Enhanced Hierarchical Bayesian Modeling of Mortality Rates Incorporating Jumps: A Multi-Country Perspective

by

Yi-Rong Zhu

M.Sc., National Chengchi University, 2016
B.Sc., Soochow University, 2014

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

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Yi-Rong Zhu**
Degree:	**Master of Science**
Thesis title:	**Enhanced Hierarchical Bayesian Modeling of Mortality Rates Incorporating Jumps: A Multi-Country Perspective**
Committee:	**Chair:** Liangliang Wang Associate Professor, Statistics and Actuarial Science
	Cary Chi-Liang Tsai Supervisor Professor, Statistics and Actuarial Science
	Yi Lu Committee Member Professor, Statistics and Actuarial Science
	Himchan Jeong Examiner Assistant Professor, Statistics and Actuarial Science

Abstract

The rapid aging of populations and global connectivity necessitates considering factors such as geographical proximity, socioeconomic conditions, and cultural connections in mortality forecasting. Additionally, catastrophes, pandemics, and political turmoil can disrupt the delicate balance of this complex system, introducing further layers of uncertainties to the already challenging task of predicting the future course of life expectancy. This project builds upon a hierarchical Bayesian random walk with drift (HBS-RW) model, inheriting its advantage of capturing the dependency structure of mortality rates across ages and populations. We further extend the model by adding jump components to account for extreme events in the data. From the numerical illustrations, the proposed model with jump components generally demonstrates superior performance compared to the HBS-RW model without a jump component and other Lee-Carter-based models in most conducted tests. Among the jump models, those incorporating trajectory jump effects yield the lowest average of mean absolute percentage error (AMAPE) overall. The choice between independent and dependent trajectory jump effects depends on the characteristics of the data. However, from a practical perspective, opting for the independent trajectory jump effect is favored, as it is significantly time efficient, compared to the models incorporating the dependent trajectory jump effect, in the maximum likelihood estimation for parameters.

Keywords: Hierarchical Bayesian; mortality modeling; multi-population mortality; transitory jump effects; permanent jump effects

Dedication

To my family, whose constant support and unconditional understanding mean a lot to me and fuel my journey forward.

Acknowledgements

First of all, I would like to express my sincere gratitude to Dr. Cary Tsai for his outstanding guidance over the past few years. His invaluable advice and patient assistance in revising every detail of the research are crucial to my completion of the project. I am also deeply grateful to Dr. Lu for the fundamental knowledge gained from her Advanced Actuarial Models course, which provided important nourishment for this research. Furthermore, her careful review and valuable feedback greatly contributed to the refinement of this project. I would also like to thank Dr. Jeong and Dr. Wang for their insightful comments during the project defense, which provided valuable insights into addressing similar challenges in my future work. Additionally, I express my heartfelt appreciation to Dr. Lin at National Chengchi University of Taiwan for his strong recommendation to Dr. Cary Tsai, which gave me the opportunity to study at Simon Fraser University and land a job in Canada. Special thanks to the managers at Aon for offering me with flexible and remote work, which allowed me to balance my studies and work. Last but not least, I would like to thank my family again for their unconditional support and sacrifice. They are always my strongest support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Developing a robust predictive model for future mortality rates has been an ongoing topic of interest in life insurance industry. The Lee-Carter (LC) model, proposed by Lee and Carter (1992), is frequently referred due to its straightforward and easily extensible nature. The the Joint-k (JK) model, the augmented common factor (ACF) model, and the co-integrated (CI) model, presented by Carter and Lee (1992), Li and Lee (2005), and Li and Hardy (2011), respectively, are three extensions of the Lee-Carter model. These three advancements enable more applications in fitting and forecasting mortality rates across multiple populations. However, two-stage fitting process of the Lee-Carter model, involving separate model parameter estimation and trend fitting using time series models, can lead to potential inconsistencies, resulting from multiple sources, including parameter estimation, mortality projection, and potential model misspecification. This, in turn, affects the accuracy of long-term mortality predictions.

Aiming to mitigate potential inconsistencies arising from the two-stage model fitting process within the Lee-Carter model and its extensions, Lin and Tsai (2022) proposed a solution that simplifies the model specification to a random walk with drift model, and then applies Bayesian theory to model the logarithmic dynamics of central death rates across a range of ages within a single population. Additionally, Lin and Tsai (2022) considered cross-population co-movements prompted by the impact of rapid globalization by integrating the hierarchical Bayesian theory into the modeling of mortality rates across multiple populations. Further details on the Bayesian mortality model and its extensions are discussed in Chapter 2.

While the Jump model is widely employed in financial mathematics to capture high-severity but low-frequency events, relatively few studies have applied it to jump events in mortality data. Most studies discussing mortality models still overlook the impact of adverse

mortality risk. Cox et al. (2006) is the pioneer in recognizing the impact of extreme events on mortality rates, by directly applying the geometric Brownian motion and a compound Poisson process, widely used in financial models, to mortality models. Following Cox et al. (2006), Lin and Cox (2008) extended their work by proposing a model that considers short-term, transitory effects of adverse mortality jumps. However, it is worth noting that their model overlooks potential correlations in mortality data. Recognizing the non-negligible correlation within mortality data, Chen and Cox (2009) developed a new Lee-Carter-based model capturing transitory effects.

Building upon the hierarchical Bayesian mortality model studied in Lin and Tsai (2022), this project employs the model's inherent capability to incorporate co-movements across populations resulting from rapid internationalization. Additionally, the new model introduces either permanent or transitory jump effects in its mortality stochastic processes to capture adverse mortality risks. Ultimately, this project demonstrates the effectiveness of the new model in mortality forecasting, highlighting its superior performance compared to conventional mortality models.

1.2 Motivation

Mortality rate plays a crucial role in determining the prices of insurance, annuity, and retirement products, as well as influencing the overall business and operational strategies of life insurance companies. Accurate mortality forecasting has traditionally been pivotal in directly impacting the prices of life insurance products. Nevertheless, the proliferation of financial innovations in recent decades has granted insurance and reinsurance companies greater flexibility in transferring or hedging their mortality risk through securitization. This makes reliable mortality predictions invaluable tools for informed decision-making, ensuring the resilience and adaptability of these companies in the face of evolving challenges.

The rapid recent aging of populations and increased global connectivity has necessitated researchers and actuaries to consider diverse aspects when predicting mortality rates. This expansion goes beyond traditional considerations like age and gender, now encompassing influences like geographical proximity, the intricacies of socioeconomic conditions, and deep-rooted cultural connections. These broader considerations are crucial, as the world becomes increasingly interconnected and populations are aging at unprecedented rates. Furthermore, catastrophes, pandemics, and political turmoil can suddenly disrupt the delicate balance of this complex system, and these occurrences can trigger dramatic spikes in mortality rates, introducing further layers of complexities and uncertainties to the already challenging task of predicting the future course of life expectancy. However, existing mortality models of-

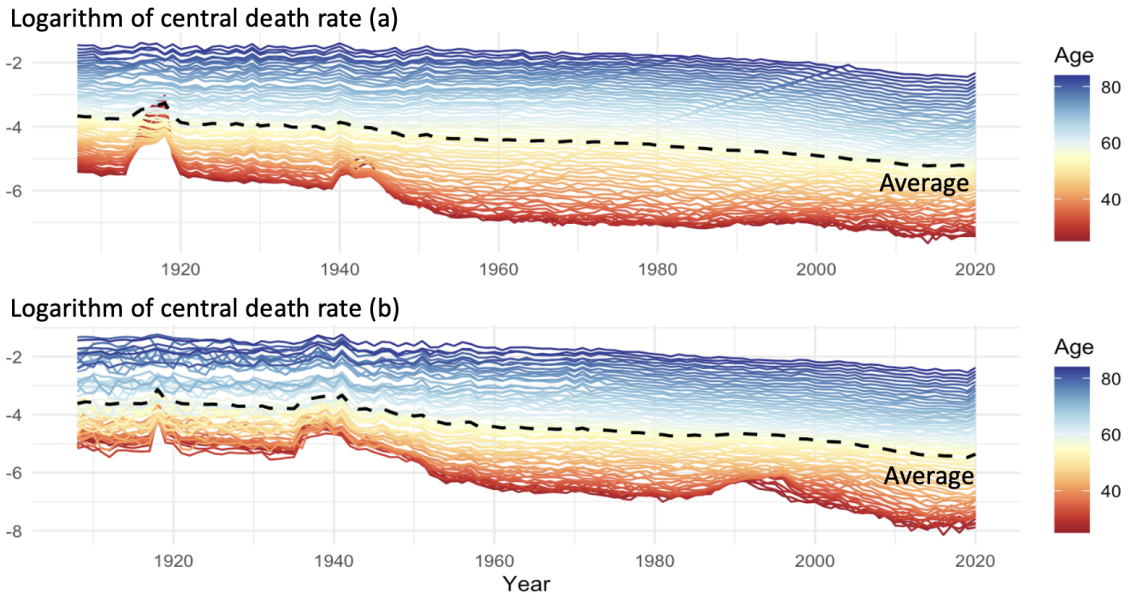

Figure 1.1: The Logarithm of central death rate. (a) UK males (b) Spain males

ten focus on single demographic groups, neglecting the potential impact of extreme events such as pandemics, catastrophes, and war. Additionally, most models, extended from the Lee-Carter model, face inconsistencies due to two-stage parameter estimation and trend fitting. To address these challenges, Lin and Tsai (2022) proposed a hierarchical Bayesian random walk model. This innovative approach not only accounts for multiple populations and mitigates estimation inconsistencies but also demonstrates superior predictive performance compared to commonly used models.

Regarding extreme events in mortality data, although some literature on mortality modeling has discussed the effectiveness of jump models in capturing high-severity but low-frequency events, the application of these models in mortality forecasting is rarely explored. The mortality datasets for the United Kingdom and Spain are comprehensive, allowing the observation of pivotal events in history that led to sharp increases in mortality rates, for example, World War, the Spanish flu pandemic, and the COVID-19 pandemic. As depicted in Figure 1.1, World War I had a significant impact on mortality rates in both the United Kingdom and Spain in the early 1900s, especially among young males who were recruited to army. Additionally, pandemics contributed to a portion of the observed mortality changes. The substantial shifts in mortality rates during the 1940s were primarily influenced by World War II. Finally, the fluctuations observed in 2019-2020 resulted from the COVID-19 pandemic. Figure 1.2, illustrating yearly mortality decrements, provides a more direct reference to these events, with clear spikes seen during the aforementioned periods in both countries.

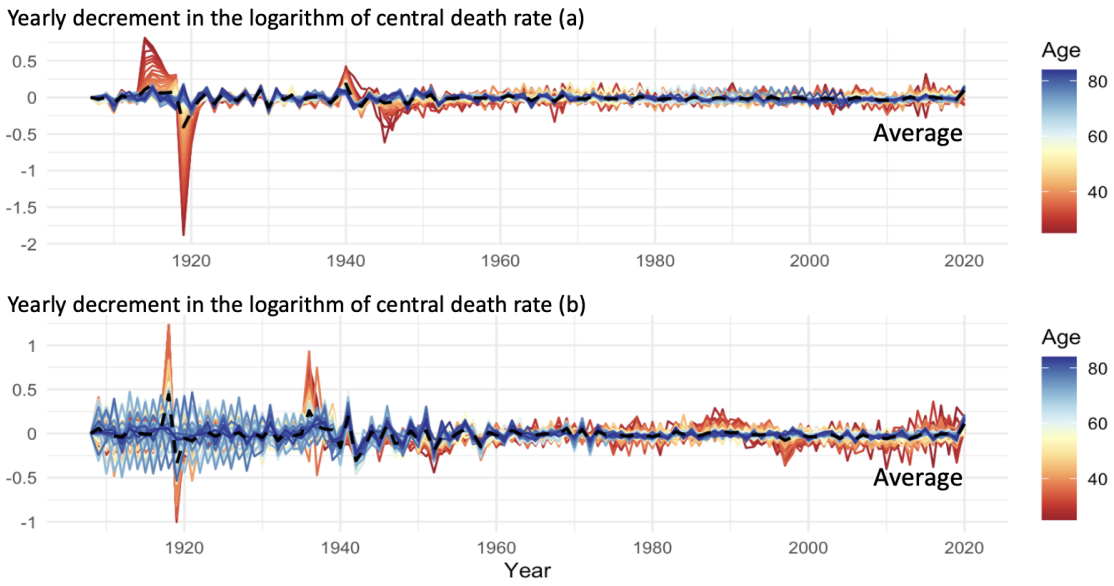

Figure 1.2: The yearly decrement in the logarithm of central death rate. (a) UK males (b) Spain males

Our approach bridges the gap by integrating the strengths of the hierarchical Bayesian model with jump models. This innovative combination aims to comprehensively capture the intricate impact of speedy internationalization and extreme events within mortality data, potentially leading to a significant advance in accurate mortality forecasting.

1.3 Outline

The remainder of this project is organized as follows. Chapter 2 presents a comprehensive literature review, covering Lee-Carter-based mortality forecasting models, Bayesian mortality and demographic modeling, as well as mortality models that incorporate jump models. Chapter 3 offers a detailed mathematical derivation of the extended Hierarchical Bayesian model, incorporating various types of jumps, including permanent and transitory jump effects. In Chapter 4, numerical illustrations demonstrate the application of our proposed model. Within the same chapter, we assess the forecasting performance of our solution, comparing it with other commonly used models for multi-population forecasting. The evaluation is conducted using the Average of Mean Absolute Percentage Error (AMAPE). Finally, Chapter 5 provides a final conclusion of our findings and insights.

Chapter 2

Literature Review

2.1 Mortality forecasting

Literature on mortality forecasting models has significantly expanded since Lee and Carter (1992) introduced their classic mortality model. This model is founded based on the assumption that the dynamics of the logarithm of central death rates for a single population is influenced by three factors: an age-specific parameter, the rate of change at each age multiplied by an overall time trend of mortality rates, plus a model error. However, the inherent limitation of the Lee-Carter model, specific to single populations, prompts scholars to extend it in order to capture and predict the dynamics of mortality rates across diverse demographic populations. Carter and Lee (1992) proposed the joint-k (JK) model, replacing the single population time trend with a common one across diverse populations, allowing for a better understanding of mortality dynamics in various situations. Li and Lee (2005) argued that closely related populations share similar mortality patterns unlikely to diverge significantly over time, and introduced the augmented common factor (ACF) model by considering both common historical experiences and individual trend differences. This approach enhances the understanding of mortality trends by capturing shared patterns and individual variations. Li and Hardy (2011) challenged the assumption of no connection between mortality rates in the Lee-Carter model for two populations of interest, proposing the co-integrated (CI) model as an improvement. Instead of employing two independent processes, they assume a linear relationship between the time trend for the base population and that for each of other studied populations, creating a model that captures correlated changes over time. This approach offers a novel perspective on the intricate relationships within diverse populations. Renshaw and Haberman (2006) enhanced the Lee-Carter model by introducing a cohort effect to capture characteristic and systematic mortality patterns. The analysis suggests that limiting the modeling to Lee-Carter alone might lead to the loss of distinctive cohort-related patterns. However, the established linear patterns in the period and cohort parameters provide a solid foundation for effective forecasting, contributing to an overall improvement in model fitting. Other studies based on the Lee-Carter model

can be referred to Renshaw and Haberman (2003), Li et al. (2009), and Mitchell et al. (2013).

While the Lee-Carter model exhibits effective and straightforward mortality forecasting capabilities, its drawbacks are also apparent. The two-stage fitting process of the Lee-Carter model, which involves separate parameter estimation and trend fitting using time series models, can lead to potential inconsistencies, resulting from multiple sources, including parameter estimation, mortality projection, and potential model misspecification. This approach may lead to discrepancies between projected trends and actual mortality rates, affecting the accuracy of long-term mortality predictions (see, for example, Leng and Peng (2016) and Lin and Tsai (2022)).

