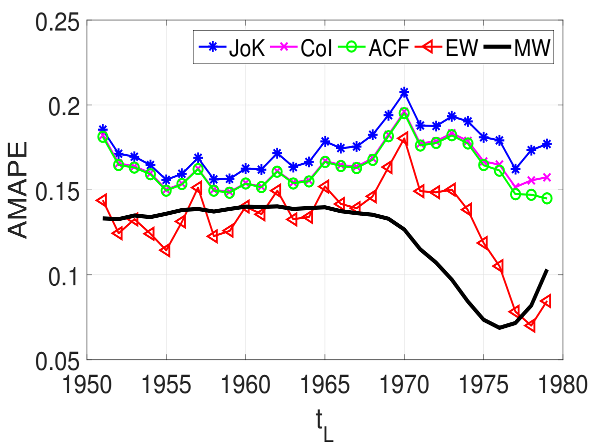
(b) U.S.A. Females



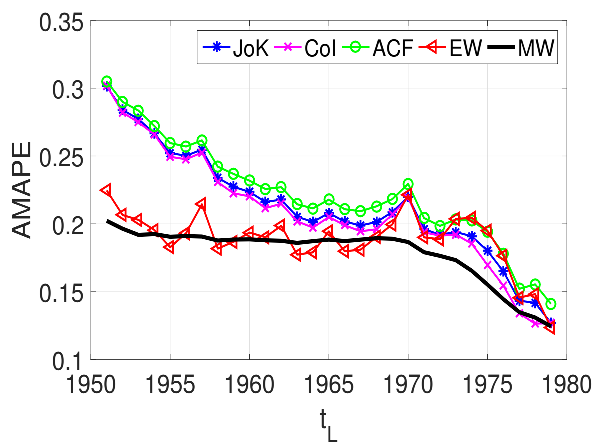
(c) U.K. Males



(d) U.K. Females



(e) Japan Males

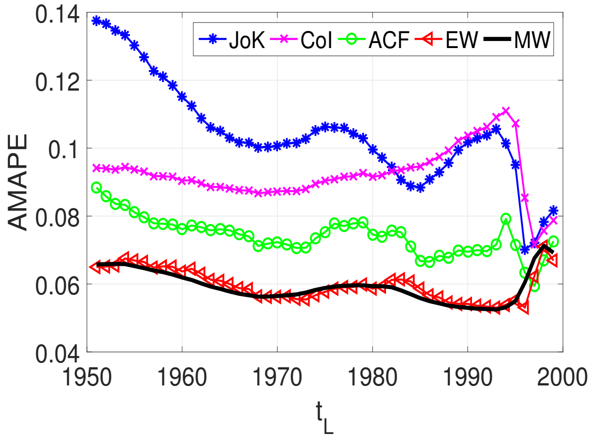


(f) Japan Females

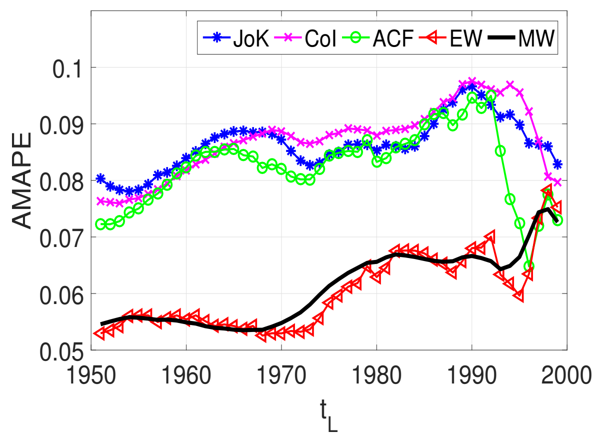
Table 4.3: $AAMAPE_{[t_U+1,\,2013],\,i}$s; six populations ($r = 6$)

		Credibility models				Lee-Carter models		
		Non		Semi				
Country	Gender	EW	MW	EW	MW	JoK	CoI	ACF
Panel A: $t_U = 2003$; forecasting $2004 - 2013$ (%)								
U.S.A.	Male	5.98	5.93	6.01	5.91	10.42	9.24	7.41
	Female	6.03	6.10	6.12	6.15	8.64	8.71	8.25
U.K.	Male	9.08	8.93	8.60	8.48	13.81	11.49	13.04
	Female	7.92	7.94	7.83	7.87	9.78	9.31	7.87
Japan	Male	5.55	5.64	5.77	5.82	7.69	7.89	6.37
	Female	8.03	7.75	8.02	7.73	10.67	9.25	10.14
Average		7.10	7.05	7.06	6.99	10.17	9.32	8.85
Panel B: $t_U = 1993$; forecasting $1994 - 2013$ (%)								
U.S.A.	Male	15.57	15.63	16.75	16.64	17.09	16.97	18.66
	Female	6.18	6.44	6.47	6.60	9.77	8.43	8.97
U.K.	Male	15.31	15.21	14.59	14.54	18.87	18.22	19.34
	Female	9.29	9.58	9.60	9.76	13.67	13.75	10.69
Japan	Male	9.93	9.48	10.43	9.81	12.72	11.16	10.91
	Female	14.36	13.62	14.43	13.66	15.33	14.58	16.21
Average		11.77	11.66	12.05	11.84	14.57	13.85	14.13
Panel C: $t_U = 1983$; forecasting $1984 - 2013$ (%)								
U.S.A.	Male	11.25	10.77	11.46	10.68	13.71	13.05	12.07
	Female	15.47	16.15	15.41	16.04	17.20	15.52	16.25
U.K.	Male	19.59	18.25	19.90	18.36	24.87	24.29	23.68
	Female	9.08	8.63	10.05	9.13	16.93	17.07	13.49
Japan	Male	13.20	12.18	13.62	12.31	18.82	15.95	14.55
	Female	18.87	17.88	19.29	18.06	22.79	20.21	22.21
Average		14.58	13.98	14.95	14.10	19.05	17.68	17.04

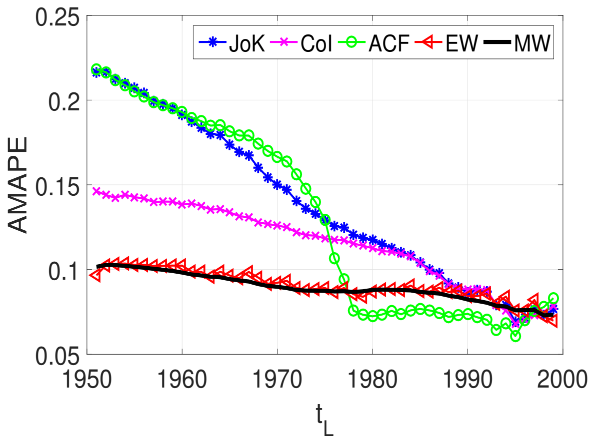
Figure 4.4: $AMAPE^{[t_L,\,2003]}_{[2004,\,2013],\,i}$ against t_L with age span $25-84$ (6 populations)