In existing literature, an alternative suggestion is to simplify the model specification by using a random walk with drift model for the dynamics of the logarithm of central death rates. Tsai and Yang (2015) introduced a linear relational model that exploits the relationship between a specific population's mortality sequence and the base one to predict future mortality rates. The sequences for each of the intercept and slope parameters from the regression model can be further modeled using either a simple linear regression or a random walk with drift approach. Compared to the Lee-Carter model, their models demonstrate overall superior forecasting performance across the empirical age span, as evidenced by lower average of root mean squared errors (RMSEs), lower average of mean absolute errors (MAEs), and lower average of mean absolute percentage errors (MAPEs). Tsai and Lin (2017) integrated the Bühlmann semi-parametric credibility approach into each of the Lee-Carter model, the CBD model, and the model proposed by Tsai and Yang (2015) to improve mortality forecasting. This approach suggests that, the predicted yearly mortality decrement for a specific age is determined by a weighted average of two yearly decrements for that age time trend and the overall group time trend for all ages. The credibility slope is suggested to more effectively capture the downward trend of the future mortality rates, leading to improve forecasting performance. Lin and Tsai (2022) proposed a random walk with drift model and further applied Bayesian theory to enhance the forecasting performance. The results show that this approach not only implies easier parameter estimation but also avoids inconsistencies found in the Lee-Carter model.

2.2 Bayesian mortality and demographic modeling

The literature related to Bayesian mortality and demographic modeling has been extensively developed (see, for example, Czado et al. (2005), Pedroza (2006), and Wong et al. (2018)). Arató et al. (2006) explored hierarchical Bayesian modeling to estimate the number of age-dependent deaths across different geographic regions. The model integrates a conditional binomial distribution for the number of age-dependent deaths and employs a

Gaussian model to accommodate correlations among age-dependent mortality rates. The consequent maps of mortality rates demonstrate reduced variability and smoother patterns when contrasted with naive estimates. Cairns et al. (2011) proposed a Bayesian framework for modeling the joint development over time of mortality rates in two related populations. This model incorporates a mean-reverting stochastic spread to capture variations in death rates between the populations. This framework is particularly valuable when the population of interest is smaller and exhibits more volatile mortality than the other population. Li et al. (2015) argued that forecasting mortality rates using mortality models is sensitive to the sample size. They proposed Bayesian learning to determine age- and gender-specific posterior distributions of the sample size, incorporating it as an extra parameter into the model's parameter space. The forecasts of the Bayesian model were proved to be robust in out-of-sample forecasting. Alexander et al. (2017) introduced a Bayesian hierarchical model for estimating sub-national mortality rates. This model leverages age patterns within mortality curves, combining information across geographic regions and smoothing over time through a Bayesian framework. Aiming to mitigate potential inconsistencies arising from the two-stage fitting process within the Lee-Carter model and its extensions, Lin and Tsai (2022) proposed a solution that simplifies model specification to a random walk with drift model. They further structured a multi-level hierarchical Bayesian framework to model mortality rates for multi-populations with shared characteristics. In the era of rapid internationalization, this model showed effective in addressing the complexities of mortality dependencies among diverse populations and improving the reliability of predictive performance.

2.3 Mortality model with jump

While the jump model has been proven effective in capturing high-severity but low-frequency events in financial mathematics, relatively few studies have applied it to jump events in mortality data. Most studies discussing mortality models still overlook the impact of adverse mortality risk. To understand the potential of the jump model in this context, we examine the existing literature on its use with mortality data. Cox et al. (2006), as well as Lin and Cox (2008), argued that the failure to account for extreme events in mortality models, a prevalent issue in most existing studies, poses a significant problem. As these models are designed to mitigate adverse mortality risks, neglecting such events weakens the life insurer's financial strength and leaves investors vulnerable to significant financial losses. Cox et al. (2006) pioneered an approach, commonly used in financial mathematics, that employs the geometric Brownian motion (GBM) with a compound Poisson process to capture the dynamics in historical mortality data. In this approach, the GBM captures the normal pattern of mortality rates, while the compound Poisson process models the dynamics of adverse events that impact mortality. Lin and Cox (2008) built upon this foundation by proposing an extended model capable of addressing the short-term, transitory effects

of such jumps. Notably, while this model showed valuable for pricing mortality-linked insurance securities, it overlooked potential correlations within mortality data. Recognizing this limitation, Chen and Cox (2009) developed a new Lee-Carter-based model that not only incorporates transitory effects, but also explicitly takes into account the correlations of data. Their comprehensive exploration of both permanent and transitory jump effects reveals that the latter may be the more suitable choice for accurate mortality predictions and securitization analysis. Zhou et al. (2013) highlighted that extreme events can significantly impact mortality-linked securities designed for hedging catastrophic mortality risk. They then developed a two-population mortality model with transitory jump effects and used it to examine how mortality jumps may affect the supply and demand of such securities. Liu and Li (2015) challenged the common assumption in existing models of identical distributions for jump effects and general mortality improvements across ages, noting a misalignment with historical data. Therefore, they proposed a model introducing a modified Lee-Carter variant that captures the age pattern of mortality jumps through a distinct set of parameters.

This project aims to construct a robust predictive mortality model for multi-population, which builds on the findings of Lin and Tsai (2022). We specifically analyze three primary jump effects: the permanent jump effect (Cox et al. (2006)), the independent-correlation transitory jump effect (Lin and Cox (2008)), and the dependent-correlation transitory jump effect (Chen and Cox (2009)). Our ultimate goal is to identify which type of jump effect yields the most significant improvement in forecasting future mortality rates.

Chapter 3

Framework

3.1 Mortality models

In this chapter, we introduce the mortality models and Bayesian model, accompanied by corresponding mathematical notations, definitions, and related derivations to set the stage for further discussion. Here are some fundamental notations and definitions.

Denote $p_{x,t,i}$ as the probability that an individual aged x in year t and population i will survive one year, and $q_{x,t,i}$ as the complementary probability that they will die within one year, i.e., $q_{x,t,i} = 1 - p_{x,t,i}$. We make the assumption that deaths are uniformly distributed within each integer age x and integer year t; that is, for any $0 \leq s < 1$, ${}_sq_{x,t,i} = s \cdot q_{x,t,i}$. This assumption allows us to derive the relationship between the central death rate, $m_{x,t,i}$, and the one-year death probability, $q_{x,t,i}$ as follows:

$$m_{x,t,i} = \frac{q_{x,t,i}}{\int_0^1 {}_sp_{x,t,i}ds} = \frac{q_{x,t,i}}{\int_0^1 (1 - {}_sq_{x,t,i})\,ds} = \frac{q_{x,t}}{\int_0^1 (1 - s \cdot q_{x,t,i})\,ds} = \frac{q_{x,t,i}}{1 - q_{x,t,i}/2},$$

which implies

$$q_{x,t,i} = \frac{m_{x,t,i}}{1 + m_{x,t,i}/2}.$$

This widely adopted assumption simplifies the conversion of mortality data between $q_{x,t,i}$ and $m_{x,t,i}$. Consider a study age-year window $[x_L, x_U] \times [T_1, T_2]$, where mortality rates are available for a specific population i, and a fitting window $[x_L, x_U] \times [t_L, t_U]$, designed for model fitting, where $T_1 \leq t_L$ and $t_U \leq T_2$.

3.1.1 Random walk with drift (RW) model

Lin and Tsai (2022) demonstrated that the commonly used age-period (AP) specification for modeling the dynamics of the logarithm of central death rates, denoted as $\ln(m_{x,t,i})$, can be represented as a random walk with a drift term $\theta_{x,i}$, as follows:

$$\ln(m_{x,t,i}) = \ln(m_{x,t-1,i}) + \theta_{x,i} + \sigma_{x,i} \cdot \epsilon_{x,t,i}, \quad (x,\, t) \in [x_L,\, x_U] \times [t_L + 1,\, t_U]. \tag{3.1}$$

Assume that the dynamics of the central death rate are driven by a geometric Brownian motion (GBM) as

$$\frac{dm_{x,t,i}}{m_{x,t,i}} = \alpha_{x,i} \cdot dt + \sigma_{x,i} \cdot dW_{x,t,i}, \tag{3.2}$$

where $\alpha_{x,i}$ is the drift term, $\sigma_{x,i}$ is the volatility, and $W_{x,t,i}$ is a Wiener process (that is, $W_{x,t,i}$ is normally distributed with mean 0 and variance t). Then the logarithms of central death rate in years t and $t-1$, by the *Itô*'s Lemma, become

$$\ln(m_{x,t,i}) = \ln(m_{x,0,i}) + \left(\alpha_{x,i} - \frac{1}{2}\sigma_{x,i}^2\right) \cdot t + \sigma_{x,i} \cdot W_{x,t,i}$$

and

$$\ln(m_{x,t-1,i}) = \ln(m_{x,0,i}) + \left(\alpha_{x,i} - \frac{1}{2}\sigma_{x,i}^2\right) \cdot (t-1) + \sigma_{x,i} \cdot W_{x,t-1,i},$$

respectively, which imply

$$\ln(m_{x,t,i}) = \ln(m_{x,t-1,i}) + \left(\alpha_{x,i} - \frac{1}{2}\sigma_{x,i}^2\right) + \sigma_{x,i} \cdot \epsilon_{x,t,i}, \ (x,\, t) \in [x_L,\, x_U] \times [t_L+1,\, t_U], \tag{3.3}$$

where $\epsilon_{x,t,i} = W_{x,t,i} - W_{x,t-1,i}$ for $t = t_L + 1,\ t_L + 2, \ldots$, by the properties of independent and stationary increments of the Wiener process, are independent and identical normal random variables with mean 0 and variance 1.

Note that Equation (3.3) closely resembles Equation (3.1), and forms the basis of the jump models introduced by Cox et al. (2006) and Lin and Cox (2008). To maintain consistency with the jump models, we transform the specification proposed by Lin and Tsai (2022)) from Equation (3.1) into Equation (3.3).

- $\ln(m_{x,t,i})$: the logarithm of the central death rate at age x and time t for population i.
- $\ln(m_{x,t-1,i})$: the logarithm of the central death rate at the same age x and the previous time point, $t-1$ for population i.
- $\alpha_{x,i}$: the instantaneous expected central death rate for age x and population i, which characterizes the expected rate of change in death rates at this age and population.
- $\sigma_{x,i}$: the instantaneous volatility of the central death rate for age x and population i, which quantifies the level of randomness or fluctuations in central death rates at this age and population.
- $\epsilon_{x,t,i}$: the time trend errors which are assumed independent and identically distributed standard normal random variables for fixed values of x and i, where t ranges from t_L to t_U.

Equivalently, the yearly decrements in the logarithm of the central death rate, denoted as $Y_{x,t,i}|\Lambda \triangleq \ln(m_{x,t+1,i}) - \ln(m_{x,t,i})$ for $t = t_L, t_L + 1, \ldots$, are also assumed to be independent and identically normally distributed with mean $(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i})$ and variance $\sigma^2_{x,i}$, where Λ is a random variable, representing risk parameter(s), for a specific study population.

Forecasting future mortality rates with Equation (3.2) only requires estimating the values of $\alpha_{x,i}$ and $\sigma_{x,i}$, which is relatively simple compared to other models. However, as Lin and Tsai (2022) mentioned, directly forecasting mortality rates with this model would likely lead to inaccurate results. Therefore, in accordance with their proposal, we incorporate Bayesian frameworks to capture the co-movement across ages and populations, aiming to improve forecasting performance. This approach is discussed in detail in Section 3.4.

3.1.2 Random walk model with permanent jump effects (RWPJ)

Combining Equation (3.2) with a compound Poisson process, we model jump events following the methodologies proposed by Cox et al. (2006) and Lin and Cox (2008). For the sake of simplicity[1], we assume that there is at most one jump event in each year for all ages and population i, denoting $N_{x,t,i}$ the jump frequency variable with a jump probability $\eta_{x,i}$.

$$N_{x,t+1,i} = \begin{cases} 1, & \text{with probability } \eta_{x,i}; \\ 0, & \text{with probability } 1 - \eta_{x,i}. \end{cases}$$

When an abnormal mortality shock occurs, its severity is captured by a jump severity variable denoted as $\gamma_{x,t,i}$, representing the magnitude of the mortality increase during the shock. In our model, we assume that $\gamma_{x,t,i}$s are independent and identically distributed log-normal random variables with mean $\mu_{x,i}$ and standard deviation $s_{x,i}$ for fixed values of x and i. Furthermore, $\gamma_{x,t,i}$ is assumed to be independent of the jump frequency variable $N_{x,t,i}$. In other words, the yearly decrement in the logarithm of the central death rate can be redefined as follows:

$$\begin{aligned} Y_{x,t,i} &= \ln(m_{x,t+1,i}) - \ln(m_{x,t,i}) \\ &= \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+1,i} + N_{x,t+1,i} \cdot \ln(\gamma_{x,t+1,i}) \\ &= \begin{cases} \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+1,i}, & \text{(Scenario 1);} \\ \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+1,i} + \ln(\gamma_{x,t+1,i}), & \text{(Scenario 2).} \end{cases} \end{aligned}$$

[1]While we can use a Poisson random variable to model the jump frequency, Chen and Cox (2009) argued that this approach does not offer new insights into the comparison between permanent and transitory jump effects and may unnecessarily complicate the mathematical framework.

The distribution of $Y_{x,t,i}$ can be analyzed into two scenarios. In the first scenario, when $N_{x,t,i} = 0$, $Y_{x,t,i}$ follows an independent and identically normal distribution with a mean of $(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i})$ and a variance of $\sigma^2_{x,i}$. Conversely, in the second scenario, when $N_{x,t,i} = 1$, the mean of $Y_{x,t,i}$ becomes $(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}) + \mu_{x,i}$, and its variance increases to $\sigma^2_{x,i} + s^2_{x,i}$.

3.1.3 Random walk with independent and transitory jump effects (RWITJ)

Lin and Cox (2008) argued that most mortality jumps, such as the 1918 flu or the 2004 Indian Ocean earthquake and tsunami, are one-time events. However, mortality generally improves over a long period. Additionally, they argued that the probability of experiencing a catastrophic mortality event is extremely low, making data correlations negligible. Consequently, this implies that the mortality pattern following a shock is entirely independent of the pattern observed during the shock itself. The following section introduces their modeling framework. We first define jump frequencies for years t and $t+1$ as

$$N_{x,t,i} = \begin{cases} 1, & \text{with probability } \eta_{x,i}, \\ 0, & \text{with probability } 1-\eta_{x,i}; \end{cases} \quad N_{x,t+1,i} = \begin{cases} 1, & \text{with probability } \eta_{x,i}, \\ 0, & \text{with probability } 1-\eta_{x,i}. \end{cases} \tag{3.4}$$

Considering Equation (3.4), as a jump impulse in $[t-1,\, t]$ is independent of that in $[t,\, t+1]$, four distinct scenarios emerge. These scenarios include two consecutive periods with and without a mortality jump. We demonstrate that all four scenarios conform to normal distributions, each with distinct means and variances.

$$\begin{aligned} Y_{x,t,i} &= \ln(m_{x,t+1,i}) - \ln(m_{x,t,i}) \\ &= \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i}\cdot\epsilon_{x,t+1,i} + N_{x,t+1,i}\cdot\ln(\gamma_{x,t+1,i}) - N_{x,t,i}\cdot\ln(\gamma_{x,t,i}) \\ &= \begin{cases} \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i}\cdot\epsilon_{x,t+1,i}, & \text{(Scenario 1);} \\ \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i}\cdot\epsilon_{x,t+1,i} - \ln(\gamma_{x,t,i}), & \text{(Scenario 2);} \\ \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i}\cdot\epsilon_{x,t+1,i} + \ln(\gamma_{x,t+1,i}), & \text{(Scenario 3);} \\ \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i}\cdot\epsilon_{x,t+1,i} + \ln(\gamma_{x,t+1,i}) - \ln(\gamma_{x,t,i}), & \text{(Scenario 4).} \end{cases} \end{aligned}$$

In Table 3.1, we summarize all the scenarios alongside their corresponding means, variances, and probabilities.

3.1.4 Random walk with dependent and transitory jump effects (RWDTJ)

Chen and Cox (2009) suggested that transitory jump effects introduce data dependence, and neglecting this dependence in the model of Lin and Cox (2008) could lead to inaccurate parameter estimation. They proposed a model accounting for this dependent data struc-

Table 3.1: Potential jump events under RWITJ

Scenario	$\boldsymbol{N_{x,t,i}}$	$\boldsymbol{N_{x,t+1,i}}$	$\boldsymbol{E[Y_{x,t,i}]}$	$\boldsymbol{Var[Y_{x,t,i}]}$	***Probability***
1	**0**	**0**	$\alpha_{x,i} - \sigma^2_{x,i}/2$	$\sigma^2_{x,i}$	$(1-\eta_{x,i})^2$
2	**1**	**0**	$\alpha_{x,i} - \sigma^2_{x,i}/2 - \mu_{x,i}$	$\sigma^2_{x,i} + s^2_{x,i}$	$\eta_{x,i} \cdot (1-\eta_{x,i})$
3	**0**	**1**	$\alpha_{x,i} - \sigma^2_{x,i}/2 + \mu_{x,i}$	$\sigma^2_{x,i} + s^2_{x,i}$	$(1-\eta_{x,i}) \cdot \eta_{x,i}$
4	**1**	**1**	$\alpha_{x,i} - \sigma^2_{x,i}/2$	$\sigma^2_{x,i} + 2\,s^2_{x,i}$	$(\eta_{x,i})^2$

ture. This section adopts a similar approach by incorporating the random walk model with dependent and transitory jump effects.

Generally, the model follows a structure similar to RWITJ. However, the inherent dependence between consecutive data points necessitates analyzing two consecutive periods simultaneously. In other words, we need to consider the year-on-year changes in the yearly decrement of the logarithm of the central death rate, represented by $Y_{x,t,i}$ at time t and $Y_{x,t+1,i}$ at time $t+1$,

$$\begin{aligned} Y_{x,t,i} &= \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+1,i} + N_{x,t+1,i} \cdot \ln(\gamma_{x,t+1,i}) - N_{x,t,i} \cdot \ln(\gamma_{x,t,i}), \\ Y_{x,t+1,i} &= \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+2,i} + N_{x,t+2,i} \cdot \ln(\gamma_{x,t+2,i}) - N_{x,t+1,i} \cdot \ln(\gamma_{x,t+1,i}). \end{aligned} \tag{3.5}$$

Observing Equation (3.5), when $N_{x,t+1,i} = 0$, $\ln(\gamma_{x,t+1,i})$ disappears. In this case, both $Y_{x,t,i}$ and $Y_{x,t+1,i}$ are not affected by the same-period jump factor, and $Y_{x,t,i}$ and $Y_{x,t+1,i}$ are assumed to be independent. On the other hand, when $N_{x,t+1,i} = 1$, both $Y_{x,t,i}$ and $Y_{x,t+1,i}$ are driven by $\ln(\gamma_{x,t+1,i})$, leading a correlation between $Y_{x,t,i}$ and $Y_{x,t+1,i}$.

Next, we sequentially discuss the dynamics of $Y_{x,t+1,i}$ under the scenarios of $N_{x,t+1,i} = 0$ and $N_{x,t+1,i} = 1$. When $N_{x,t+1,i} = 0$, $Y_{x,t+1,i}$ is only driven by $N_{x,t+2,i} \cdot \ln(\gamma_{x,t+2,i})$, that is,

$$\begin{aligned} Y_{x,t+1,i} &= \ln(m_{x,t+2,i}) - \ln(m_{x,t+1,i}) \\ &= \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+2,i} + N_{x,t+2,i} \cdot \ln(\gamma_{x,t+2,i}) \\ &= \begin{cases} \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+2,i}, & \text{(Scenario 1);} \\ \left(\alpha_{x,i} - \frac{1}{2}\sigma^2_{x,i}\right) + \sigma_{x,i} \cdot \epsilon_{x,t+2,i} + \ln(\gamma_{x,t,i}), & \text{(Scenario 2).} \end{cases} \end{aligned} \tag{3.6}$$

Conversely, when $N_{x,t+1,i} = 1$, $Y_{x,t+1,i}$ becomes correlated with $Y_{x,t+2,i}$ and is driven by $\ln(\gamma_{x,t,i})$, $\ln(\gamma_{x,t+1,i})$, and $\ln(\gamma_{x,t+2,i})$. To illustrate this, summing $Y_{x,t,i}$ and $Y_{x,t+1,i}$ in Equation (3.5) and retaining only $Y_{x,t+1,i}$ on the left-hand side of the equation gives

$$
\begin{aligned}
Y_{x,\,t+1,\,i} = &- Y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \sigma_{x,\,i}\cdot(\epsilon_{x,\,t+1,\,i} + \epsilon_{x,\,t+2,\,i}) \\
&+ N_{x,\,t+2,\,i}\cdot\ln(\gamma_{x,\,t+2,\,i}) - N_{x,\,t,\,i}\cdot\ln(\gamma_{x,\,t,\,i}).
\end{aligned} \tag{3.7}
$$

Using Equation (3.7), we analyze four different combinations of $N_{x,\,t,\,i}$ and $N_{x,\,t+2,\,i}$ under the condition $N_{x,\,t+1,\,i} = 1$ to derive four distinct scenarios. In each scenario, $y_{x,\,t,\,i}$ represents the observed value of $Y_{x,\,t,\,i}$. The details of these scenarios are outlined below:

$$
Y_{x,\,t+1,\,i} = \begin{cases}
-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \sigma_{x,\,i}\cdot(\epsilon_{x,\,t+1,\,i} + \epsilon_{x,\,t+2,\,i}), & \text{(Scenario 3);} \\
-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \sigma_{x,\,i}\cdot(\epsilon_{x,\,t+1,\,i} + \epsilon_{x,\,t+2,\,i}) - \ln(\gamma_{x,\,t,\,i}), & \text{(Scenario 4);} \\
-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \sigma_{x,\,i}\cdot(\epsilon_{x,\,t+1,\,i} + \epsilon_{x,\,t+2,\,i}) + \ln(\gamma_{x,\,t+2,\,i}), & \text{(Scenario 5);} \\
-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \sigma_{x,\,i}\cdot(\epsilon_{x,\,t+1,\,i} + \epsilon_{x,\,t+2,\,i}) + \ln(\gamma_{x,\,t+2,\,i}) - \ln(\gamma_{x,\,t,\,i}), & \\
& \text{(Scenario 6).}
\end{cases}
$$

Building upon the normal distribution property, all scenarios within the RWDTJ model can be represented as normal distributions with their respective means and variances. Table 3.2 summarizes the means, variances, and corresponding probabilities for all possible scenarios. It is worth recalling that when $N_{x,\,t+1,\,i} = 0$, $Y_{x,\,t+1,\,i}$ is only driven by $\ln(\gamma_{x,\,t+2,\,i})$. In other words, it does not matter what the value of $N_{x,\,t,\,i}$ is in the first two scenarios.

Table 3.2: Potential jump events under RWDTJ

Scenario	$\boldsymbol{N_{x,\,t,\,i}}$	$\boldsymbol{N_{x,\,t+1,\,i}}$	$\boldsymbol{N_{x,\,t+2,\,i}}$	$\boldsymbol{E[Y_{x,\,t+1,\,i}]}$	$\boldsymbol{Var[Y_{x,\,t+1,\,i}]}$	***Probability***
1	–	**0**	**0**	$\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}$	$\sigma^2_{x,\,i}$	$(1-\eta_{x,\,i})^2$
2	–	**0**	**1**	$\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i}$	$\sigma^2_{x,\,i} + s^2_{x,\,i}$	$\eta_{x,\,i}\,(1-\eta_{x,\,i})$
3	**0**	**1**	**0**	$-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right)$	$2\,\sigma^2_{x,\,i}$	$\eta_{x,\,i}\,(1-\eta_{x,\,i})^2$
4	**1**	**1**	**0**	$-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) - \mu_{x,\,i}$	$2\,\sigma^2_{x,\,i} + s^2_{x,\,i}$	$(\eta_{x,\,i})^2\,(1-\eta_{x,\,i})$
5	**0**	**1**	**1**	$-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right) + \mu_{x,\,i}$	$2\,\sigma^2_{x,\,i} + s^2_{x,\,i}$	$(\eta_{x,\,i})^2\,(1-\eta_{x,\,i})$
6	**1**	**1**	**1**	$-y_{x,\,t,\,i} + 2\left(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}\right)$	$2\,\sigma^2_{x,\,i} + 2\,s^2_{x,\,i}$	$(\eta_{x,\,i})^3$

3.2 Mortality forecasting

In Lin and Tsai (2022)'s random walk with drift model, estimating the drift as the slope of the linear prediction function, $\hat{E}[Y_{x,\,t,\,i}]$, is relatively straightforward. It can be calculated by averaging the past observations $\{y_{x,\,1,\,i},\, y_{x,\,2,\,i}, \ldots\}$ of $\{Y_{x,\,1,\,i},\, Y_{x,\,2,\,i}, \ldots\}$. In contrast, our proposed framework for forecasting future mortality rates requires the estimation of model parameters, including $\alpha_{x,\,t}$, $\sigma_{x,\,t}$, $\eta_{x,\,t}$, $\mu_{x,\,t}$, and $s_{x,\,t}$. As demonstrated in the previous section, the distribution of $Y_{x,\,t,\,i}$ within RW, RWPJ, RWITJ, and RWDTJ models follows a mixture of normal distributions, each characterized by distinct means and variances. To estimate these model parameters, we employ the maximum likelihood estimation method. Detailed

derivations of the likelihood function for each individual model can be found in Appendix B.

After obtaining the maximum likelihood estimates (MLEs), we proceed to calculate the mean of the yearly decrement in the logarithm of the central death rate under each model. This yields an estimate for the slope parameter, denoted as $\hat{\theta}_{x,i}$. The specific prediction model for the logarithm of the predicted central death rate for lives aged x in year $t_U + \tau$ and population i can be formulated as follows:

$$\ln(\hat{m}_{x,t_U+\tau,i}) = \ln(m_{x,t_U,i}) + \hat{\theta}_{x,i} \cdot \tau, \quad \tau = 1, 2, \ldots, \tag{3.8}$$

where $\ln(m_{x,t_U,i})$ and $\hat{\theta}_{x,i}$ represent the intercept and slope of the linear prediction function, respectively. The slope varies across different models and can be estimated using $\hat{E}[Y_{x,t,i}]$ under the RW, RWPJ and RWITJ models, and $\hat{E}[Y_{x,t,i}|Y_{x,t-1,i} = y_{x,t-1,i}]$ under the RWDTJ model. The following provide estimates of the slope under each respective model.

- The slope under the RW model is given by

$$\hat{\theta}_{x,i}^{RW} = \hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2.$$

- The slope under the RWPJ model is given by

$$\begin{aligned}\hat{\theta}_{x,i}^{RWPJ} &= \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) \cdot (1 - \hat{\eta}_{x,i}) + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 + \hat{\mu}_{x,i}\right) \cdot \hat{\eta}_{x,i} \\ &= \hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 + \hat{\eta}_{x,i} \cdot \hat{\mu}_{x,i} = \hat{\theta}_{x,i}^{RW} + \hat{\eta}_{x,i} \cdot \hat{\mu}_{x,i}.\end{aligned}$$

- The slope under the RWITJ model is given by

$$\begin{aligned}\hat{\theta}_{x,i}^{RWITJ} &= \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) \cdot (1 - \hat{\eta}_{x,i})^2 + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 - \hat{\mu}_{x,i}\right) \cdot \hat{\eta}_{x,i}\,(1 - \hat{\eta}_{x,i}) \\ &\quad + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 + \hat{\mu}_{x,i}\right) \cdot \hat{\eta}_{x,i}\,(1 - \hat{\eta}_{x,i}) + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) \cdot (\hat{\eta}_{x,i})^2 \\ &= \hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 = \hat{\theta}_{x,i}^{RW}.\end{aligned}$$

- The slope under the RWDTJ model is given by

$$\begin{aligned}
\hat{\theta}_{x,i}^{RWDTJ} = \frac{1}{t_U - t_L} \cdot \sum_{t=t_{L+1}}^{t_U} \Bigg\{ & \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) \cdot (1-\hat{\eta}_{x,i})^2 \\
& + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2 + \hat{\mu}_{x,i}\right) \cdot \hat{\eta}_{x,i}\,(1-\hat{\eta}_{x,i}) \\
& + \left[-y_{x,t,i} + 2\left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right)\right] \cdot \hat{\eta}_{x,i}\,(1-\hat{\eta}_{x,i})^2 \\
& + \left[-y_{x,t,i} + 2\left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) - \hat{\mu}_{x,i}\right] \cdot (\hat{\eta}_{x,i})^2\,(1-\hat{\eta}_{x,i}) \\
& + \left[-y_{x,t,i} + 2\left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) + \hat{\mu}_{x,i}\right] \cdot (\hat{\eta}_{x,i})^2\,(1-\hat{\eta}_{x,i}) \\
& + \left[-y_{x,t,i} + 2\left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right)\right] \cdot (\hat{\eta}_{x,i})^3 \Bigg\} \\
= {} & -\overline{y}_{x,\bullet,i} \cdot \hat{\eta}_{x,i} + \left(\hat{\alpha}_{x,i} - \frac{1}{2}\hat{\sigma}_{x,i}^2\right) \cdot (1+\hat{\eta}_{x,i}) + \hat{\mu}_{x,i} \cdot \hat{\eta}_{x,i} \cdot (1-\hat{\eta}_{x,i}),
\end{aligned}$$

where $\overline{y}_{x,\bullet,i} = \frac{1}{t_U - t_L} \cdot \sum_{t=t_{L+1}}^{t_U} y_{x,t,i}$ is the sample mean of $y_{x,t_L+1,i}, y_{x,t_L+2,i}, \ldots, y_{x,t_U,i}$.

Our extension leverages the hierarchical Bayesian framework (Lin and Tsai (2022)) to improve forecasting performance by incorporating age and population co-movement. This approach overcomes the limitations of the random walk with drift model, which may result in inferior forecasting performance.

3.3 Bayesian (BS) model

Building upon the four random walk models introduced earlier, with or without jump events, and classified based on the nature of jump events, this section explores their potential applications in forecasting mortality rates. The new models in this section address the limitations of previous approaches by considering the co-movement across the study ages and multiple populations using the Bayesian framework proposed by Lin and Tsai (2022).

3.3.1 Bayesian model for single population

Observing Equation (3.8), the logarithm of the predicted central death rate can be modeled as a linear function of year τ, with the slope representing the downward time trend in mortality. To refine this slope estimate, we follow the Bayesian methodology of Lin and Tsai (2022).

Assuming that we have $\tilde{y} = (y_1, y_2, \ldots, y_n)$, a realization from $\tilde{Y} = (Y_1, Y_2, \ldots, Y_n)$. Let $Y_t|\Lambda, t = 1, 2, \ldots, n, n+1$ be random variables, each having a probability density $f_{Y_t|\Lambda}(y_t|\lambda_x)$ at y_t, where Λ is assumed to be a discrete random variable in this project, representing

the risk parameter for a specific study population. Here, we skip the age and population subscripts, x and i, in Y_t and y_t to simplify the notations. We also introduce and define mathematical notations for clarity in subsequent derivations:

- $f_{Y_{n+1}|\widetilde{Y}}(y_{n+1}|\widetilde{y})$: the predictive probability density function, providing the conditional probability density of Y_{n+1} at y_{n+1} given observations $\widetilde{Y} = \widetilde{y}$.
- $\pi_\Lambda(\lambda_x)$: the prior probability mass function of Λ at λ_x, representing the initial belief about the possible values of Λ before observing any data.
- $\pi_{\Lambda|\widetilde{y}}(\lambda_x|\widetilde{y})$: the posterior probability mass function, the conditional probability density of Λ at λ_x given $\widetilde{Y} = \widetilde{y}$, representing our updated belief about Λ after observing the data $\widetilde{Y} = \widetilde{y}$.