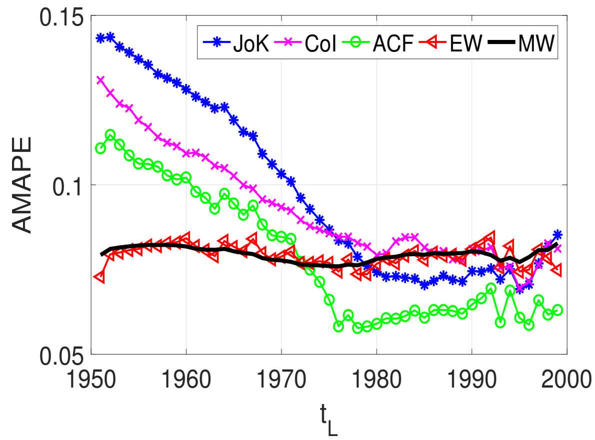
(a) U.S.A. Males



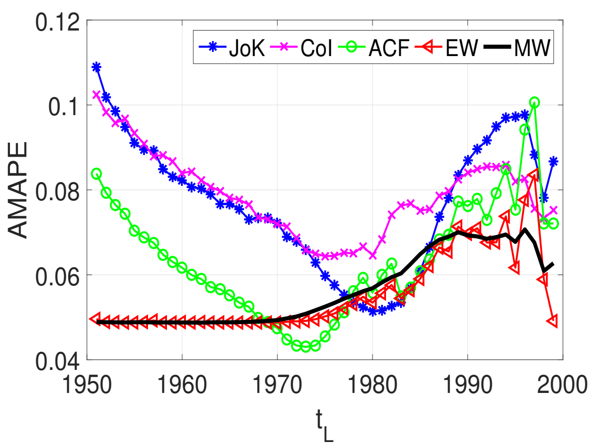
(b) U.S.A. Females



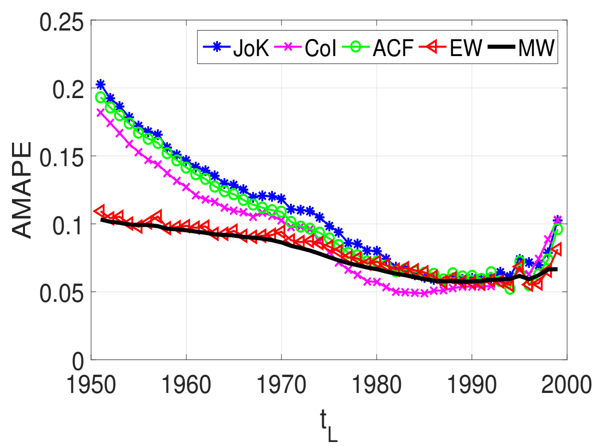
(c) U.K. Males



(d) U.K. Females

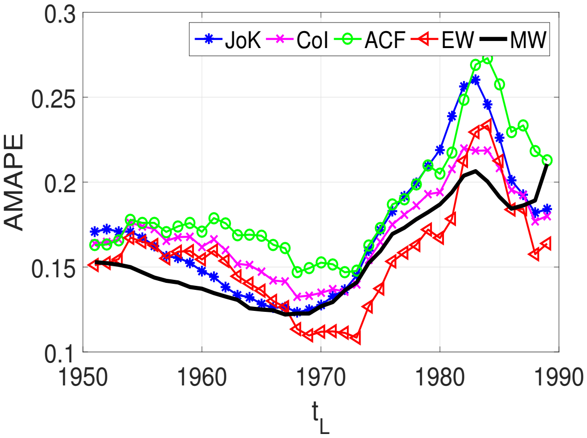


(e) Japan Males

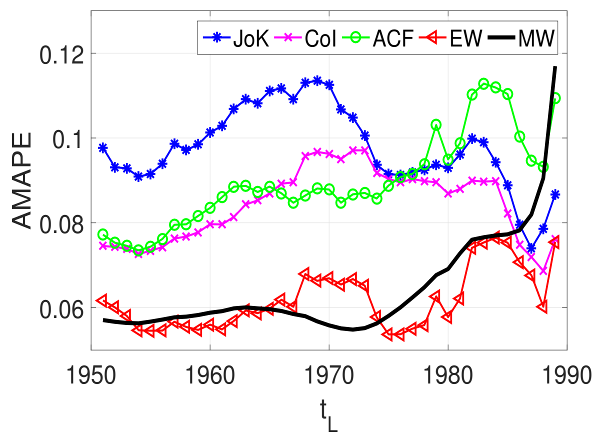


(f) Japan Females

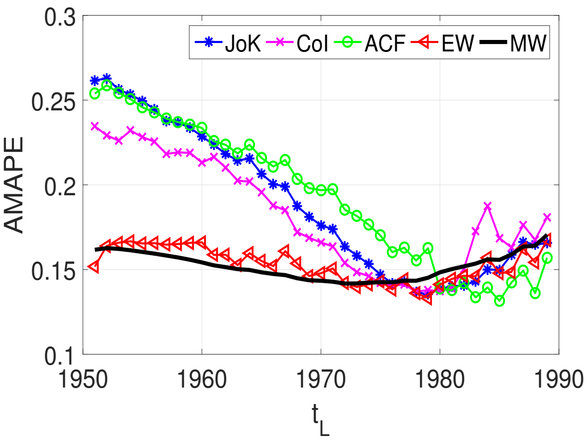
Figure 4.5: $AMAPE^{[t_L,\,1993]}_{[1994,\,2013],\,i}$ against t_L with age span $25-84$ (6 populations)