To calculate the mean of the predictive distribution, the Bayesian estimate, we begin by deriving the predictive probability density function of Y_{n+1} at y_{n+1} given $\widetilde{Y} = \widetilde{y}$. Depending on whether $\tilde{Y}$ is an independent or dependent random vector, we distinguish two cases.

Case 1: $\tilde{Y}$ is an independent random vector

$$\begin{aligned}
f_{Y_{n+1}|\widetilde{Y}}(y_{n+1}|\widetilde{y}) &= \frac{f_{\widetilde{Y},\, Y_{n+1}}(\widetilde{y},\, y_{n+1})}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{\widetilde{Y},\, Y_{n+1}|\Lambda}(\widetilde{y},\, y_{n+1}|\lambda_x) \cdot \pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&\overset{\text{ind.}}{=} \frac{\sum_x \prod_{t=1}^{n+1} f_{Y_t|\Lambda}(y_t|\lambda_x) \cdot \pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{Y_{n+1}|\Lambda}(y_{n+1}|\lambda_x) \cdot f_{\widetilde{Y},\,\Lambda}(\widetilde{y},\, \lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&\overset{2}{=} \frac{\sum_x f_{Y_{n+1}|\Lambda}(y_{n+1}|\lambda_x) \cdot \pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y}) \cdot f_{\widetilde{Y}}(\widetilde{y})}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \sum_x f_{Y_{n+1}|\Lambda}(y_{n+1}|\lambda_x) \cdot \pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y}).
\end{aligned} \tag{3.9}$$

Case 2: $\tilde{Y}$ is a dependent random vector

[2]Applying Bayes's theorem, we can derive $f_{\widetilde{Y}|\Lambda}(\widetilde{y}|\lambda_x) \cdot \pi_\Lambda(\lambda_x) = f_{\widetilde{Y},\,\Lambda}(\widetilde{y}, \lambda_x) = \pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y}) \cdot f_{\widetilde{Y}}(\widetilde{y})$.

$$
\begin{aligned}
f_{Y_{n+1}|\widetilde{Y}}(y_{n+1}|\widetilde{y}) &= \frac{f_{\widetilde{Y},\,Y_{n+1}}(\widetilde{y},\,y_{n+1})}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{\widetilde{Y},\,Y_{n+1}|\Lambda}(\widetilde{y},\,y_{n+1}|\lambda_x)\cdot\pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&\overset{\text{(B.3)}}{=} \frac{\sum_x\{[\prod_{t=1}^{n} f_{Y_{t+1}|Y_t,\,\Lambda}(y_{t+1}|y_t,\,\lambda_x)]\cdot f_{Y_1|\Lambda}(y_1|\lambda_x)\}\cdot\pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{Y_{n+1}|Y_n,\,\Lambda}(y_{n+1}|y_n,\,\lambda_x)\cdot\{[\prod_{t=1}^{n-1} f_{Y_{t+1}|Y_t,\,\Lambda}(y_{t+1}|y_t,\,\lambda_x)]\cdot f_{Y_1|\Lambda}(y_1|\lambda_x)\}\cdot\pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&\overset{\text{(B.3)}}{=} \frac{\sum_x f_{Y_{n+1}|Y_n,\,\Lambda}(y_{n+1}|y_n,\,\lambda_x)\cdot f_{\widetilde{Y}|\Lambda}(\widetilde{y}|\lambda_x)\cdot\pi_\Lambda(\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{Y_{n+1}|Y_n,\,\Lambda}(y_{n+1}|y_n,\,\lambda_x)\cdot f_{\widetilde{Y},\,\Lambda}(\widetilde{y},\,\lambda_x)}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \frac{\sum_x f_{Y_{n+1}|Y_n,\,\Lambda}(y_{n+1}|y_n,\,\lambda_x)\cdot\pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y})\cdot f_{\widetilde{Y}}(\widetilde{y})}{f_{\widetilde{Y}}(\widetilde{y})} \\
&= \sum_x f_{Y_{n+1}|Y_n,\,\Lambda}(y_{n+1}|y_n,\,\lambda_x)\cdot\pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y}). \qquad (3.10)
\end{aligned}
$$

Next, building upon the insights from Equations (3.9) and (3.10), we can derive the Bayesian estimate under each case as follows:

Case 1: $\tilde{Y}$ is an independent random vector

$$
\begin{aligned}
E\left[Y_{n+1}|\tilde{Y}=\tilde{y}\right] &= \int y_{n+1}\cdot f_{Y_{n+1}|\tilde{Y}}\left(y_{n+1}|\tilde{y}\right)\,dy_{n+1} \\
&\overset{(3.9)}{=} \int y_{n+1}\cdot\left[\sum_x f_{Y_{n+1}|\Lambda}\left(y_{n+1}|\lambda_x\right)\cdot\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y})\right]\,dy_{n+1} \\
&= \sum_x\left[\int y_{n+1}\cdot f_{Y_{n+1}|\Lambda}\left(y_{n+1}|\lambda_x\right)\,dy_{n+1}\right]\cdot\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y}) \\
&= \sum_x\left[E\left(Y_{n+1}|\Lambda=\lambda_x\right)\right]\cdot\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y}). \qquad (3.11)
\end{aligned}
$$

Case 2: $\tilde{Y}$ is a dependent random vector

$$\begin{aligned} E\left[Y_{n+1}|\tilde{Y}=\tilde{y}\right] &= \int y_{n+1}\cdot f_{Y_{n+1}|\tilde{Y}}\left(y_{n+1}|\tilde{y}\right)\, dy_{n+1} \\ &\overset{(3.10)}{=} \int y_{n+1}\cdot\left[\sum_x f_{Y_{n+1}|Y_n,\,\Lambda}\left(y_{n+1}|y_n,\,\lambda_x\right)\cdot\pi_{\Lambda|\widetilde{Y}}(\lambda_x|\widetilde{y})\right] dy_{n+1} \\ &= \sum_x\left[\int y_{n+1}\cdot f_{Y_{n+1}|Y_n,\,\Lambda}\left(y_{n+1}|y_n,\,\lambda_x\right)\, dy_{n+1}\right]\cdot\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y}) \\ &= \sum_x\left[E\left(Y_{n+1}|Y_n,\,\Lambda=\lambda_x\right)\right]\cdot\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y}). \end{aligned} \tag{3.12}$$

Observing the results of Equation (3.11) and (3.12), we need to calculate $E(Y_{n+1}|\Lambda = \lambda_x) = \theta_{x,\,i}$ for the independent case and $E(Y_{n+1}|Y_n,\,\Lambda=\lambda_x) = \theta_{x,\,i}$ for the dependent case. As outlined in the previous section, the value of $\theta_{x,\,i}$ depends on the model chosen. In the independent case, the potential candidates include $\theta^{RW}_{x,\,i}$, $\theta^{RWPJ}_{x,\,i}$, and $\theta^{RWITJ}_{x,\,i}$, while the dependent case includes $\theta^{RWDTJ}_{x,\,i}$ as a candidate. The other component is the posterior distribution, defined by Bayes' Theorem as $\pi_{\Lambda|\tilde{Y}}(\lambda_x|\tilde{y}) = f_{\tilde{Y},\,\Lambda}(\tilde{y},\,\lambda_x)/f_{\tilde{Y}}(\tilde{y})$. The numerator and denominator can be derived as follows:

$$f_{\tilde{Y},\,\Lambda}(\tilde{y},\lambda_x) = f_{\tilde{Y}|\Lambda}(y_1,\,y_2,\,\ldots,y_n|\lambda_x)\cdot\pi_\Lambda(\lambda_x) \overset{\text{if ind.}}{=} \left[\prod_{t=1}^{n} f_{Y_t|\Lambda}\left(y_t|\lambda_x\right)\right]\cdot\pi_\Lambda(\lambda_x) \tag{3.13}$$

and

$$f_{\tilde{Y}}(\tilde{y}) = \sum_x f_{\tilde{Y}|\Lambda}(y_1,\,y_2,\,\ldots,y_n|\lambda_x)\cdot\pi_\Lambda(\lambda_x) \overset{\text{if ind.}}{=} \sum_x\left[\prod_{t=1}^{n} f_{Y_t|\Lambda}\left(y_t|\lambda_x\right)\right]\cdot\pi_\Lambda(\lambda_x). \tag{3.14}$$

Now, to explore further into random variable Λ for population i, we assume that $Y_t|(\Lambda = \lambda_z)$ follows a density $f_{Y_t|\Lambda}(y_t|\lambda_z)$ for $t = 1, 2, \ldots$ and $z = x_L, \ldots, x_U$, which varies depending on the selected model. For the Random Walk (RW) model, $Y_t|(\Lambda=\lambda_z)$ follows a normal density function characterized by a set of parameters, $\Omega_{z,\,i} = \{\alpha_{z,\,i},\,\sigma_{z,\,i}\}$. In this case, the mean and variance of the normal distribution are $(\alpha_{z,\,i} - \frac{1}{2}\sigma^2_{z,\,i})$ and $\sigma^2_{z,\,i}$, respectively. In contrast, the models incorporating jumps (RWPJ, RWITJ, and RWDTJ) exhibit a mixture normal distribution characterized by a set of parameters, $\Omega_{z,\,i} = \{\alpha_{z,\,i},\,\sigma_{z,\,i},\,\eta_{z,\,i},\,\mu_{z,\,i},\,s_{z,\,i}\}$. The values of $\Omega_{z,\,i}$ under each model are governed by the realization of the risk parameter $\Lambda = \lambda_z$ and can be estimated using the observation $\tilde{y}_{z,\,i} = (y_{z,\,1,\,i}, y_{z,\,2,\,i}, \ldots, y_{z,\,n,\,i})$.

Assuming that a mortality vector $\tilde{y}_{x,\,i} = (y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U,\,i})$ is observed from $\tilde{Y}_{x,\,i} = (Y_{x,\,t_L+1,\,i}, \ldots, Y_{x,\,t_U,\,i})$, the parameters for age x and population i, $\Omega_{x,\,i}$, can be estimated using maximum likelihood estimation. The Bayesian estimates, as given by Equations (3.11)

and (3.12), take the following specific form:

$$
\begin{aligned}
\hat{\theta}^{BS}_{x,\,i} &= \hat{E}[Y_{x,\,n+1,\,i}|\tilde{Y}_{x,\,i} = \tilde{y}_{x,\,i}] \\
&= \sum_{z=x_L}^{x_U} \hat{\theta}_{z,\,i} \cdot \pi_{\Lambda|\tilde{Y}_{x,\,i}}(\lambda_z|\tilde{y}_{x,\,i}) \\
&= \sum_{z=x_L}^{x_U} \hat{\theta}_{z,\,i} \cdot \frac{f_{\tilde{Y}_{x,\,i},\,\Lambda}(\tilde{y}_{x,\,i},\,\lambda_z)}{f_{\tilde{Y}_{x,\,i}}(\tilde{y}_{x,\,i})} \\
&= \begin{cases}
\displaystyle\sum_{z=x_L}^{x_U} \hat{\theta}_{z,\,i} \cdot \frac{\pi_\Lambda(\lambda_z) \cdot \displaystyle\prod_{t=t_L+1}^{t_U} f(y_{x,\,t,\,i}|\hat{\Omega}_{z,\,i})}{\displaystyle\sum_{w=x_L}^{x_U} \pi_\Lambda(\lambda_w) \cdot \prod_{t=t_L+1}^{t_U} f\left(y_{x,\,t,\,i}|\hat{\Omega}_{w,\,i}\right)}, & \text{Case 1,} \\
\displaystyle\sum_{z=x_L}^{x_U} \hat{\theta}_{z,\,i} \cdot \frac{\pi_\Lambda(\lambda_z) \cdot f(y_{x,\,t_L+1,\,i},\, y_{x,\,t_L+2,\,i},\,\ldots,\, y_{x,\,t_U,\,i}|\hat{\Omega}_{z,\,i})}{\displaystyle\sum_{w=x_L}^{x_U} \pi_\Lambda(\lambda_w) \cdot f(y_{x,\,t_L+1,\,i},\, y_{x,\,t_L+2,\,i},\,\ldots,\, y_{x,\,t_U,\,i}|\hat{\Omega}_{w,\,i})}, & \text{Case 2,}
\end{cases}
\end{aligned}
\tag{3.15}
$$

where Case 1 is for the RW, RWPJ and RWITJ models based on an independent random vector $\tilde{Y}_{x,\,i}$, whereas Case 2 is for the RWDTJ model based on a dependent random vector $\tilde{Y}_{x,\,i}$, and their corresponding density functions, f, can be referred to Equations (B.1), (B.2), and (B.3) in the Appendix B for the RWPJ, RWITJ and RWDTJ models, respectively.

Regarding the prior probability function, $\pi_\Lambda(\lambda_z)$, in Equation (3.15), Lin and Tsai (2022) propose two options for the prior probability function:

1. Uniform Distribution: It assumes that Λ is uniformly distributed over all study ages with $\pi_\Lambda(\lambda_z) = 1/m$ for $z = x_L, \ldots, x_U$, where $m = x_U - x_L + 1$.

2. Introducing population size: This approach introduces population size information for age x in year t_U (the last year of the fitting year span $[t_L,\, t_U]$) and the study population i to construct the prior probability function π_Λ.

However, as noted by Lin and Tsai (2022), there is no clear evidence supporting the superiority of one approach over the other. Therefore, in this project, we directly adopt the uniform distribution as our proxy for the prior probability function.

This derived Bayesian slope from Equation (3.15) can then be incorporated into Equation (3.8) to facilitate mortality forecasting. Observing Equation (3.15), we note that it calculates a weighted average of the original $\hat{\theta}_{x,\,i}$s, defined in the previous section, across all study ages within the year range $[t_L,\, t_U]$. It places more emphasis on contributions from

ages that yield higher posterior probabilities at λ_z, while conversely assigning less weight to those with lower posterior probabilities at λ_z. In the numerical analysis section, we will use a prefix of "BS-" to represent the Bayesian slopes. Specifically, we will use $\hat{\theta}_{x,i}^{BS-RW}$, $\hat{\theta}_{x,i}^{BS-RWPJ}$, $\hat{\theta}_{x,i}^{BS-RWITJ}$, and $\hat{\theta}_{x,i}^{BS-RWDTJ}$ to distinguish among our four models that incorporate the Bayesian framework.

3.3.2 Hierarchical Bayesian model for multiple populations

In this subsection, we generalize the Bayesian model to a hierarchical Bayesian model, which not only incorporates the co-movement across ages but also considers multiple populations to refine the slope. Following Lin and Tsai (2022), the available mortality data are stratified into hierarchical levels. The topmost level encompassed multiple population groups, represented by I populations $(1, 2, \ldots, I)$. For example, mortality data might originate from populations with different genders and countries. Moving down one level, the focus could be placed on a specific individual population. Drilling down further, mortality data for the study ages $x_L, \ldots, x_U$ could be observed. At the bottom level, the mortality vector $\tilde{y}_{x,\,i} = (y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U,\,i})$ for each age within that population is observed.

Recalling in the previous subsection, the introduction of the risk parameter Λ aimed to capture the co-movement across different ages within a single population. To extend this concept and account for co-movement across multiple populations, the hyperparameter Ψ is also introduced. Denote $\pi_{\Lambda,\,\Psi}\,(\lambda_z,\,\psi_j) = \pi_{\Lambda|\Psi}\,(\lambda_z|\psi_j) \cdot \pi_{\Psi}(\psi_j)$ the joint prior probability of $(\Lambda,\,\Psi)$ at $(\lambda_z,\,\psi_j)$. We assume that λ_z for age $z = x_L, \ldots, x_U$ is equally likely selected, given $\Psi = \psi_j$ for population j. Similarly, ψ_j for $j = 1, \ldots, I$ is also equally likely chosen. In other words, both $\Lambda|\Psi$ and Ψ are assumed to be uniformly distributed with $\pi_{\Lambda|\Psi} = 1/m$ and $\pi_{\Psi}(\psi_j) = 1/I$ for $z = x_L, \ldots, x_U$ and $j = 1, \ldots, I$. Therefore, $\pi_{\Lambda,\,\Psi}(\lambda_z,\,\psi_j) = 1/(m \cdot I)$ is uniform over the age-population product space $[x_L,\,x_U] \times [1,\,I]$.