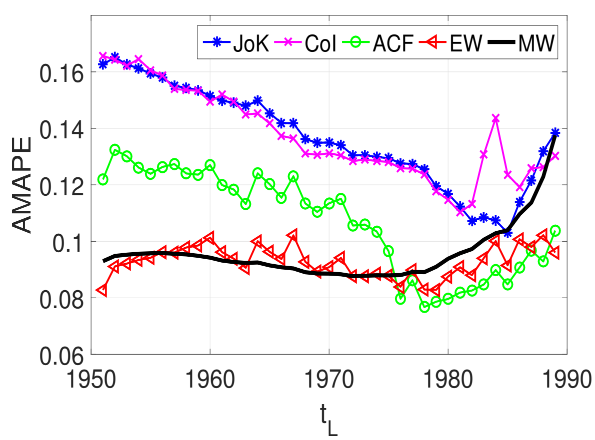
(a) U.S.A. Males



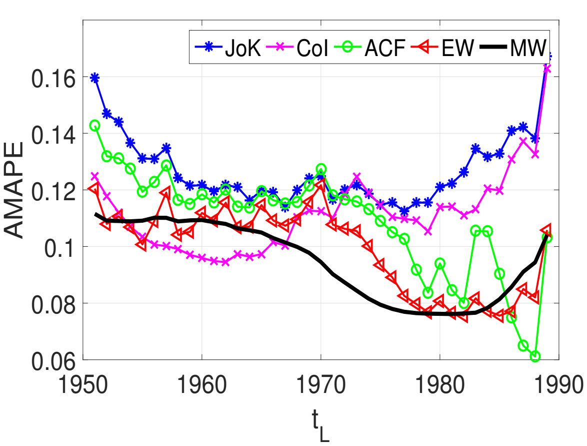
(b) U.S.A. Females



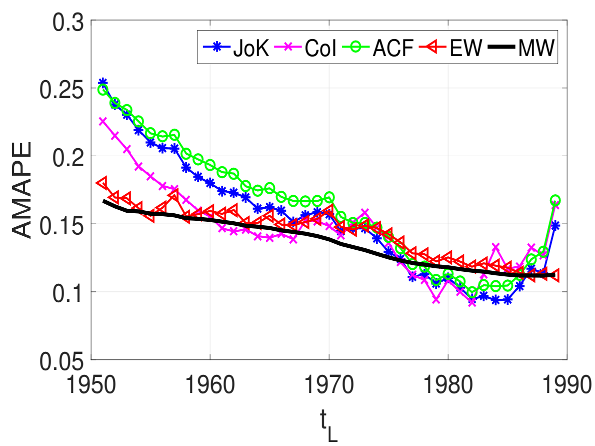
(c) U.K. Males



(d) U.K. Females

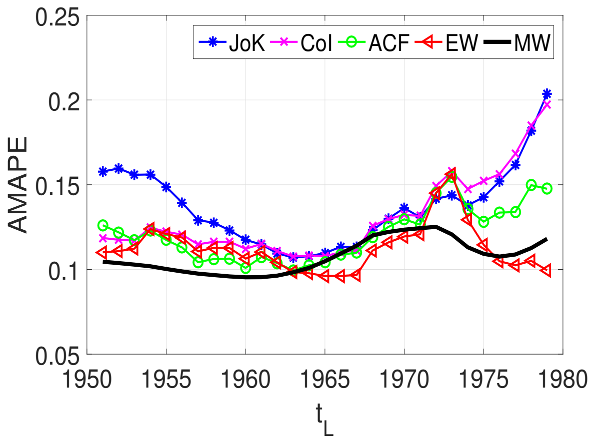


(e) Japan Males

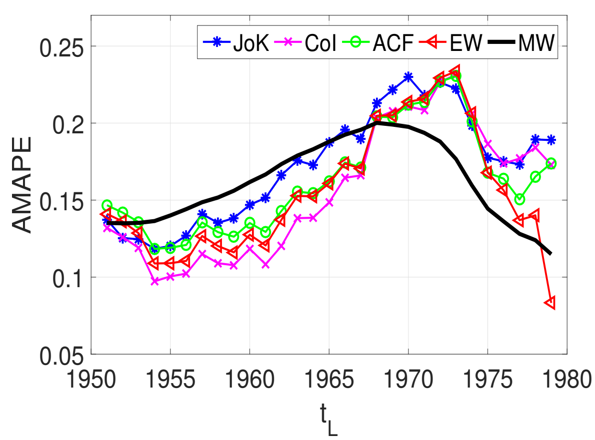


(f) Japan Females

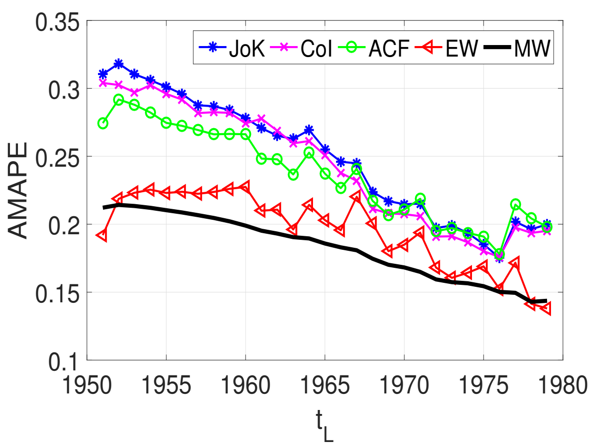
Figure 4.6: $AMAPE^{[t_L,\,1983]}_{[1984,\,2013],\,i}$ against t_L with age span $25 - 84$ (6 populations)



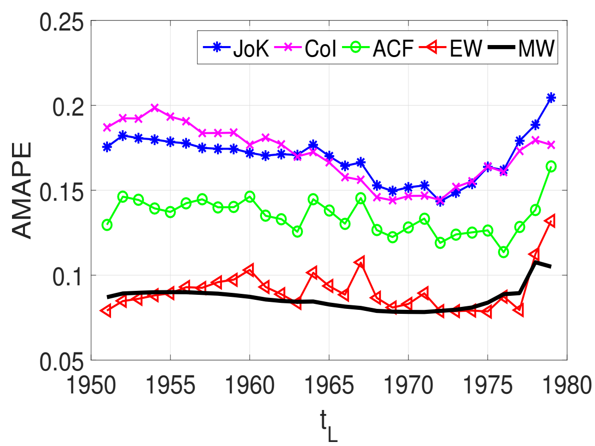
(a) U.S.A. Males



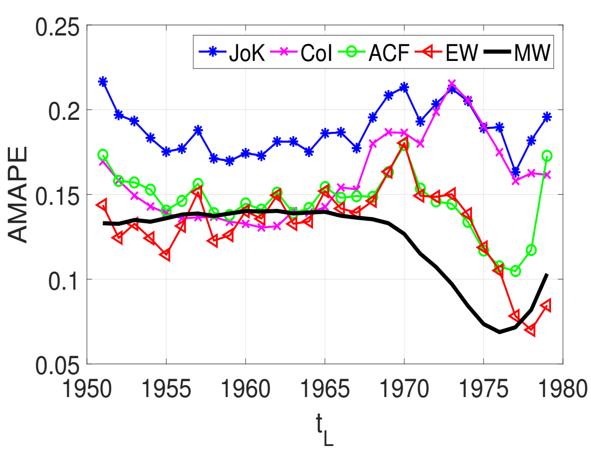
(b) U.S.A. Females



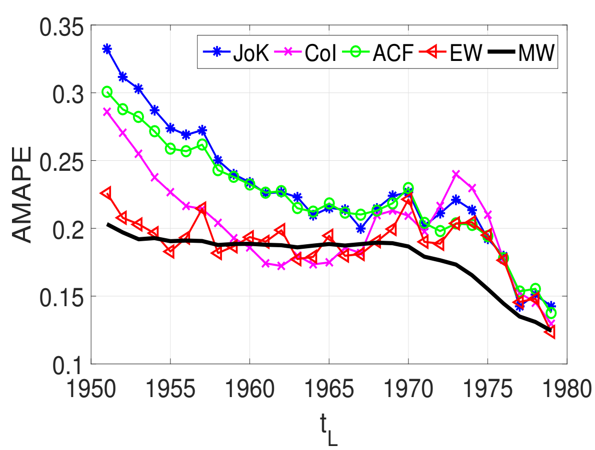
(c) U.K. Males



(d) U.K. Females



(e) Japan Males



(f) Japan Females

Chapter 5

Conclusion

In this project we employ a multi-dimensional Bühlmann credibility approach to modeling the mortality dynamics of multiple populations jointly under both non-parametric and semi-parametric frameworks. We further propose two strategies for forecasting mortality levels for two or more years; one is expanding mortality data window by one year (EW strategy) and the other is moving mortality data window by one year (MW strategy). Moreover, the formulas for calculating corresponding stochastic mortality rates are also provided in this project to simulate future mortality rates for applications and construct predictive intervals.

We also give an informative credibility interpretation that the future decrement mortality rate per year for age x is the weighted average of the the sample mean of the past decrement rates per year for the individual time trend for x and that of the past decrement rates per year for the group time trend for all ages. We prove that credibility forecasts of the natural logarithm of the central death rate for age x in year $t_U + \tau$ under the EW strategy is a linear functions of τ with slope parameter $\hat{\boldsymbol{Y}}_{x,t_U+1}$, the decrement mortality rate for age x and the first forecast year, and intercept parameter $\ln(\boldsymbol{m}_{x,t_U})$, the natural logarithm of the true (observed) central death rate for age x and the last fitting year.

Three well-known multi-population extensions of the Lee-Carter model and credibility models proposed in this project are applied to mortality data of the Human Mortality Database for both genders of the U.S.A, the U.K. and Japan. Empirical results show that the credibility approach generally produces better forecasting performances, measured by the MAPE, than those based on the Lee-Carter model. In addition to providing a simple approach that produces satisfactory and stable forecasts, another major contribution of this project is that it innovatively applies the multi-dimensional credibility approach, an approach that is predominantly used in property and casualty insurance, to mortality fitting and forecasting for multiple populations.

Appendix A

Single Population Lee-Carter Model

A.1 The Model

The natural logarithm of central death rates, $\ln(m_{x,t})$, for lives aged x in year t under the well-known Lee-Carter model (see Lee and Carter, 1992) is expressed by:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \times k_t + \varepsilon_{x,t}, \quad x = x_L, \ldots, x_U, \quad t = t_L, \ldots, t_U,$$

where

- α_x is the average age-specific mortality factor at age x,
- k_t is the index of the mortality level in year t,
- β_x is the age-specific reaction to k_t at age x, and
- $\varepsilon_{x,t}$ is the model error and $\{\varepsilon_{x,t} : t = t_L, t_L + 1, \ldots\} \overset{i.i.d.}{\sim} N(0, \sigma^2_{\varepsilon_x})$.

For uniqueness of the model specification, the following two constraints are imposed:

$$\sum_{x=x_L}^{x_U} \beta_x = 1 \text{ and } \sum_{t=t_L}^{t_U} k_t = 0.$$

A.2 Fitting the Model

According to the two constrains, estimates of α_x, k_t and β_x can be obtained as follows:

- $\hat{\alpha}_x$ can be derived by averaging the sum of $\ln(m_{x,t})$ over the fitting year span $[t_L, t_U]$,

$$\sum_{t=t_L}^{t_U} \ln(m_{x,t}) = n \times \alpha_x + \beta_x \times \sum_{t=t_L}^{t_U} k_t = n \times \alpha_x$$
$$\Rightarrow \hat{\alpha}_x = \frac{\sum\limits_{t=t_L}^{t_U} \ln(m_{x,t})}{n}, \quad x = x_L, \ldots, x_U,$$

  where $n = t_U - t_L + 1$;

- $\hat{k}_t$ is equal to the sum of $[\ln(m_{x,t}) - \hat{\alpha}_x]$ over the age span $[x_L, x_U]$,

$$\sum_{x=x_L}^{x_U} [\ln(m_{x,t}) - \hat{\alpha}_x] = k_t \times \sum_{x=x_L}^{x_U} \hat{\beta}_x = k_t \times 1$$
$$\Rightarrow \hat{k}_t = \sum_{x=x_L}^{x_U} [\ln(m_{x,t}) - \hat{\alpha}_x], \quad t = t_L, \ldots, t_U;$$

- $\hat{\beta}_x$ can be obtained by regressing $[\ln(m_{x,t}) - \hat{\alpha}_x]$ on $\hat{k}_t$ without the constant term for each age x; specifically,

$$\hat{\beta}_x = \sum_{t=t_L}^{t_U} \frac{[\ln(m_{x,t}) - \hat{\alpha}_x] \times k_t}{(\hat{k}_t)^2}.$$

A.3 Modeling and Forecasting the Mortality Level Index k_t

Empirical analyses (see Lee and Carter, 1992) shown that $\{\hat{k}_t : t = t_L, \ldots, t_U\}$ displays a linear trend, so we apply the ARIMA(0,1,0) times series model, a random walk with drift θ, to model $\hat{k}_t$ by

$$\hat{k}_t = \hat{k}_{t-1} + \theta + \epsilon_t,$$

where

- the time trend error $\epsilon_t \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$, that is, $\hat{k}_t - \hat{k}_{t-1} \overset{i.i.d.}{\sim} N(\theta, \sigma_\epsilon^2)$, for all t.

- the time trend errors, $\{\epsilon_t\}$, are assumed to be independent of the model errors, $\{\varepsilon_{x,t}\}$.

The unbiased estimator of the mean of the i.i.d. $(\hat{k}_t - \hat{k}_{t-1})$, θ, is given by

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

Thus, $k_{t_U+\tau}$ can be deterministically projected as $\hat{k}_{t_U+\tau} = \hat{k}_{t_U} + \tau \cdot \hat{\theta}$, and stochastically projected as $\tilde{k}_{t+\tau} = \hat{k}_{t+\tau} + \sum_{t=t_U+1}^{t_U+\tau} \epsilon_t$, where $\sum_{t=t_U+1}^{t_U+\tau} \epsilon_t$ has the same distribution as $\sqrt{\tau} \times \epsilon_{t_U+\tau}$.

A.4 Forecasting Mortality Rates

The natural logarithm of the deterministic central death rates, $\ln(m_{x,\,t_U+\tau})$, for lives aged x in year $t_U + \tau$ can be predicted as

$$\begin{aligned}
\ln(\hat{m}_{x,\,t_U+\tau}) &= \hat{\alpha}_x + \hat{\beta}_x \times \hat{k}_{t_U+\tau} \\
&= \hat{\alpha}_x + \hat{\beta}_x \times (\hat{k}_{t_U} + \tau \times \hat{\theta}) \\
&= (\hat{\alpha}_x + \hat{\beta}_x \times \hat{k}_{t_U}) + \hat{\beta}_x \times \tau \times \hat{\theta} \\
&= \ln(\hat{m}_{x,\,t_U}) + (\hat{\beta}_x \times \hat{\theta}) \times \tau, \quad \tau = 1, \ldots, T_2 - t_U,
\end{aligned}$$

where $\ln(\hat{m}_{x,\,t}) = \hat{\alpha}_x + \hat{\beta}_x \times \hat{k}_{t_U}$.

It follows that

$$\hat{m}_{x,\,t_U+\tau} = \exp\left[\hat{\alpha}_x + \hat{\beta}_x \times (\hat{k}_{t_U} + \tau \times \hat{\theta})\right].$$

By Equation (3.1), the predicted deterministic one-year death rate, $\hat{q}_{x,\,t_U+\tau}$, for age x in year $t_U + \tau$, is given by

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

We can add two error terms, the model error $\varepsilon_{x,\,t_U+\tau}$ and the time trend errors ϵ_{t_U+t}, $t = 1, \ldots, \tau$, to the natural logarithm of the predicted deterministic central death rates, $\ln(\hat{m}_{x,\,t_U+\tau})$, to form the natural logarithm of the stochastic central death rates, $\ln(\tilde{m}_{x,\,t_U+\tau})$, for age x in year $t_U + \tau$. Specifically,

$$\begin{aligned}
\ln(\tilde{m}_{x,\,t_U+\tau}) &= \hat{\alpha}_x + \hat{\beta}_x \times \tilde{k}_{t_U+\tau} + \varepsilon_{x,\,t_U+\tau} \\
&= \hat{\alpha}_x + \hat{\beta}_x \times (\hat{k}_{t_U} + \tau \times \hat{\theta} + \sum_{t=t_U+1}^{t_U+\tau} \epsilon_t) + \varepsilon_{x,\,t_U+\tau} \\
&= \ln(\hat{m}_{x,\,t_U+\tau}) + \sqrt{\tau} \times \hat{\beta}_x \times \epsilon_{t_U+\tau} + \varepsilon_{x,\,t_U+\tau},
\end{aligned}$$

which follows that $\ln(\tilde{m}_{x,\,t_U+\tau}) \sim N\left(\ln(\hat{m}_{x,\,t_U+\tau}),\, \tau \times \hat{\beta}_x^2 \times \sigma_\epsilon^2 + \sigma_{\varepsilon_x}^2)\right)$ and

$$\tilde{m}_{x,\,t_U+\tau} = \exp\left(\ln(\hat{m}_{x,\,t_U+\tau}) + \sqrt{\tau} \times \hat{\beta}_x \times \epsilon_{t_U+\tau} + \varepsilon_{x,\,t_U+\tau}\right).$$

From Equation (3.1), the stochastic one-year death rate, $\tilde{q}_{x,\,t_U+\tau}$, for age x in year $t_U+\tau$ is given by

$$\tilde{q}_{x,\,t_U+\tau} = 1 - \exp\left[-\exp\left(\ln(\hat{m}_{x,\,t_U+\tau}) + \sqrt{\tau} \times \hat{\beta}_x \times \epsilon_{t_U+\tau} + \varepsilon_{x,\,t_U+\tau}\right)\right].$$

The estimate of the variance of the model error, $\hat{\sigma}^2_{\varepsilon_x}$, is obtained by

$$\hat{\sigma}^2_{\varepsilon_x} = \frac{1}{n-2}\sum_{t=t_L}^{t_U}(\varepsilon_{x,\,t})^2 = \frac{\sum_{t=t_L}^{t_U}\left[\ln(m_{x,\,t}) - \hat{\alpha}_x - \hat{\beta}_x \times \hat{k}_t\right]^2}{n-2},$$

and the estimate of the variance of the time trend error, $\hat{\sigma}^2_{\epsilon}$, is given by

$$\hat{\sigma}^2_{\epsilon} = \frac{1}{n-2}\sum_{t=t_L+1}^{t_U}(\epsilon_t)^2 = \frac{\sum_{t=t_L+1}^{t_U}\left(\hat{k}_t - \hat{k}_{t-1} - \hat{\theta}\right)^2}{n-2}.$$

Therefore, the estimate of the variance of the natural logarithm of the stochastic central death rate, $\sigma^2\left(\ln(\tilde{m}_{x,\,t_U+\tau})\right)$, is given as

$$\hat{\sigma}^2\left(\ln(\tilde{m}_{x,\,t_U+\tau})\right) = \tau \times \hat{\beta}^2_x \times \hat{\sigma}^2_{\epsilon} + \hat{\sigma}^2_{\varepsilon_x},$$

and a $100(1-\gamma)\%$ predictive interval on $q_{x,\,t_U+\tau}$ is given by

$$1 - \exp\left\{-\exp\left[\ln(\hat{m}_{x,\,t_U+\tau}) \pm z_{\frac{\gamma}{2}} \times \hat{\sigma}^2\left(\ln(\tilde{m}_{x,\,t_U+\tau})\right)\right]\right\}.$$

Appendix B

One-dimensional Bühlmann Credibility Model

The one-dimensional Bühlmann credibility model for forecasting mortality rates for single population can also be referred to Tsai and Lin (2017b).