Assuming a mortality vector $\tilde{y}_{x,\,i} = (y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U,\,i})$ is observed from $\tilde{Y}_{x,\,i} = (Y_{x,\,t_L+1,\,i}, \ldots, Y_{x,\,t_U,\,i})$, the hierarchical Bayesian estimate, similar to Equation (3.15) for

the Bayesian estimate, takes the following specific form:

$$
\begin{aligned}
\hat{\theta}_{x,i}^{HBS} &= \hat{E}\left[Y_{x,n+1,i}|\tilde{Y}_{x,i} = \tilde{y}_{x,i}\right] \\
&= \sum_{j=1}^{I}\sum_{z=x_L}^{x_U} \hat{\theta}_{z,j} \cdot \pi_{(\Lambda,\Psi)|\tilde{Y}_{x,i}}(\lambda_z,\, \psi_j|\tilde{y}_{x,i}) \\
&= \sum_{j=1}^{I}\sum_{z=x_L}^{x_U} \hat{\theta}_{z,j} \cdot \frac{f_{\tilde{Y}_{x,i},\Lambda,\Psi}(\tilde{y}_{x,i},\, \lambda_z,\, \psi_j)}{f_{\tilde{Y}_{x,i}}(\tilde{y}_{x,i})} \\
&= \begin{cases}
\dfrac{\displaystyle\sum_{j=1}^{I}\sum_{z=x_L}^{x_U} \hat{\theta}_{z,j} \cdot \pi_{\Lambda,\Psi}(\lambda_z,\, \psi_j) \cdot \prod_{t=t_L+1}^{t_U} f(y_{x,t,i}|\hat{\Omega}_{z,j})}{\displaystyle\sum_{h=1}^{I}\sum_{w=x_L}^{x_U} \pi_{\Lambda,\Psi}(\lambda_w,\, \psi_h) \cdot \prod_{t=t_L+1}^{t_U} f(y_{x,t,i}|\hat{\Omega}_{w,h})}, & \text{Case 1,} \\
\dfrac{\displaystyle\sum_{j=1}^{I}\sum_{z=x_L}^{x_U} \hat{\theta}_{z,j} \cdot \pi_{\Lambda,\Psi}(\lambda_z,\, \psi_j) \cdot f(y_{x,t_L+1,i}, y_{x,t_L+2,i}, \ldots, y_{x,t_U,i}|\hat{\Omega}_{z,j})}{\displaystyle\sum_{h=1}^{I}\sum_{w=x_L}^{x_U} \pi_{\Lambda,\Psi}(\lambda_w,\, \psi_h) \cdot f(y_{x,t_L+1,i}, y_{x,t_L+2,i}, \ldots, y_{x,t_U,i}|\hat{\Omega}_{w,h})}, & \text{Case 2,}
\end{cases}
\end{aligned}
\tag{3.16}
$$

where Case 1 is for the RW, RWPJ and RWITJ models based on an independent random vector $\tilde{Y}_{x,i}$, whereas Case 2 is for the RWDTJ model based on a dependent random vector $\tilde{Y}_{x,i}$, and their corresponding density functions, f, can be referred to Equations (B.1), (B.2), and (B.3) in the Appendix B for the RWPJ, RWITJ and RWDTJ models, respectively.

As demonstrated in Equation (3.16), the hierarchical Bayesian estimate can be expressed as a weighted average of the slopes $\hat{\theta}_{z,j}$s, with the weights equal to the posterior probabilities, $\pi_{(\Lambda,\Psi)|\tilde{Y}_{x,i}}(\lambda_z,\, \psi_j|\tilde{y}_{x,i})$, for all ages and populations. Consistent with the Bayesian estimate, we propose four slope candidates of $\hat{\theta}_{x,i}^{HBS}$ based on the nature of the jump: $\hat{\theta}_{x,i}^{HBS-RW}$, $\hat{\theta}_{x,i}^{HBS-RWPJ}$, $\hat{\theta}_{x,i}^{HBS-RWITJ}$, and $\hat{\theta}_{x,i}^{HBS-RWDTJ}$, which will be used for reference in the following chapters. Similarly, by replacing the slope $\hat{\theta}_{x,i}$ in Equation (3.8) with $\hat{\theta}_{x,i}^{HBS}$ in Equation (3.16), we can obtain the logarithm of the predicted central death rate in year $t_U + \tau$ for age x and population i under the hierarchical Bayesian (HBS) model.

Chapter 4

Numerical Illustrations

The mortality data used in this section is sourced from the Human Mortality Database (HMD, www.mortality.org), covering the study ages ranging from 25 to 84 for both male and female populations. The numerical illustration consists of four main parts:

1. Apply the Bayesian model with four jump models to analyze a single population ($I = 1$), considering both genders of the US, the UK, and Japan separately.
2. Apply the hierarchical Bayesian model to assess two populations ($I = 2$), exploring potential co-movements between both genders of each country.
3. Apply the hierarchical Bayesian model to assess six populations ($I = 6$), exploring potential co-movements between both genders of the US, the UK, and Japan.
4. Conduct a global analysis, expanding the study to 32 populations ($I = 32$), including 16 countries worldwide. To comprehensively assess forecasting performance, employ three different grouping methods:

    (a) One-group method: Consider each of the 32 populations individually.
    (b) Two-group method: Classify 32 populations into two groups: developed countries and developing countries.
    (c) Three-group method: Classify 32 populations into three groups: North America, Oceania-Asia, and Europe.

As specified in Chapter 3, consistent data lengths across populations are essential for multi-population analysis. One of our goals is to capture jump phenomena in the dataset. To achieve this, we leverage all accessible data, selecting a shared time interval universally available across different countries. The data intervals for the first and second analyses are presented in Table 4.1. Due to data availability limitations in each country, adjustments are necessary for the starting years in the third and fourth analyses. In the third analysis with $I = 6$, the starting year of the training data is set to 1947; for the fourth analysis with $I = 32$, the starting year is adjusted to 1951.

Table 4.1: Age-year data windows for single/two-population modeling.

Country	**Forecast Years**		
	10 years	**20 years**	**30 years**
US			
Training Set	$[25, 84] \times [1933, 2010]$	$[25, 84] \times [1933, 2000]$	$[25, 84] \times [1933, 1990]$
Test Set	$[25, 84] \times [2011, 2020]$	$[25, 84] \times [2001, 2020]$	$[25, 84] \times [1991, 2020]$
UK			
Training Set	$[25, 84] \times [1900, 2010]$	$[25, 84] \times [1900, 2000]$	$[25, 84] \times [1900, 1990]$
Test Set	$[25, 84] \times [2011, 2020]$	$[25, 84] \times [2001, 2020]$	$[25, 84] \times [1991, 2020]$
Japan			
Training Set	$[25, 84] \times [1947, 2010]$	$[25, 84] \times [1947, 2000]$	$[25, 84] \times [1947, 1990]$
Test Set	$[25, 84] \times [2011, 2020]$	$[25, 84] \times [2001, 2020]$	$[25, 84] \times [1991, 2020]$

4.1 Comparisons of models for a single population

In this section, we analyze the predictive accuracy of various mortality models, including LC, LCr, BS-RW, BS-RWPJ, BS-RWITJ, and BS-RWDTJ, specifically for both genders of the US, the UK, and Japan. Figure 4.1 for 10 years of forecast presents a subset of slopes for each model plotted against age x for the age-year windows [25, 84] $\times$ [t_L, 2010], where t_L varies by country due to data availability, and the details can be referred to Table 4.1.

The analysis in Figure 4.1 reveals negative slopes for all study ages in both genders of the US, the UK, and Japan, indicating that all ages and populations can expect a decrease in the projected central death rates over time. The specific rate of improvement, as indicated by the steepness of the negative slope, varies. A steeper slope implies faster decline in mortality rates. In each country (US, UK, Japan), females exhibit a more rapid rate of improvement compared to males. Generally, Japan experiences the fastest decline, followed by the UK, and then the US.

Figures 4.2 and 4.3 present analyses of the Bayesian posterior distribution Λ given $\tilde{Y} = \tilde{y}_{x,\,i}$ for both genders of three countries with various jump models, using the same age-year windows as in Figure 4.1. By applying the Bayesian theory to calculate the posterior probabilities $\pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_{x,\,i})$ for age $z = 25, \ldots, 84$ with $x = 25$ and $x = 84$, respectively, we examine the posterior probabilities $\pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_{x,\,i})$. As expected, the largest values occur at $z = x = 25$ and $z = x = 84$, respectively, supporting the hypothesis that the distribution behind the observation $\tilde{y}_{x,\,i}$ is most likely from density f with parameter set $\Omega_{x,\,i}$ under each jump model. Figures 4.2 and 4.3 also display posterior likelihoods $\pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_{x,i})$ for other ages z. The results align with the notions introduced in Chapter 3. The Bayesian estimates are derived by calculating a weighted average over all corresponding slope values of $\hat{\theta}^{RW-*}_{z,i}$, which is explicitly represented as $\hat{\theta}^{BS-*}_{x,\,i} = \sum_{z=25}^{84} \pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_{x,\,i}) \cdot \hat{\theta}^{RW-*\ 2}_{z,\,i}$ for age x and the

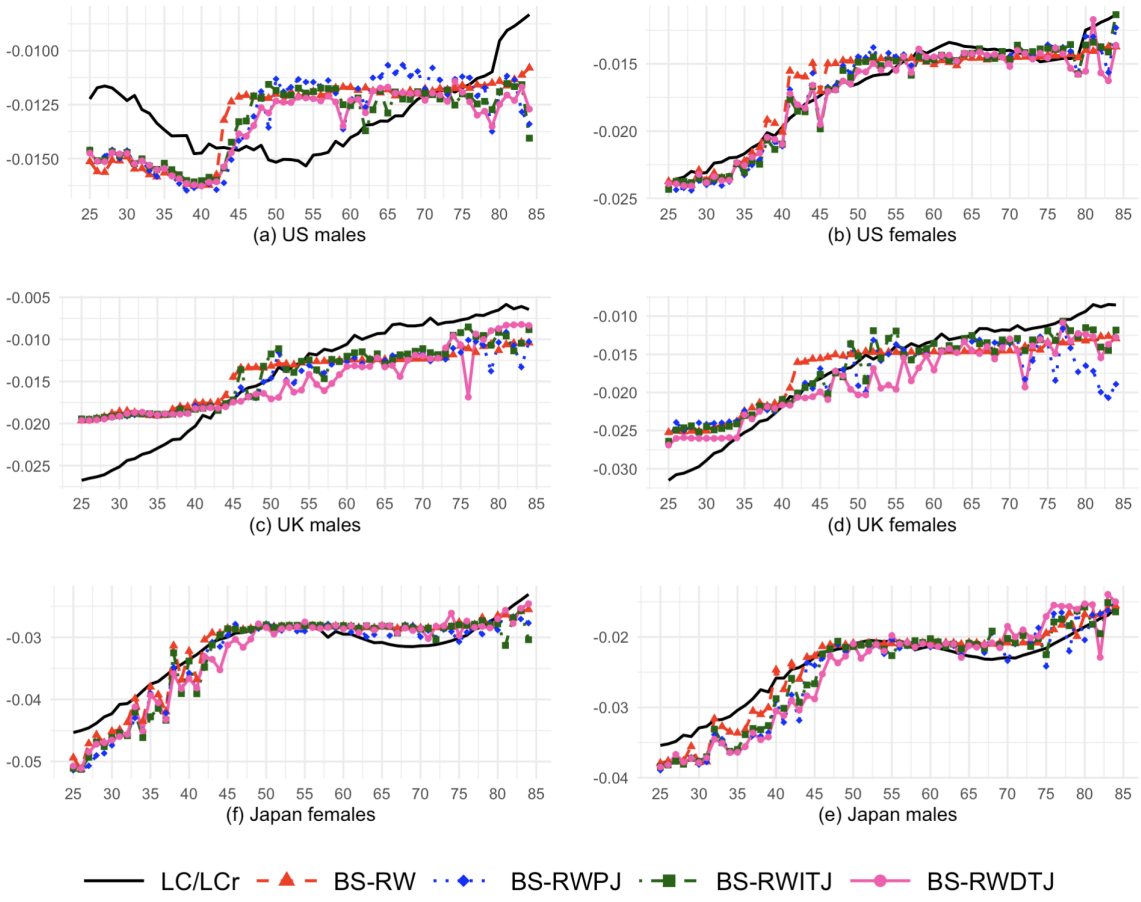

Figure 4.1: Slope against age x for the US, the UK, and Japan

study population i

To assess and compare the predictive performance of different models, the Mean Absolute Percentage Error (MAPE), a commonly used statistical metric, is employed. The forecasting error is measured by quantifying the percentage difference between the observed one-year death probabilities (q's) and the corresponding projected one-year death probabilities ($\hat{q}$'s). To be more specific, considering an age-year window $[x_L, x_U] \times [t_U + 1, T_2]$ for population i, with a fitting age-year window $[x_L, x_U] \times [t_L, t_U]$, the MAPE is defined as follows:

$$\text{MAPE}^{[x_L, x_U] \times [t_L, t_U]}_{[t_U+1, T_2], i} = \frac{1}{(T_2 - t_U) \times (x_U - x_L + 1)} \sum_{t=1}^{T_2 - t_U} \sum_{x=x_L}^{x_U} \left| \frac{\hat{q}_{x, t_U+t, i} - q_{x, t_U+t, i}}{q_{x, t_U+t, i}} \right|.$$

[2]The asterisk (*) is used to collectively represent the four models throughout the text, streamlining references to RW, RWPJ, RWITJ, and RWDTJ.

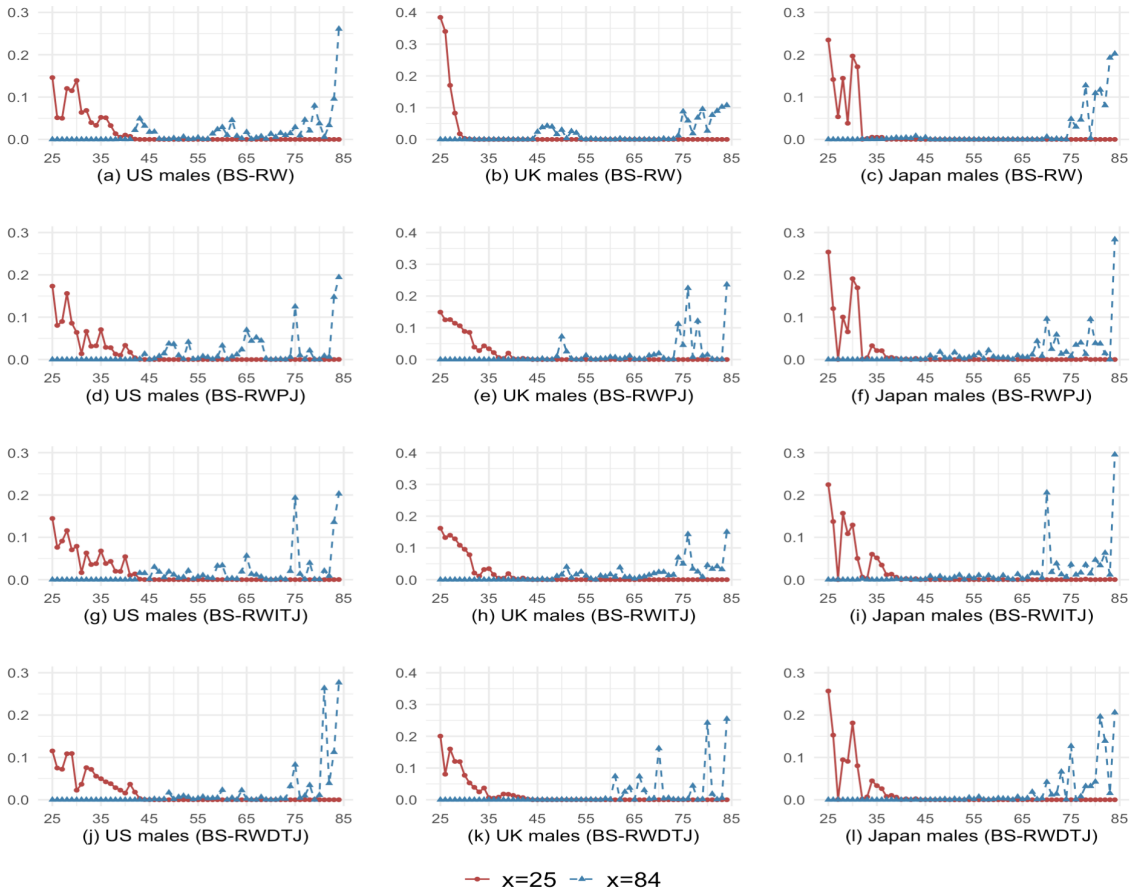

Figure 4.2: $\pi_{\Lambda|\tilde{y}}(\lambda_z|\tilde{y}_{x,\,i})$, $x = 25, 84$, against age z, for males of the US, the UK, and Japan

This formula quantifies the average percentage difference across the specified age-year window, providing a comprehensive evaluation of forecasting performance.

To comprehensively assess forecasting performance, the average of $\text{MAPE}^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[t_U+1,\,T_2],\,i}$ across a series of fitting age-year windows $[x_L,\,x_U]\times[t_L,\,t_U]$, $t_L = T_1, T_1+1, \ldots, t_U - 22$, is calculated. Initially, we employ the training data time interval defined in Table 4.1 to estimate parameters, predict mortality rates, and calculate the $\text{MAPE}^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[t_U+1,\,T_2],\,i}$ on the test set. At each iteration, the first column of data points for the smallest fitting year are discarded, and the process of estimating parameters, predicting mortality rats, and calculating the $\text{MAPE}^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[t_U+1,\,T_2],\,i}$ on the test set is repeated. This iterative process continues until only 23 columns of data points for 23 fitting years remain for modeling. The retention of these data points is crucial to ensuring the quality of maximum likelihood estimation. Finally, we calculate the average of the $\text{MAPE}^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[t_U+1,\,T_2],\,i}$ values to obtain the

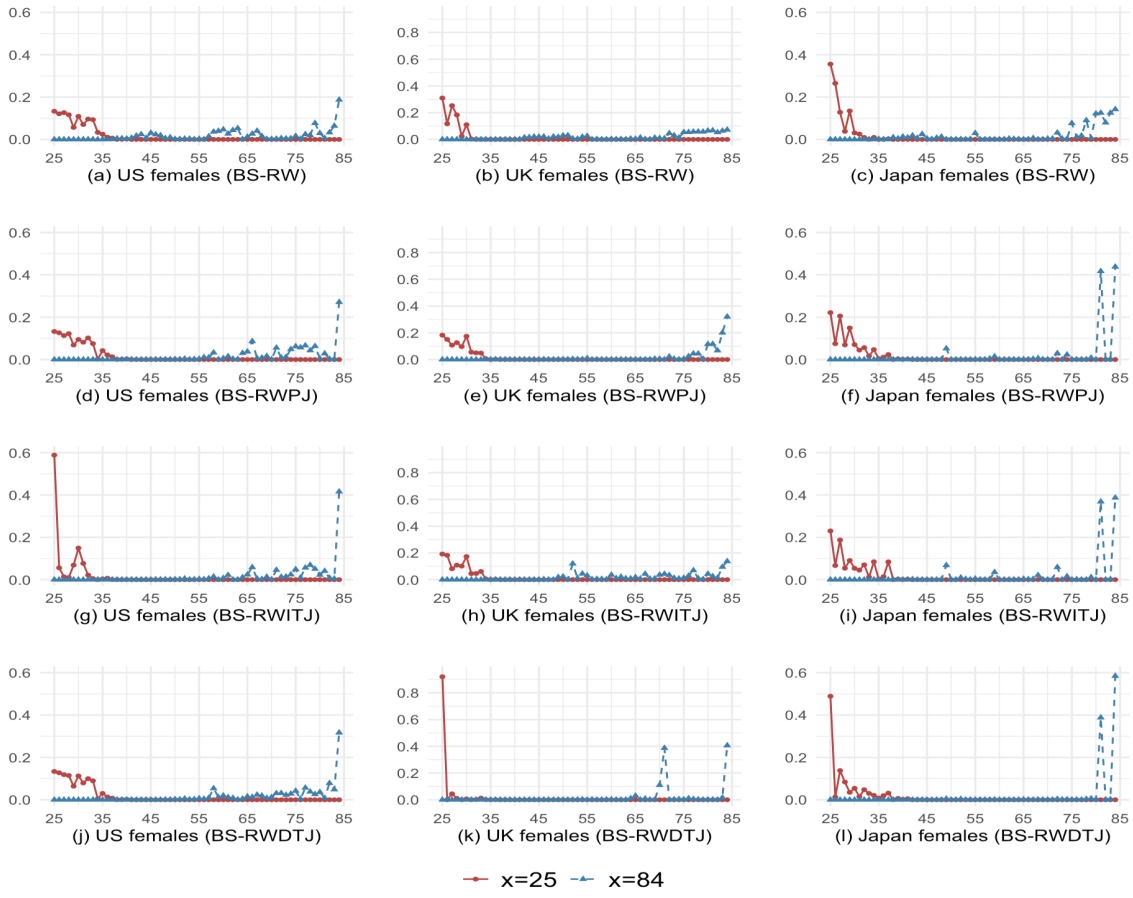

Figure 4.3: $\pi_{\Lambda|\tilde{y}}(\lambda_z|\tilde{y}_{x,i})$, $x = 25, 84$, against age z, for females of the US, the UK, and Japan

$\text{AMAPE}^{[x_L, x_U]}_{[t_U+1, T_2], i}$, with its specific mathematical definition provided below.

$$\text{AMAPE}^{[x_L, x_U]}_{[t_U+1, T_2], i} = \frac{1}{(t_U - 22 - T_1 + 1)} \sum_{t_L=T_1}^{t_U-22} \text{MAPE}^{[x_L, x_U]\times[t_L, t_U]}_{[t_U+1, T_2], i}.$$

Table 4.2 lists the values of Avg-6, Avg-2[3], and $\text{AMAPE}^{[25, 84]}_{[t_U+1, 2020], i}$ for single population i ($i = 1, 2, \ldots, 6$) across three distinct forecasting periods. The smallest values in each column are highlighted in bold. Typically, a longer forecasting period corresponds to a larger $\text{AMAPE}^{[25, 84]}_{[t_U+1, 2020], i}$ value. To conduct a thorough comparative analysis, the classic Lee-Carter (LC) model and its adjusted version, denoted as LCr, are also included for comparisons. The LCr model is designed to avoid jump-off bias by adjusting the intercept from the fitted value $\ln(\hat{m}^{LC}_{x, t_U, i})$ to the observed value $\ln(m_{x, t_U, i})$. This adjustment not

[3]Avg-6 denotes the average AMAPE across six populations, including all combinations of gender and country, whereas Avg-2 represents the average AMAPE across both genders of each country.

Model	Avg-6	Avg-2	US-M	US-F	Avg-2	UK-M	UK-F	Avg-2	Japan-M	Japan-F
Panel A: $t_U = 2010$ for forecasting year span [2011, 2020]										
LC	12.42	12.21	12.78	11.63	17.03	19.97	14.08	8.01	8.21	7.82
LCr	8.14	10.18	10.87	9.48	7.02	6.88	7.16	7.21	7.20	7.22
BS-RW	7.87	10.04	10.70	**9.38**	6.72	6.39	7.05	6.85	6.89	6.82
BS-RWPJ	7.51	10.04	10.39	9.69	5.85	**5.62**	6.07	6.65	6.70	6.61
BS-RWITJ	**7.36**	**9.76**	**10.13**	9.38	**5.77**	5.70	**5.85**	**6.56**	6.65	**6.46**
BS-RWDTJ	7.46	9.94	10.30	9.58	5.80	5.66	5.95	6.65	**6.58**	6.72
Panel B: $t_U = 2000$ for forecasting year span [2001, 2020]										
LC	16.36	12.02	12.47	11.58	24.91	30.86	18.96	12.14	9.21	15.08
LCr	11.22	11.23	11.23	11.23	12.99	15.57	10.41	9.44	6.55	12.33
BS-RW	10.29	10.63	10.75	10.51	11.19	13.49	**8.88**	9.04	5.82	12.26
BS-RWPJ	**9.41**	**9.97**	**10.45**	**9.50**	**11.02**	13.11	8.92	**7.23**	5.57	**8.89**
BS-RWITJ	9.91	10.46	10.65	10.28	11.05	12.93	9.17	8.20	**5.57**	10.84
BS-RWDTJ	9.79	10.40	10.60	10.19	11.03	**12.83**	9.24	7.94	5.59	10.29
Panel C: $t_U = 1990$ for forecasting year span [1991, 2020]										
LC	22.34	14.86	15.43	14.29	30.74	40.07	21.41	21.42	17.07	25.76
LCr	18.57	13.82	13.56	14.08	21.68	26.27	17.09	20.22	16.46	23.99
BS-RW	17.15	13.21	12.97	13.45	18.49	22.46	14.53	19.76	16.25	23.27
BS-RWPJ	**14.68**	10.18	**10.62**	9.74	**14.45**	**17.86**	**11.05**	19.41	15.50	23.31
BS-RWITJ	15.22	10.03	11.04	**9.02**	16.70	20.33	13.08	**18.93**	**15.04**	**22.83**
BS-RWDTJ	14.93	**9.98**	10.84	9.12	15.69	18.84	12.53	19.12	15.20	23.04

Table 4.2: $\text{AMAPE}^{[25,84]}_{[t_U+1,\,2020],\,i}(\%)$ for $I = 1$

only results in significantly improved forecasting performance in most cases compared to the classic LC model but also facilitates a specific examination of the impact of the slope on forecasting mortality rates. Among the models LCr, BS-RW, BS-RWPJ, BS-RWITJ, and BS-RWDTJ, all sharing the same intercept ($\ln(m_{x,\,t_U,\,i})$), the BS-RWPJ model exhibits superior performance in long-term (20 and 30 years) forecasting, achieving the lowest Avg-6, while the BS-RWITJ model generally demonstrates accurate forecasting for a short-term period (10 years) by providing the smallest Avg-6. These results highlight that RWPJ and RWITJ provide more precise slopes for mortality forecasting compared to other models in a single-population test. Diving into the details, there are exceptions to this general trend. For instance, in the 30-year forecast period for Japan, the BS-RWITJ model slightly outperforms the BS-RWPJ model, and in the 10-year forecast period for Japan males, the BS-RWDTJ model exhibits a slight superiority over the BS-RWITJ model. In almost all scenarios, models incorporating jumps generally yield lower $\text{AMAPE}^{[25,84]}_{[t_U+1,\,2020],\,i}$ compared to the LC, LCr, and BS-RW (models without considering jumps) models, emphasizing the importance of incorporating jump components into a model.

4.2 Comparisons of models for multiple populations

This subsection employs the hierarchical Bayesian theory to comprehensively compare the forecasting performance for multiple populations. Ten models are evaluated, including the random walk with drift model and its three extended models incorporating jump components, and six Lee-Carter-based models (JK, CI, and ACF, along with their corresponding JKr, CIr, and ACFr; refer to Appendices A.1.1-A.1.3). Table 4.3 presents a test with $I = 2$, considering both genders of a country for each of the US, the UK, and Japan, while Table 4.4 conducts a test with $I = 6$, encompassing all six populations. Similar to the single population test, both Tables 4.3 and 4.4 display the values of Avg-6, Avg-2, and $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ for three different forecasting periods, with the smallest value being highlighted in bold in each column.

Similar to the classical Lee-Carter model, adjusting the intercept from the fitted value $\ln(\hat{m}^{LC}_{x,t_U,i})$ to the observed value $\ln(m_{x,t_U,i})$ for JKr, CIr, and ACFr models significantly reduces the $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ in most cases. This ensures that all models forecast mortality rates starting from the same baseline (intercept), allowing us to focus solely on which model produces more accurate slopes to better capture the downward mortality trend compared to the others. Among the ten models, either the HBS-RWDTJ or HBS-RWITJ model exhibits the superior performance in terms of Avg-6 in Table 4.3. To elaborate, the HBS-RWDTJ model achieves the lowest Avg-2, and Avg-6, and $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ for the 10-year and 30-year forecast period with few exceptions. The HBS-RWPJ and HBS-RWITJ models also exhibit promising results, delivering overall lower Avg-2 and Avg-6 values compared to other models without jumps. Generally, the models incorporating jump components yield lower $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values, except for the 10-year and 20-year forecasts for US females, as well as the 20-year forecasts for UK females and Japan males.

In Table 4.4 for $I = 6$, the HBS-RWDTJ model displays the lowest Avg-2, Avg-6, and $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values, with a few exceptions, in the 10-year and 30-year forecasting periods. In the 20-year forecasting, the HBS-RWDTJ model also generally outperforms other models in terms of the Avg-6 and Avg-2 values, but $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ favors the HBS-RWPJ or HBS-RWITJ model for some cases. Overall, the models with jump components tend to dominate this test with a few exceptions, including 10-year forecasting for US females and 20-year forecasting for UK females, as well as 30-year forecasting for both genders of Japan. Summarizing the results from Tables 4.3 and 4.4, the models with jump components exhibit the lowest Avg-6, Avg-2, and $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values in most cases. However, in certain cases for female populations, the models without jump components may surpass those with jump components.