Suppose that we have $(n-1)$ past observed values, $Y_{x,\,t_L+1}, \ldots, Y_{x,\,t_U}$, where $t_U - t_L = n-1$, and would like to get the credibility estimator $\hat{Y}_{x,\,t_U+1}$ for ages $x = x_L, \ldots, x_U$ and the next year t_U+1. To implement the Bühlmann credibility approach, we further assume that $Y_{x,\,t}$, where $t = t_L, \ldots, t_U$, is characterized by the risk parameter, Θ_x, associated with age x.

Among the various possible predictors of $Y_{x,\,t_U+1}$, we choose to project $Y_{x,\,t_U+1}$ with a linear function of the past data $Y_{x,t_L+1}, \ldots, Y_{x,t_U}$ (see Bühlmann, 1967), i.e., $c_{x,\,0}+\sum_{t=1}^{n-1} c_{x,\,t} Y_{x,\,t_L+t}$. That is, we would like to choose $c_{x,\,0},\, c_{x,\,1}, \ldots, c_{x,\,n-1}$ to minimize the quadratic loss function Q, where

$$\mathrm{Q} = \mathrm{E}\left\{\left[Y_{x,\,t_U+1} - c_{x,\,0} - \sum_{t=1}^{n-1} c_{x,\,t} Y_{x,\,t_L+t}\right]^2\right\}. \tag{B.1}$$

To minimize Q, we take the derivative of Q with respect to $c_{x,\,t}$, for $t = 0, \ldots, n-1$, and set to 0,

- $\frac{\partial Q}{\partial c_{x,\,0}}$ implies that $\mathrm{E}[Y_{x,\,t_U+1}] = \hat{c}_{x,\,0} + \sum_{t=1}^{n-1} \hat{c}_{x,\,t} \mathrm{E}[Y_{x,\,t_L+t}]$;
- $\frac{\partial Q}{\partial c_{x,\,u}}$ implies that $\mathrm{E}[Y_{x,\,t_U+1} \cdot Y_{x,\,t_L+u}] = \hat{c}_{x,0}\mathrm{E}[Y_{x,\,t_L+u}] + \sum_{t=1}^{n-1} \hat{c}_{x,\,t} \mathrm{E}[Y_{x,\,t_L+t} \cdot Y_{x,\,t_L+u}]$, for $u = 1, \ldots, n-1$.

Special Case: (see Klugman et al., 2012) If $\mu_x = \mathrm{E}[Y_{x,\,t}] = \mu$, $\sigma_x^2(t,\,t) = \mathrm{Var}[Y_{x,\,t}] = \sigma^2$, and $\sigma_x^2(t_1,\,t_2) = \mathrm{Cov}[Y_{x,\,t_1},\,Y_{x,\,t_2}] = \rho \cdot \sigma^2$ for $t_1 \neq t_2$, where the correlation coefficient

$\rho \in [-1, 1]$, then we have

$$\hat{c}_{x,\,0} = \frac{(1-\rho)\cdot \mu}{1-\rho+(n-1)\rho},$$

and

$$\hat{c}_{x,\,t} = \frac{\rho \cdot \hat{c}_{x,\,0}}{(1-\rho)\cdot \mu} = \frac{\rho}{1-\rho+(n-1)\rho}.$$

Thus, the credibility estimate of $Y_{x,\,t_U+1}$ is given by

$$\hat{Y}_{x,\,t_U+1} = \frac{(1-\rho)\cdot \mu}{1-\rho+(n-1)\rho} + \frac{\rho}{1-\rho+(n-1)\rho} \cdot \sum_{t=1}^{n-1} Y_{x,\,t_L+t} = (1-Z)\,\mu + Z\overline{Y}_{x,\,\bullet},$$

where $Z = \frac{(n-1)\rho}{1-\rho+(n-1)\rho}$ and $\overline{Y}_{x,\,\bullet} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U} Y_{x,\,t}$. Note that if $\rho \in [0, 1]$ then $Z \in [0, 1]$, and the credibility estimate is a weighted average of the expectation μ and the sample mean $\bar{Y}_{x,\,\bullet}$.

B.1 Parametric Bühlmann Model

The following specifies the additional assumptions of the distributions of $Y_{x,t}|\Theta_x$ and Θ_x to construct the one-dimensional parametric Bühlmann credibility model,

Assumption B.1. *Conditional on the risk parameter Θ_x, $\{Y_{x,t}\}$ are independent and identically distributed for $t = t_L, \ldots, t_U$ with*

$$\begin{cases} E\left[Y_{x,t}|\Theta_x\right] & = \mu\left(\Theta_x\right), \\ Var\left[Y_{x,t}|\Theta_x\right] & = v\left(\Theta_x\right). \end{cases}$$

Assumption B.2. *Θ_x are independent and identically distributed as Θ for $x = x_L, \ldots, x_U$ with*

$$\begin{cases} \mu\left(\Theta_x\right) & = \mu\left(\Theta\right), \\ v\left(\Theta_x\right) & = v\left(\Theta\right). \end{cases}$$

Assumption B.3. *The distribution of Θ is such that*

$$\begin{cases} \mu & = E\left[\mu\left(\Theta\right)\right], \\ v & = E\left[v\left(\Theta\right)\right], \\ a & = Var\left[\mu\left(\Theta\right)\right]. \end{cases}$$

Denote the structural parameters as follows:

- the hypothetical mean, $\mu(\Theta_x) = \mathrm{E}[Y_{x,t}|\Theta_x]$;
- the process variance, $v(\Theta_x) = \mathrm{Var}[Y_{x,t}|\Theta_x]$;
- the expected value of the hypothetical mean, $\mu = \mathrm{E}\left[\mu\left(\Theta_x\right)\right]$;
- the expected value of the process variance, $v = \mathrm{E}\left[v\left(\Theta_x\right)\right]$; and
- the variance of the hypothetical mean, $a = \mathrm{Var}\left[\mu\left(\Theta_x\right)\right]$.

Under Assumptions B.1 to B.3, and the notations above, it follows directly that

- $\mathrm{E}[Y_{x,t}] = \mathrm{E}\left[\mathrm{E}(Y_{x,t}|\Theta_x)\right] = \mathrm{E}[\mu(\Theta_x)] = \mu$, where $t = t_L + 1, \ldots, t_U$;
- $\mathrm{Var}[Y_{x,t}] = \mathrm{E}[\mathrm{Var}(Y_{x,t}|\Theta_x)] + \mathrm{Var}[\mathrm{E}(Y_{x,t}|\Theta_x)] = \mathrm{E}[v(\Theta_x)] + \mathrm{Var}[\mu(\Theta_x)] = v + a$, where $t = t_L + 1, \ldots, t_U$; and
- $\mathrm{Cov}[Y_{x,t_1}, Y_{x,t_2}] = \mathrm{E}[\mathrm{Cov}(Y_{x,t_1}, Y_{x,t_2}|\Theta_x)] + \mathrm{Cov}[\mathrm{E}(Y_{x,t_1}|\Theta_x), \mathrm{E}(Y_{x,t_2}|\Theta_x)] = 0 + \mathrm{Var}[\mu(\Theta_x)] = a$, where $t_1,\ t_2 = t_L + 1, \ldots, t_U$ and $t_1 \neq t_2$.