Model	Avg-6	Avg-2	US-M	US-F	Avg-2	UK-M	UK-F	Avg-2	Japan-M	Japan-F
Panel A: $t_U = 2010$ for forecasting year span [2011, 2020]										
JK	12.93	12.68	12.54	12.83	17.31	20.76	13.87	8.80	9.01	8.58
JKr	8.17	10.19	11.03	9.34	7.04	6.86	7.23	7.29	7.28	7.30
CI	12.40	12.09	12.78	11.39	17.03	19.97	14.09	8.10	8.21	7.98
CIr	8.11	9.96	10.87	**9.05**	7.08	6.88	7.28	7.29	7.20	7.38
ACF	12.43	11.91	12.94	10.88	17.16	19.91	14.42	8.21	8.72	7.70
ACFr	8.14	10.17	10.83	9.52	7.03	6.79	7.28	7.22	7.27	7.16
HBS-RW	7.71	10.03	10.55	9.50	6.66	6.37	6.95	6.44	6.47	6.40
HBS-RWPJ	7.43	10.25	10.49	10.01	5.97	5.71	6.22	6.06	6.11	6.01
HBS-RWITJ	7.35	10.11	10.35	9.87	5.93	5.72	6.14	**6.00**	6.02	**5.97**
HBS-RWDTJ	**7.29**	**9.89**	**10.26**	9.52	**5.80**	**5.67**	**5.93**	6.19	**6.02**	6.37
Panel B: $t_U = 2000$ for forecasting year span [2001, 2020]										
JK	16.87	13.63	12.96	14.30	25.10	31.71	18.48	11.90	9.14	14.65
JKr	11.21	11.27	11.13	11.42	12.99	15.59	10.40	9.37	6.61	12.13
CI	16.33	12.11	12.47	11.76	24.89	30.86	18.93	11.99	9.21	14.76
CIr	11.17	11.31	11.23	11.40	13.02	15.57	10.48	9.18	6.55	11.81
ACF	16.41	11.89	13.04	10.73	25.33	31.43	19.23	12.00	9.04	14.96
ACFr	11.16	11.06	11.35	10.77	12.97	15.29	10.65	9.44	6.53	12.34
HBS-RW	10.02	10.63	10.92	**10.34**	10.91	13.07	**8.76**	8.51	**6.42**	10.59
HBS-RWPJ	10.05	**10.60**	**10.63**	10.56	10.89	12.90	8.88	8.66	6.97	10.36
HBS-RWITJ	**10.00**	10.81	10.91	10.71	**10.87**	**12.77**	8.97	**8.33**	6.57	**10.08**
HBS-RWDTJ	10.25	10.92	11.39	10.46	10.94	12.89	8.99	8.90	6.92	10.87
Panel C: $t_U = 1990$ for forecasting year span [1991, 2020]										
JK	22.51	15.40	15.86	14.93	30.71	40.16	21.27	21.43	17.28	25.58
JKr	18.53	13.72	13.49	13.94	21.63	26.16	17.10	20.25	16.68	23.82
CI	22.17	14.51	15.43	13.60	30.76	40.07	21.46	21.24	17.07	25.40
CIr	18.39	13.46	13.56	13.36	21.71	26.27	17.15	20.01	16.46	23.57
ACF	22.41	14.84	15.81	13.88	30.76	38.72	22.80	21.64	17.32	25.96
ACFr	18.60	13.79	13.65	13.94	21.71	25.87	17.54	20.30	16.55	24.05
HBS-RW	16.69	11.93	11.58	12.28	18.14	21.77	14.50	20.01	18.40	21.62
HBS-RWPJ	13.69	10.67	11.06	10.27	17.28	21.12	13.44	13.13	**9.55**	16.70
HBS-RWITJ	13.97	10.71	11.86	9.55	18.00	21.66	14.33	13.21	11.45	14.96
HBS-RWDTJ	**12.75**	**10.00**	**10.90**	**9.09**	**15.85**	**19.59**	**12.10**	**12.42**	10.74	**14.10**

Table 4.3: $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}(\%)$ for $I = 2$

4.3 Three grouping analyses of models for multiple populations worldwide

In the previous subsection, mortality data from the US, the UK, and Japan are used as representatives of America, Europe, and Asia, respectively. In this section, we extend the application of the hierarchical Bayesian models with jump components to include mortality data from an additional 14 countries, resulting in a total of 17 countries. These countries include Canada and the US for North America; Australia and Japan for Oceania-Asia; Sweden, Finland and Norway for North Europe; Belgium, Ireland, Netherlands and the UK for West Europe; France, Spain and Switzerland for South Europe; and Bulgaria, Czechia and Hungary for East Europe. The selection of these countries is constrained by the complete-

Model	Avg-6	Avg-2	US-M	US-F	Avg-2	UK-M	UK-F	Avg-2	Japan-M	Japan-F
Panel A: $t_U = 2010$ for forecasting year span [2011, 2020]										
JK	10.94	11.52	12.07	10.98	10.67	12.59	8.75	10.61	7.89	13.33
JKr	8.14	10.06	11.25	8.88	6.99	7.07	6.92	7.37	7.31	7.43
CI	9.72	11.63	12.55	10.71	9.42	10.94	7.90	8.11	8.82	7.40
CIr	7.89	**9.77**	10.99	**8.55**	6.63	6.70	6.56	7.27	7.88	6.65
ACF	9.72	11.20	12.13	10.28	9.52	11.91	7.12	8.45	8.56	8.34
ACFr	7.98	9.98	10.95	9.02	6.77	6.61	6.93	7.18	7.34	7.02
HBS-RW	7.96	10.53	11.15	9.90	6.64	6.39	6.88	6.72	7.18	6.27
HBS-RWPJ	8.00	10.42	10.83	10.01	7.26	6.99	7.52	**6.33**	**6.59**	6.07
HBS-RWITJ	8.02	10.28	10.66	9.89	7.35	7.05	7.66	6.43	6.68	6.18
HBS-RWDTJ	**7.70**	10.20	**10.62**	9.77	**5.96**	**5.68**	**6.24**	6.94	8.48	**5.40**
Panel B: $t_U = 2000$ for forecasting year span [2001, 2020]										
JK	14.47	12.54	12.99	12.09	16.92	20.14	13.71	13.95	10.24	17.66
JKr	10.74	10.86	10.80	10.91	11.70	14.34	9.06	9.67	6.70	12.63
CI	12.99	11.67	12.08	11.26	15.18	17.37	12.99	12.12	9.46	14.78
CIr	10.65	11.06	11.16	10.96	11.76	14.37	9.14	9.13	6.69	11.58
ACF	12.60	11.00	11.00	11.01	15.41	19.58	11.23	11.37	7.19	15.56
ACFr	10.30	11.03	11.11	10.94	10.77	13.13	8.41	9.11	5.91	12.31
HBS-RW	9.44	10.96	11.20	10.72	9.73	11.82	**7.64**	7.63	6.00	9.26
BS-RWPJ	9.55	**10.65**	10.62	**10.68**	9.70	11.05	8.34	8.32	6.00	10.63
HBS-RWITJ	9.60	10.67	**10.61**	10.72	9.61	**10.78**	8.43	8.52	6.20	10.84
HBS-RWDTJ	**9.30**	11.05	11.19	10.90	**9.51**	10.88	8.15	**7.35**	**5.90**	**8.80**
Panel C: $t_U = 1990$ for forecasting year span [1991, 2020]										
JK	19.59	14.75	14.21	15.30	22.44	26.68	18.19	21.57	17.53	25.61
JKr	17.02	13.68	13.68	13.69	17.69	20.09	15.29	19.68	16.19	23.18
CI	17.55	13.24	14.49	11.99	22.36	26.32	18.41	17.04	13.48	20.60
CIr	15.51	12.72	13.82	11.62	18.97	21.94	15.99	**14.83**	**11.99**	**17.67**
ACF	18.61	13.83	15.11	12.55	20.60	26.82	14.38	21.41	16.92	25.89
ACFr	17.02	13.81	15.11	12.52	17.00	20.30	13.70	20.23	16.43	24.03
HBS-RW	15.39	11.69	11.29	12.09	16.47	17.26	15.67	18.00	14.55	21.45
HBS-RWPJ	15.44	11.75	10.59	12.91	16.37	**17.14**	15.60	18.20	14.95	21.46
HBS-RWITJ	15.47	11.36	10.68	12.03	16.55	17.18	15.93	18.50	14.93	22.08
HBS-RWDTJ	**14.10**	**10.64**	**10.29**	**10.99**	**15.46**	17.79	**13.12**	16.20	12.71	19.70

Table 4.4: $\text{AMAPE}^{[25,84]}_{[t_U+1,\,2020],\,i}(\%)$ for $I = 6$

ness of mortality data in the Human Mortality Database. These countries provide the most complete and continuous data throughout the study window $[25, 84] \times [1950, 2020]$. The 17 selected countries are classified into distinct groups based on specific criteria. We assess the forecasting performance of the hierarchical Bayesian jump models through three grouping analyses. Following that, the models are utilized to analyze trends in the average expected yearly mortality changes worldwide.

For the one-group analysis, we adopt an approach similar to the one detailed in Table 4.4. This analysis involves mortality data for both genders across 17 countries ($I = 34$). The forecasting performances of the hierarchical Bayesian models, with and without jump components, are evaluated in comparison to the underlying Lee-Carter-Based models. To

facilitate a comprehensive comparison, the averages of the $\text{AMAPE}^{[25,\,84]}_{[t_U+1,\,2020],\,i}$ values across 34 populations ($i = 1, 2, \ldots, 34$) are first calculated and then presented in Panel A of Table 4.5, where the HBS-RWITJ model outperforms all other models in terms of the average $\text{AMAPE}^{[25,\,84]}_{[t_U+1,\,2020],\,i}$ across forecasting intervals of 10, 20, and 30 periods.

For the two-group analysis, the selected 34 populations are divided into two groups based on their development status: developing and developed. Developing countries comprised Bulgaria, Czechia, and Hungary, while developed countries included Canada, the US, Australia, Japan, Belgium, Switzerland, Spain, Finland, France, Ireland, Netherlands, Norway, Sweden, and the UK. Panel B (B_1 and B_2) of Table 4.5 summarizes the averages of $\text{AMAPE}^{[25,\,84]}_{[t_U+1,\,2020],\,i}$ values for forecasting across both groups and three forecasting periods. In Panel B_1 for the developed countries, the HBS-RWITJ model outperform all other models. In Panel B_2 for the developing countries, the HBS-RWITJ model shows a slightly higher average of $\text{AMAPE}^{[25,\,84]}_{[t_U+1,\,2020],\,i}$ compared to the HBS-RWDTJ model in the 20-year forecast period; however, the performance gap is minimal, emphasizing that the HBS-RWITJ model remains competitive overall.

For the three-group analysis, the 34 countries are grouped, based on geographic regions, into three categories: the North America group, including Canada and the US; the Oceania-Asia group, comprising Australia and Japan; and the Europe group, consisting of the remaining countries. Panel C (C_1 to C_3) presents the respective averages of $\text{AMAPE}^{[25,\,84]}[t_U + 1,\, 2020],\, i$ for each group. Analyzing these values, the HBS-RWITJ model also generally outperforms all other models. However, in the North America group, its averages of $\text{AMAPE}^{[25,\,84]}[t_U + 1,\, 2020],\, i$ for the 20-year and 30-year forecasts is slightly higher than the other two jump models, but the difference is minimal. It is noteworthy that, in the 30-year forecast for the Oceania-Asia group, the model without a jump component (the HBS-RW model) achieves the smallest average of $\text{AMAPE}^{[25,\,84]}_{[t_U+1,\,2020],\,i}$.

As observed in the preceding test, all the hierarchical Bayesian models demonstrate superior performance compared to the Lee-Carter-based models. Furthermore, the models incorporating jump components generally outperform those without jump components. The subsequent analysis focuses on evaluating the patterns in the averages of the expected yearly mortality changes across 34 populations by refitting the hierarchical Bayesian models, both with and without jump components, and using the mortality data across the same three grouping analyses. More precisely, we re-estimate $\theta^{HBS-*}_{(x,\,i,\,t_U)}$ by using a shifting window of $[25,\, 84] \times [t_U - 30,\, t_U]$, rolling it forward by one year at a time from $t_U = 1980$ to $t_U = 2020$. Throughout this iterative process, we repeatedly conduct model fitting, yielding a series of $\hat{\theta}^{HBS-*}_{x,\,i,\,t_U}$ values for each model. These values represent the expected yearly changes in the logarithm of the central death rate beyond each year t_U for age x and population

Model	HBS-RW	HBS-RWPJ	HBS-RWITJ	HBS-RWDTJ	JK	JKr	CI	CIr	ACF	ACFr
t_U	Panel A: 34 populations under one-group method									
2010	11.61	11.87	**11.51**	11.76	16.03	12.38	13.02	12.31	13.62	12.35
2000	17.41	18.45	**16.68**	19.82	25.37	20.76	21.16	20.59	21.88	20.28
1990	23.35	25.78	**21.17**	23.49	29.40	28.95	29.25	29.49	28.62	28.79
t_U	Panel B_1: Developed-country group under two-group method									
2010	11.46	11.73	**11.37**	11.65	14.90	12.20	12.78	12.14	13.09	12.17
2000	16.30	17.44	**15.86**	19.62	22.01	18.99	19.72	18.92	20.17	18.82
1990	19.84	22.91	**19.09**	21.55	25.05	24.29	25.17	25.27	24.37	24.46
t_U	Panel B_2: Developing-country group under two-group method									
2010	12.63	12.80	**12.58**	12.70	14.24	12.78	15.51	14.57	15.39	13.04
2000	25.14	24.14	22.78	**22.52**	26.25	24.76	30.54	30.89	26.62	26.63
1990	45.85	39.79	**35.31**	37.50	47.05	46.55	49.89	50.89	48.25	49.13
t_U	Panel C_1: North American group under three-group method									
2010	9.75	9.54	**9.45**	9.67	11.65	9.86	11.27	9.61	11.35	9.83
2000	10.55	10.68	10.40	**10.39**	12.76	11.13	11.82	11.16	11.51	11.13
1990	11.19	**11.16**	11.20	11.33	14.54	13.21	14.03	13.46	14.05	13.02
t_U	Panel C_2: Oceania-Asia group under three-group method									
2010	7.34	7.67	**7.19**	7.74	10.80	8.40	9.96	8.39	9.47	8.16
2000	10.04	11.76	**9.78**	11.27	15.43	12.35	14.22	12.25	13.64	11.86
1990	**16.30**	18.47	16.80	17.84	21.21	19.00	19.93	19.19	20.23	18.70
t_U	Panel C_3: Europe group under three-group method									
2010	12.49	12.82	**12.42**	12.85	16.20	13.25	13.71	13.23	14.42	13.28
2000	20.06	20.98	**19.05**	22.31	27.52	23.05	23.53	23.19	24.55	22.72
1990	27.18	28.69	**24.25**	26.78	32.55	32.46	31.61	32.15	31.95	32.37

Table 4.5: Averages of $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}s$ (%) for three grouping methods.

i, under each model. At each time step, the average of $\hat{\theta}^{HBS-*}_{x,i,t_U}$ is calculated across ages x from 25 to 84, which can be expressed mathematically as $\bar{\theta}^{HBS-*}_{i,t_U} = (1/60)\sum_{x=25}^{84}\hat{\theta}^{HBS-*}_{x,i,t_U}$ for population i within the fitting year span $[t_U - 30, t_U]$. Besides, we calculate $\bar{\theta}^{HBS-*}_{i,t_U}$ for each time step from year 1980 to 2020 for every model and illustrate the results from Figures 4.4 to 4.21.

From Figures 4.4 to 4.21, we observe that the $\bar{\theta}^{HBS-*}_{i,t_U}$ values tend to be negative for all populations throughout the considered periods, with only a few exceptions. The first exception is in Figure 4.9; the HBS-RW model exhibits three positive $\bar{\theta}^{HBS-*}_{i,t_U}$ values in Hungary males around year 1995, while the HBS-RWPJ, HBS-RWITJ, and HBS-RWDTJ model show negative values, though very close to zero. This suggests that the models with jump components expect a slightly higher mortality improvement compared to models without jump components. Another exception is in the developing countries of two-group analysis (see Figure 4.15). Interestingly, for both genders from approximately year 1985 to 2000, all four models display positive $\bar{\theta}^{HBS-*}_{i,t_U}$ values. Specifically, the hierarchical Bayesian method incorporates a mechanism to adjust the impact of study ages within a group of populations on a specific age and population. This adjustment accounts for co-movements across ages

and populations.

As illustrated in Figure 4.9, Hungary male population shows subtle positive $\overline{\theta}^{HBS-*}_{i,t_U}$ values. It seems that Hungary male population has a relatively worse improvement in mortality rates compared to other populations within the group, where the overall trend indicates better mortality improvement. As a result, the subtle positive value in this context is somewhat diluted by the prevailing trend of superior mortality improvement in the group. On the contrary, observing from Figure 4.15 which includes only six populations, the absence of other populations dilutes the impact results in Hungary populations displaying larger $\overline{\theta}^{HBS-*}_{i\,,t_U}$ values compared to Figure 4.9. Furthermore, other populations in the developing country group, including both genders of Bulgaria and Czechia, appear to be affected by Hungary populations, leading to positive $\overline{\theta}^{HBS-*}_{i,\,t_U}$ values.