Then, using the special case, we have $\rho \cdot \mathrm{Var}[Y_{x,t}] = \mathrm{Cov}[Y_{x,t_1}, Y_{x,t_2}]$ for $t_1 \neq t_2$, that is, $\rho \cdot (a + v) = a$. Thus, the Bühlmann credibility factor, Z, is given by

$$Z = \frac{(n-1)\rho}{1 - \rho + (n-1)\rho} = \frac{(n-1)a/(v+a)}{v/(v+a) + (n-1)a/(v+a)} = a\left(\frac{1}{n-1}v + a\right)^{-1},$$

and the Bühlmann credibility estimate of Y_{x,t_U+1} is

$$\hat{Y}_{x,t_U+1} = Z\overline{Y}_x + (1 - Z)\,\mu,$$

where $\overline{Y}_x = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U} Y_{x,t}$ is the sample mean for age x,.

B.2 Non-parametric Bühlmann Model

To get the non-parametric estimators of μ, v and a, we further assume that

Assumption B.4. *The pairs* $\{(\Theta_x, Y_x), x = x_L, \ldots, x_U\}$, *where* $\mathbf{Y}_x = (Y_{x,t_L+1}, \ldots, Y_{x,t_U})$, *are independent.*

Table B.1: Non-parametric Bühlmann estimation for μ, v and a

$\boldsymbol{Y}_{x_L} = (Y_{x_L,t_L+1}, \ldots, Y_{x_L,t_U})$,	$\hat{\mu}_{x_L} = \overline{Y}_{x_L} = \frac{1}{n-1} \sum_{t=t_L+1}^{t_U} Y_{x_L,t}$, $\hat{v}_{x_L} = \frac{1}{n-2} \sum_{t=t_L+1}^{t_U} \left(Y_{x_L,t} - \overline{Y}_{x_L}\right)\left(Y_{x_L,t} - \overline{Y}_{x_L}\right)$;
$\vdots$	$\vdots$
$\boldsymbol{Y}_{x_U} = (Y_{x_U,t_L+1}, \ldots, Y_{x_U,t_U})$.	$\hat{\mu}_{x_U} = \overline{Y}_{x_U} = \frac{1}{n-1} \sum_{t=t_L+1}^{t_U} Y_{x_U,t}$, $\hat{v}_{x_U} = \frac{1}{n-2} \sum_{t=t_L+1}^{t_U} \left(Y_{x_U,t} - \overline{Y}_{x_U}\right)\left(Y_{x_U,t} - \overline{Y}_{x_U}\right)$.
$\hat{a} = \frac{1}{m-1} \sum_{x=x_L}^{x_U} (\overline{Y}_x - \overline{Y})(\overline{Y}_x - \overline{Y}) - \frac{\hat{v}}{n-1}$.	$\hat{\mu} = \overline{Y} = \frac{1}{m} \sum_{x=x_L}^{x_U} \overline{Y}_x$, $\hat{v} = \frac{1}{m} \sum_{x=x_L}^{x_U} \hat{v}_x, \quad m = x_U - x_L + 1.$

The preliminary unbiased estimators of μ, v and a are presented in Table B.1

Note that the preliminary non-parametric estimate of the variance of the hypothetical mean, $\hat{a}$, could be negative. Thus, the ultimate estimate of $\hat{a}$ is set by

$$\hat{a} = \begin{cases} \hat{a}, & \text{if } \hat{a} \geqslant 0, \\ 0, & \text{if } \hat{a} < 0. \end{cases}$$

Therefore, the Bühlmann non-parametric estimate of Y_{x,t_U+1}, for age x in year $t_U + 1$, obtained by minimizing the quadratic loss function in (B.1), is given by

$$\hat{Y}_{x,\,t_U+1} = \hat{Z}\overline{Y}_{x,\bullet} + \left(1 - \hat{Z}\right)\hat{\mu}, \quad x = x_L, \ldots, x_U, \tag{B.2}$$

where $\hat{Z} = \hat{a}\left(\frac{1}{n-1}\hat{v} + \hat{a}\right)^{-1}$ and $\overline{Y}_{x,\bullet} = \frac{1}{n-1}\sum_{t=t_L+1}^{t_U} Y_{x,t}$.

Finally, the Bühlmann non-parametric estimate of $\ln(m_{x,\,t_U+1})$ for age x in year $t_U + 1$ is $\ln(\hat{m}_{x,\,t_U+1}) = \ln(m_{x,\,t_U}) + \hat{Y}_{x,\,t_U+1}$ provided that $Y_{x,\,t+1} = \ln(m_{x,\,t+1}) - \ln(m_{x,\,t}),\ t = t_L, t_L + 1, \ldots$.

Appendix C

Kronecker Product

C.1 Definition and Example

Definition 1. *Let $\boldsymbol{F}$ be an $m \times n$ matrix with elements $f_{i,j}$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$, and $\boldsymbol{G}$ be a $p \times q$ matrix; then the Kronecker product of $\boldsymbol{F}$ and $\boldsymbol{G}$, denoted by $\boldsymbol{F} \otimes \boldsymbol{G}$, is defined as the $mp \times nq$ block matrix:*

$$\boldsymbol{F} \otimes \boldsymbol{G} = \begin{bmatrix} f_{1,1}\boldsymbol{G} & \cdots & f_{1,n}\boldsymbol{G} \\ \vdots & \ddots & \vdots \\ f_{m,1}\boldsymbol{G} & \cdots & f_{m,n}\boldsymbol{G} \end{bmatrix}.$$

Example: For $m = n = p = q = 2$, we have

$$\begin{aligned} \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \end{bmatrix} \otimes \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{bmatrix} &= \begin{bmatrix} f_{1,1} \cdot \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{bmatrix} & f_{1,2} \cdot \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{bmatrix} \\ f_{2,1} \cdot \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{bmatrix} & f_{2,2} \cdot \begin{bmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} f_{1,1} \cdot g_{1,1} & f_{1,1} \cdot g_{1,2} & f_{1,2} \cdot g_{1,1} & f_{1,2} \cdot g_{1,2} \\ f_{1,1} \cdot g_{2,1} & f_{1,1} \cdot g_{2,2} & f_{1,2} \cdot g_{2,1} & f_{1,2} \cdot g_{2,2} \\ f_{2,1} \cdot g_{1,1} & f_{2,1} \cdot g_{1,2} & f_{2,2} \cdot g_{1,1} & f_{2,2} \cdot g_{1,2} \\ f_{2,1} \cdot g_{2,1} & f_{2,1} \cdot g_{2,2} & f_{2,2} \cdot g_{2,1} & f_{2,2} \cdot g_{2,2} \end{bmatrix}. \end{aligned}$$

C.2 Properties of the Kronecker Product

Property 1. *Bilinearity and associativity: The Kronecker product is bilinear and associative:*

- $\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$,
- $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$,
- $(k\boldsymbol{A}) \otimes \boldsymbol{B} = \boldsymbol{A} \otimes (k\boldsymbol{B}) = k(\boldsymbol{A} \otimes \boldsymbol{B})$,
- $(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$,

where $\boldsymbol{A}$, $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *are matrices and* k *is a scalar.*

Property 2. *Non-commutative: In general,* $\boldsymbol{A} \otimes \boldsymbol{B}$ *and* $\boldsymbol{B} \otimes \boldsymbol{A}$ *are different matrices.*

Property 3. *The mixed-product property and the inverse of a Kronecker product: If* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ *and* $\boldsymbol{D}$ *are matrices of such size that one can form the matrix products* $\boldsymbol{AC}$ *and* $\boldsymbol{BD}$, *then*

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}.$$

This is called the mixed-product property, because it mixes the ordinary matrix product and the Kronecker product. It follows that $\boldsymbol{A} \otimes \boldsymbol{B}$ *is invertible if and only if both* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are invertible, in which case the inverse is given by*

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}.$$

Appendix D

Proofs of Lemma 1 and Theorem 1

The proofs of Lemma 1 and Theorem 1 can also be referred to Poon and Lu (2015).

D.1 Proof of Lemma 1

Lemma 1. *Under Assumptions 1 to 3, and the notations in Chapter* 3*,*

1. *The means of $\boldsymbol{Y}_x$ and $\boldsymbol{Y}_{x,t_U+1}$ are given by*

$$\boldsymbol{\mu}_{\boldsymbol{Y}_x} = E(\boldsymbol{Y}_x) = \boldsymbol{\mu} \otimes \mathbf{1}_{n-1} \text{ and } \boldsymbol{\mu}_{\boldsymbol{Y}_{x,t_U+1}} = E(\boldsymbol{Y}_{x,t_U+1}) = \boldsymbol{\mu},$$

where $\otimes$ is the Kronecker product operator and $\mathbf{1}_{n-1} = (1, \ldots, 1)'$, a column vector of length $(n-1)$.

2. *The $r(n-1) \times r(n-1)$ covariance matrix of $\boldsymbol{Y}_x$ is given by*

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x} = \boldsymbol{V} \otimes \boldsymbol{I}_{n-1} + \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}', \tag{D.1}$$

where

$$\boldsymbol{I}_{n-1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}_{(n-1)\times(n-1)},$$

and

$$\boldsymbol{U} = \begin{bmatrix} \mathbf{1}_{n-1} & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{1}_{n-1} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \mathbf{1}_{n-1} \end{bmatrix}_{r(n-1)\times r}.$$

3. *The $r \times r(n-1)$ covariance matrix between $\boldsymbol{Y}_x$ and $\boldsymbol{Y}_{x,t_U+1}$ is given by*

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,\,t_U+1},\boldsymbol{Y}_x} = \boldsymbol{A}\boldsymbol{U}'. \tag{D.2}$$

4. *The inverse of the $r(n-1) \times r(n-1)$ covariance matrix of $\boldsymbol{Y}_x$ is given by*

$$(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1} = (\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_{n-1}) - (\boldsymbol{V}^{-1} \otimes \mathbf{1}_{n-1})\boldsymbol{U}\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}\boldsymbol{U}'(\boldsymbol{V}^{-1} \otimes \mathbf{1}'_{n-1}).$$

Proof. **1.** From Assumptions 1 to 3, we get

$$\mathrm{E}(\boldsymbol{Y}_{x,\,t}) = \mathrm{E}\left[\mathrm{E}(\boldsymbol{Y}_{x,\,t}|\boldsymbol{\Theta}_x)\right] = \mathrm{E}[\boldsymbol{\mu}(\boldsymbol{\Theta}_x)] = \boldsymbol{\mu}, \tag{D.3}$$

implying that

$$\boldsymbol{\mu}_{\boldsymbol{Y}_x} = \mathrm{E}(\boldsymbol{Y}_x) = \boldsymbol{\mu} \otimes \mathbf{1}_{n-1} \text{ and } \boldsymbol{\mu}_{\boldsymbol{Y}_{x,t_U+1}} = \mathrm{E}(\boldsymbol{Y}_{x,t_U+1}) = \boldsymbol{\mu}.$$

2. From Assumptions 1 to 3, it follows directly that

$$\mathrm{E}\left[\mathrm{Cov}(\boldsymbol{Y}_{x,\,t_1},\, \boldsymbol{Y}'_{x,\,t_2}|\boldsymbol{\Theta}_x)\right] = \begin{cases} \mathrm{E}\left[\boldsymbol{\Sigma}\left(\boldsymbol{\Theta}_x\right)\right] = \boldsymbol{V}, & t_1 = t_2 \\ \mathbf{0}, & t_1 \neq t_2 \end{cases}, \tag{D.4}$$

and

$$\mathrm{Cov}\left[\mathrm{E}(\boldsymbol{Y}_{x,\,t_1}|\boldsymbol{\Theta}_x),\, \mathrm{E}(\boldsymbol{Y}'_{x,\,t_2}|\boldsymbol{\Theta}_x)\right] = \mathrm{Cov}\left[\boldsymbol{\mu}(\boldsymbol{\Theta}_x), \boldsymbol{\mu}(\boldsymbol{\Theta}_x)'\right] = \boldsymbol{A}. \tag{D.5}$$

From (D.4) and (D.5) and the well-known decomposition of the covariance matrix, we have

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x} &= \mathrm{Cov}\left[\boldsymbol{Y}_x, \boldsymbol{Y}'_x\right] \\ &= \mathrm{E}\left[\mathrm{Cov}(\boldsymbol{Y}_x,\, \boldsymbol{Y}'_x|\boldsymbol{\Theta}_x)\right] + \mathrm{Cov}\left[\mathrm{E}(\boldsymbol{Y}_x|\boldsymbol{\Theta}_x),\, \mathrm{E}(\boldsymbol{Y}'_x|\boldsymbol{\Theta}_x)\right] \\ &= \boldsymbol{V} \otimes \boldsymbol{I}_{n-1} + \boldsymbol{A} \otimes (\mathbf{1}_{n-1}\mathbf{1}'_{n-1}) \\ &= \boldsymbol{V} \otimes \boldsymbol{I}_{n-1} + \boldsymbol{U}\boldsymbol{A}\boldsymbol{U}'. \end{aligned}$$

3. The proof of (D.2) follows analogously to the proof of (D.1), that is,

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,\,t_U+1},\boldsymbol{Y}_x} &= \mathrm{Cov}\left(\boldsymbol{Y}_{x,\,t_U+1},\boldsymbol{Y}_x\right)\\ &= \mathrm{E}\left[\mathrm{Cov}(\boldsymbol{Y}_{x,\,t_U+1},\ \boldsymbol{Y}_x|\boldsymbol{\Theta}_x)\right]+\mathrm{Cov}\left[\mathrm{E}(\boldsymbol{Y}_{x,\,t_U+1}|\boldsymbol{\Theta}_x),\ \mathrm{E}(\boldsymbol{Y}_x|\boldsymbol{\Theta}_x)\right]\\ &= \boldsymbol{0}+\boldsymbol{A}\boldsymbol{U}'\\ &= \boldsymbol{A}\boldsymbol{U}'.\end{aligned}$$

4. Applying the matrix inversion identity $(\boldsymbol{E}+\boldsymbol{F}\boldsymbol{G}\boldsymbol{H})^{-1}=\boldsymbol{E}^{-1}-\boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{G}^{-1}+\boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F})^{-1}\boldsymbol{H}\boldsymbol{E}^{-1}$ to (D.1), where we set $\boldsymbol{E}=\boldsymbol{V}\otimes\boldsymbol{I}_{n-1}$, $\boldsymbol{F}=\boldsymbol{U}$, $\boldsymbol{G}=\boldsymbol{A}$ and $\boldsymbol{H}=\boldsymbol{U}'$, we get

$$\begin{aligned}(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1} &= (\boldsymbol{V}\otimes\boldsymbol{I}_{n-1}+\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}')^{-1}\\ &= (\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}-(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}\boldsymbol{U}\left[\boldsymbol{A}^{-1}+\boldsymbol{U}'(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}\boldsymbol{U}\right]^{-1}\boldsymbol{U}'(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}\\ &= (\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1})-(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_{n-1})\left[\boldsymbol{A}^{-1}+(n-1)\boldsymbol{V}^{-1}\right]^{-1}(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}'_{n-1}).\end{aligned}$$

In the last equality above, we have used

- $(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}=\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1}^{-1}=\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1}$,
- $(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}\boldsymbol{U}=(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1})\,\boldsymbol{U}=\left[v_{i,\,j}^{-1}\boldsymbol{I}_{n-1}\right]_{i,\,j}\boldsymbol{U}=\boldsymbol{V}^{-1}\otimes\boldsymbol{1}_{n-1}$,
- $\boldsymbol{U}'\,(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}=\boldsymbol{U}'\,(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1})=\boldsymbol{U}'\left[v_{i,\,j}^{-1}\boldsymbol{I}_{n-1}\right]_{i,\,j}=\boldsymbol{V}^{-1}\otimes\boldsymbol{1}'_{n-1}$, and
- $\boldsymbol{U}'\,(\boldsymbol{V}\otimes\boldsymbol{I}_{n-1})^{-1}\boldsymbol{U}=\boldsymbol{U}'\,(\boldsymbol{V}^{-1}\otimes\boldsymbol{I}_{n-1})\,\boldsymbol{U}=(\boldsymbol{V}^{-1}\otimes\boldsymbol{1}'_{n-1})\,\boldsymbol{U}=\left[v_{i,\,j}^{-1}\boldsymbol{1}'_{n-1}\right]_{i,\,j}\boldsymbol{U}=(n-1)\boldsymbol{V}^{-1}$,

where $v_{i,\,j}^{-1}$ is entry of matrix $\boldsymbol{V}^{-1}$. □

D.2 Proof of Theorem 1

Theorem 1. *Under Assumption 1 to 3, the parametric Bühlmann estimate of* $\boldsymbol{Y}_{x,t_U+1}$, $\hat{\boldsymbol{Y}}_{x,t_U+1}=\left(\hat{Y}_{x,t_U+1,1},\ldots,\hat{Y}_{x,t_U+1,r}\right)'$, *for age x in year t_U+1, which is obtained by minimizing the quadratic loss function in* (3.3) *is given by*

$$\hat{\boldsymbol{Y}}_{x,\,t_U+1}=\boldsymbol{Z}\overline{\boldsymbol{Y}}_{x,\bullet}+(\boldsymbol{I}_r-\boldsymbol{Z})\,\boldsymbol{\mu},$$

where

- $\boldsymbol{Z}=\boldsymbol{A}\left(\frac{1}{n-1}\boldsymbol{V}+\boldsymbol{A}\right)^{-1}$, *and*

- $\overline{\boldsymbol{Y}}_{x,\bullet} = \left(\overline{Y}_{x,\bullet,1}, \ldots, \overline{Y}_{x,\bullet,r}\right)' = \frac{1}{n-1}\left(\sum_{t=t_L+1}^{t_U} Y_{x,t,1}, \ldots, \sum_{t=t_L+1}^{t_U} Y_{x,t,r}\right)'$.

Proof. From Lemma 1 and

- $\boldsymbol{U}'\left(\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_{n-1}\right) = \boldsymbol{U}'\left[v_{i,j}^{-1}\boldsymbol{I}_{n-1}\right]_{i,j} = \boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1}$,
- $\boldsymbol{U}'\left(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}_{n-1}\right) = \boldsymbol{U}'\left[v_{i,j}^{-1}\boldsymbol{1}_{n-1}\right]_{i,j} = (n-1)\boldsymbol{V}^{-1}$, and
- $(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1})(\boldsymbol{Y}_x - \boldsymbol{\mu}_{\boldsymbol{Y}_x}) = \left[v_{i,j}^{-1}\boldsymbol{1}'_{n-1}\right]_{i,j}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{\boldsymbol{Y}_x}) = (n-1)\boldsymbol{V}^{-1}(\overline{\boldsymbol{Y}}_{x,\bullet} - \boldsymbol{\mu})$,

it follows directly that

$$\begin{aligned}
& (\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\boldsymbol{Y}_x})(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) \\
= \; & \boldsymbol{A}\boldsymbol{U}'\left\{(\boldsymbol{V}^{-1} \otimes \boldsymbol{I}_{n-1}) - (\boldsymbol{V}^{-1} \otimes \boldsymbol{1}_{n-1})\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1})\right\}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) \\
= \; & \left\{\boldsymbol{A}(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1}) - \boldsymbol{A}\boldsymbol{U}'(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}_{n-1})\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1})\right\}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) \\
= \; & \left\{\boldsymbol{A} - \boldsymbol{A}(n-1)\boldsymbol{V}^{-1}\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}\right\}(\boldsymbol{V}^{-1} \otimes \boldsymbol{1}'_{n-1})(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) \\
= \; & \left\{\boldsymbol{A} - \boldsymbol{A}(n-1)\boldsymbol{V}^{-1}\left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}\right\}(n-1)\boldsymbol{V}^{-1}(\overline{\boldsymbol{Y}}_{x,\bullet} - \boldsymbol{\mu}).
\end{aligned}$$

Applying the inverse matrix identity $(\boldsymbol{E}+\boldsymbol{F})^{-1} = \boldsymbol{E}^{-1} - \boldsymbol{E}^{-1}\boldsymbol{F}(\boldsymbol{E}+\boldsymbol{F})^{-1}$, where we set $\boldsymbol{E} = \boldsymbol{A}^{-1}$ and $\boldsymbol{F} = (n-1)\boldsymbol{V}^{-1}$, leads to

$$(\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\boldsymbol{Y}_x})(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) = \left[\boldsymbol{A}^{-1} + (n-1)\boldsymbol{V}^{-1}\right]^{-1}(n-1)\boldsymbol{V}^{-1}(\overline{Y}_{x,\bullet} - \boldsymbol{\mu}),$$

using $(\boldsymbol{E}+\boldsymbol{F})^{-1}F = \boldsymbol{E}^{-1}(\boldsymbol{F}^{-1}+\boldsymbol{E}^{-1})^{-1}$, we have

$$(\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\boldsymbol{Y}_x})(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) = \boldsymbol{A}\left(\frac{1}{n-1}\boldsymbol{V} + \boldsymbol{A}\right)^{-1}(\overline{\boldsymbol{Y}}_{x,\bullet} - \boldsymbol{\mu}) = \boldsymbol{Z}(\overline{Y}_{x,\bullet} - \boldsymbol{\mu}).$$

Plugging the equation above into (3.4), we thus get the credibility estimator of $\boldsymbol{Y}_{x,t_U+1}$,

$$\begin{aligned}
\hat{\boldsymbol{Y}}_{x,t_U+1} &= \boldsymbol{\mu} + (\boldsymbol{\Sigma}_{\boldsymbol{Y}_{x,t_U+1},\boldsymbol{Y}_x})(\boldsymbol{\Sigma}_{\boldsymbol{Y}_x,\boldsymbol{Y}_x})^{-1}(\boldsymbol{Y}_x - \boldsymbol{\mu}_{Y_x}) \\
&= \boldsymbol{\mu} + \boldsymbol{Z}(\overline{\boldsymbol{Y}}_{x,\bullet} - \boldsymbol{\mu}) \\
&= \boldsymbol{Z}\overline{\boldsymbol{Y}}_{x,\bullet} + (\boldsymbol{I}_r - \boldsymbol{Z})\boldsymbol{\mu},
\end{aligned}$$

which is a weighted average of $\overline{\boldsymbol{Y}}_{x,\bullet}$ and $\boldsymbol{\mu}$. □

Bibliography

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

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

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

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

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

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

D’Amato, Valeria, Haberman, Steven, Piscopo, Gabriella, and Russolillo, Maria (2012). “Modelling dependent data for longevity projections”. In: *Insurance: Mathematics and Economics* 51(3), pp. 694–701.

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

Dowd, Kevin, Cairns, Andrew J.G., Blake, David, Coughlan, Guy D., and Khalaf-Allah, Marwa (2011). “A gravity model of mortality rates for two related populations”. In: *North American Actuarial Journal* 15(2), pp. 334–356.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Wilson, Chris (2001). “On the scale of global demographic convergence 1950–2000”. In: *Population and Development Review* 27.1, pp. 155–171.

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