Moreover, a consistent trend of convergence towards -0.02 is observed in $\bar{\theta}^{HBS-*}_{i,\,t_U}$ for all four models, regardless of the grouping approach used. This trend holds across various analyses, except for instances involving both genders of the US population as seen in Figures 4.4, 4.10 and 4.16. Interestingly, both one-group and two-group analyses reveal that, regardless of the model type, the average mortality improvement for the US populations tends to worsen, deviating by approximately -0.02 to -0.01. However, the three-group analysis, which only includes both genders of the US and Canada ($I = 4$) within the North America group, highlights a stronger influence of the US populations on Canada populations. Consequently, the final $\overline{\theta}^{HBS-*}_{i,\,t_U}$ for Canada populations deviates from the overall trend, exhibiting a decrease of -0.02. From an economic perspective, Becker et al. (2005) argued that GDP is commonly used as a proxy to assess the quality of life for individuals in different countries, with welfare being influenced by life expectancy. This suggests that life expectancy can be considered a strong indicator of the quality of life. Additionally, they explained that rapid improvement in mortality rates in developing countries is fueled by mortality improvements in developed countries driven by advancements on the frontier of medical technology. Developing countries can adopt technology and knowledge at relatively lower costs, contributing to their own mortality improvements. The results of our two-group analysis in this study consistently align with this phenomenon. Over the same time span (e.g., 1995 to 2020), the trends in mortality rate improvement in developing countries exhibit a steeper trajectory compared to that in developed countries, which indicates a more pronounced improvement in mortality rates during this specific time period. Regarding the ongoing mortality improvement and convergence to -0.02, our findings align with the emphasis by Lin and Tsai (2022) that mortality rates worldwide have been improving over the past decades and are expected to continue doing so, but they converge to a constant level irrespective of the efforts humans put into medical breakthroughs, living environment improvement, and so on.

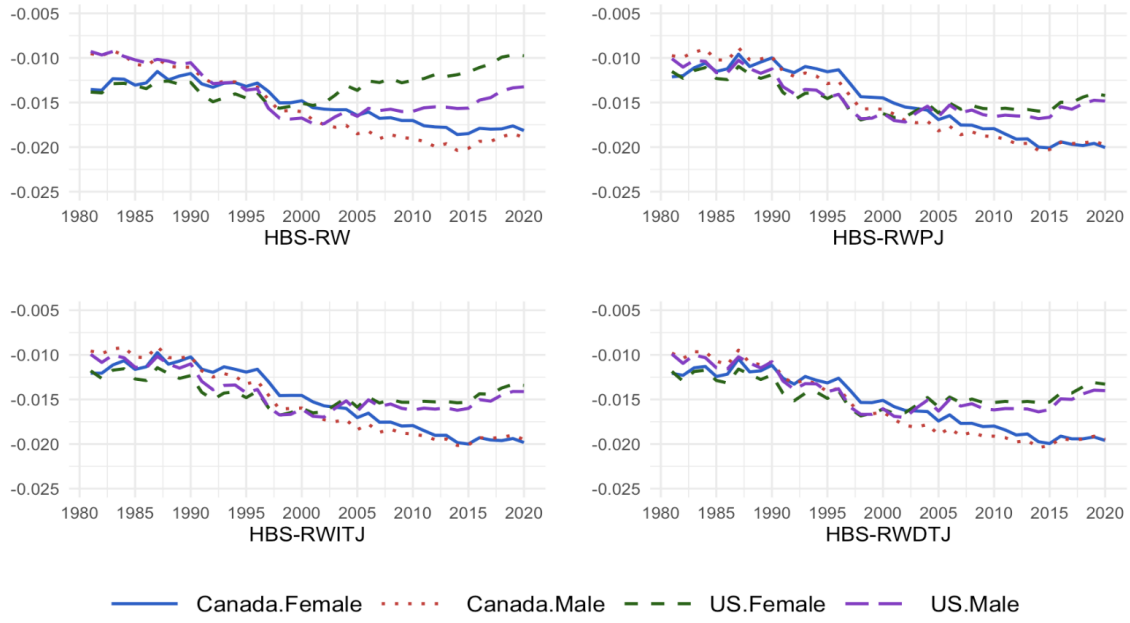

Figure 4.4: $\overline{\theta}^{HBS}_{i,t_U}$ for the North American group under the one-group method

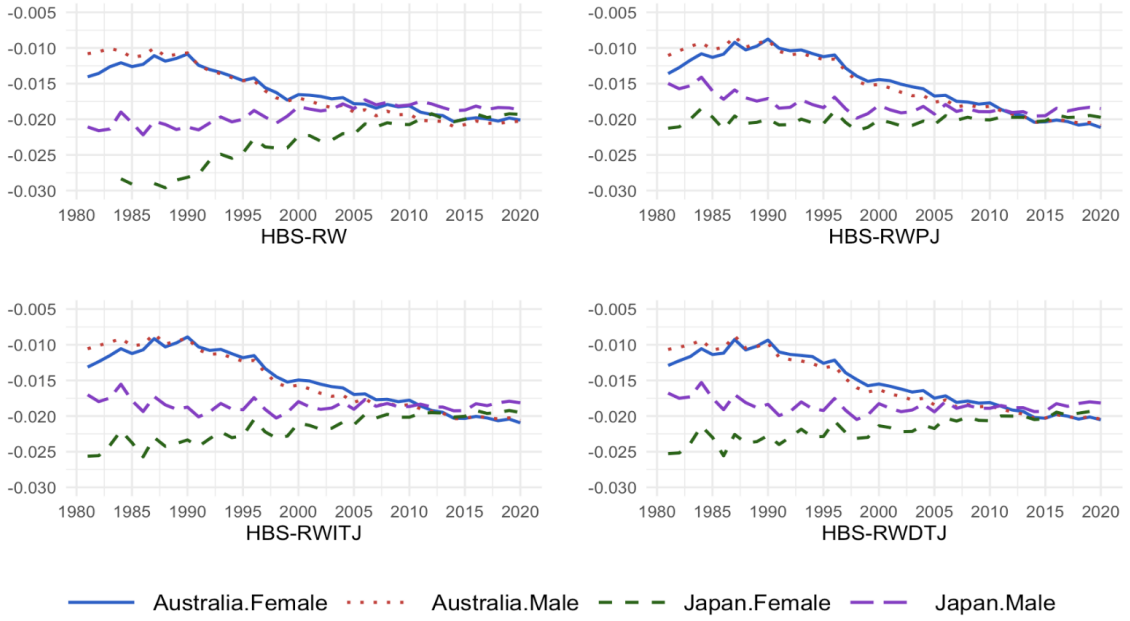

Figure 4.5: $\overline{\theta}^{HBS}_{i,t_U}$ for the Oceania-Asia group under the one-group method

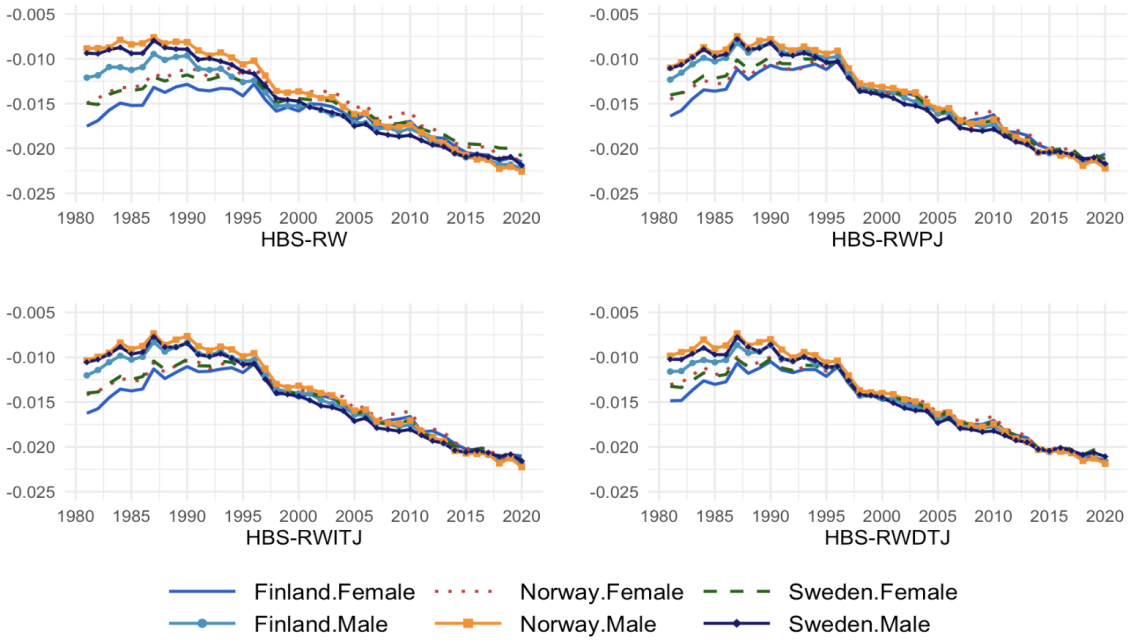

Figure 4.6: $\overline{\theta}^{HBS}_{i,t_U}$ for the North Europe group under the one-group method

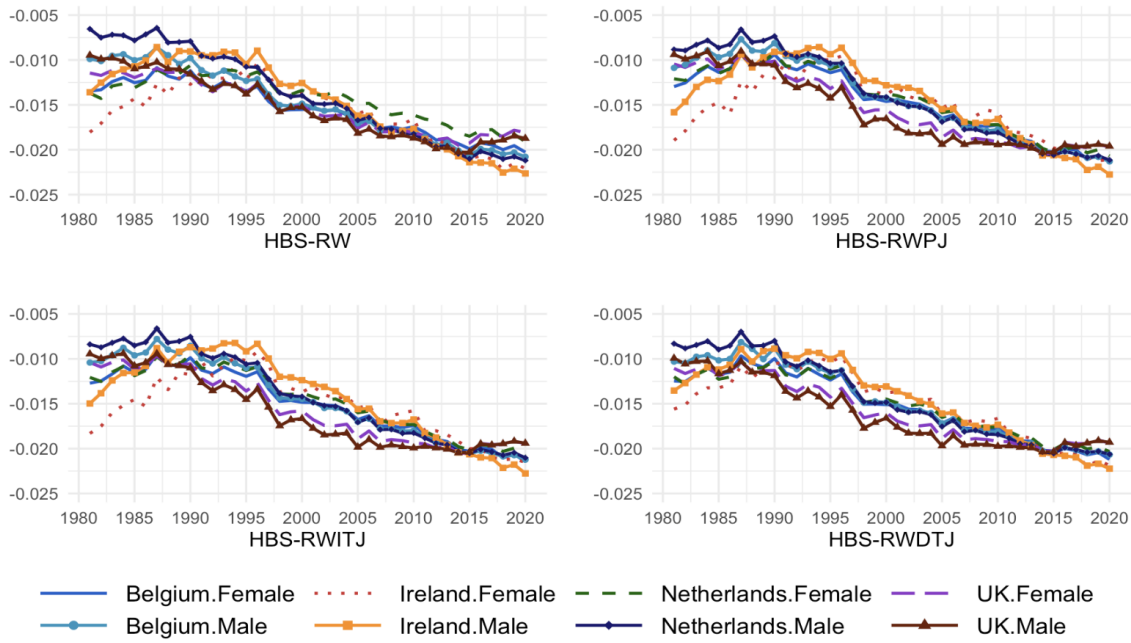

Figure 4.7: $\overline{\theta}^{HBS}_{i,t_U}$ for the West Europe group under the one-group method

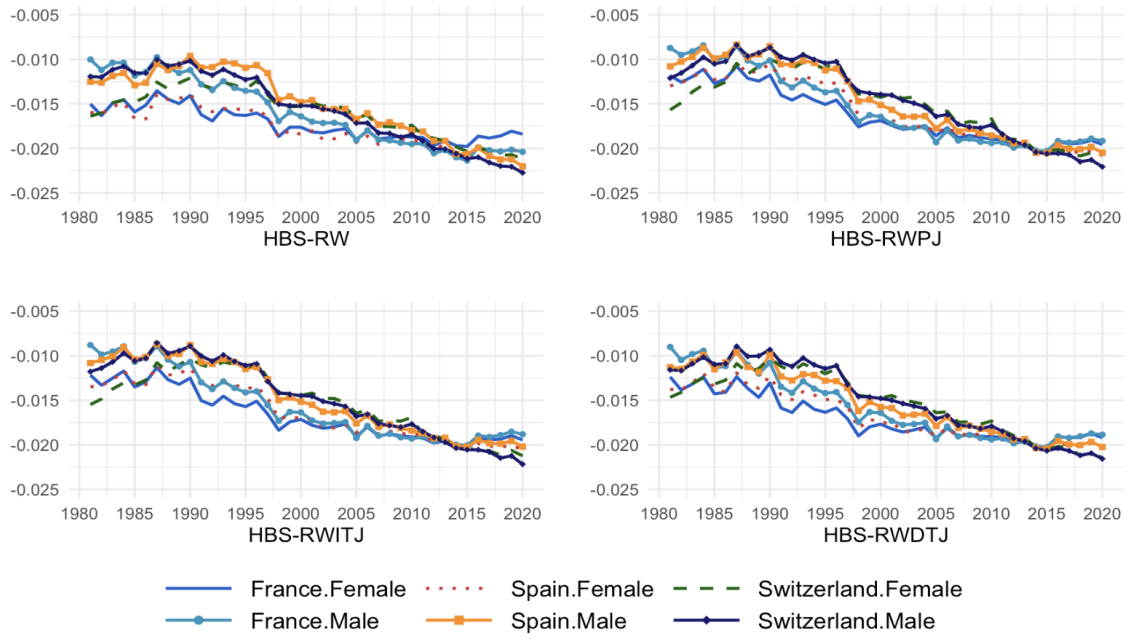

Figure 4.8: $\overline{\theta}^{HBS}_{i,t_U}$ for the South Europe group under the one-group method

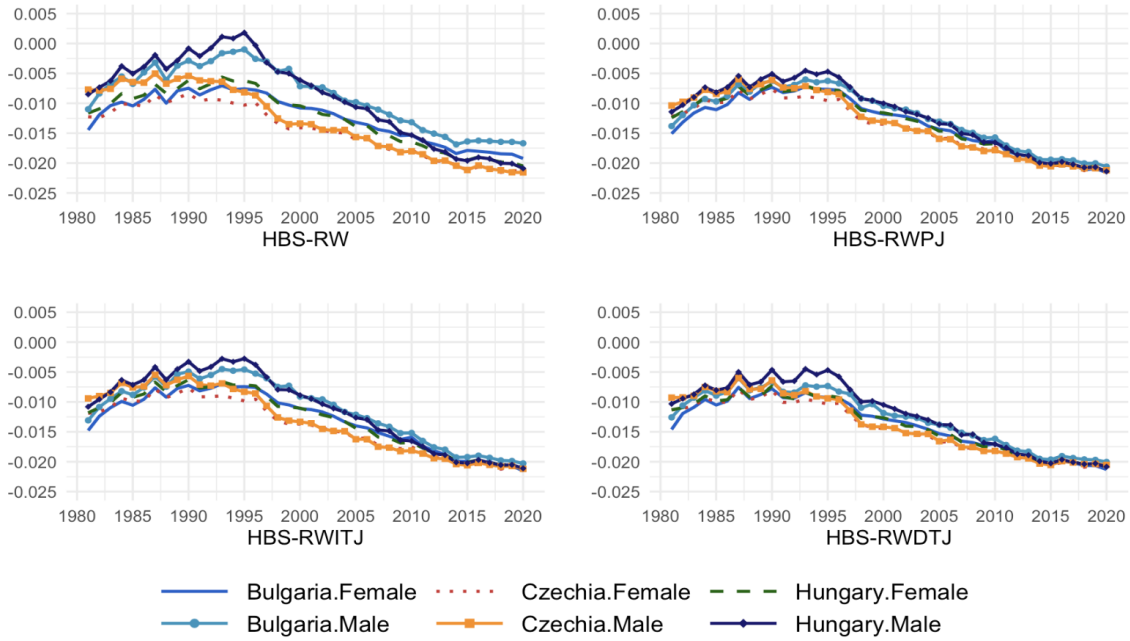

Figure 4.9: $\overline{\theta}^{HBS}_{i,t_U}$ for the East Europe group under the one-group method

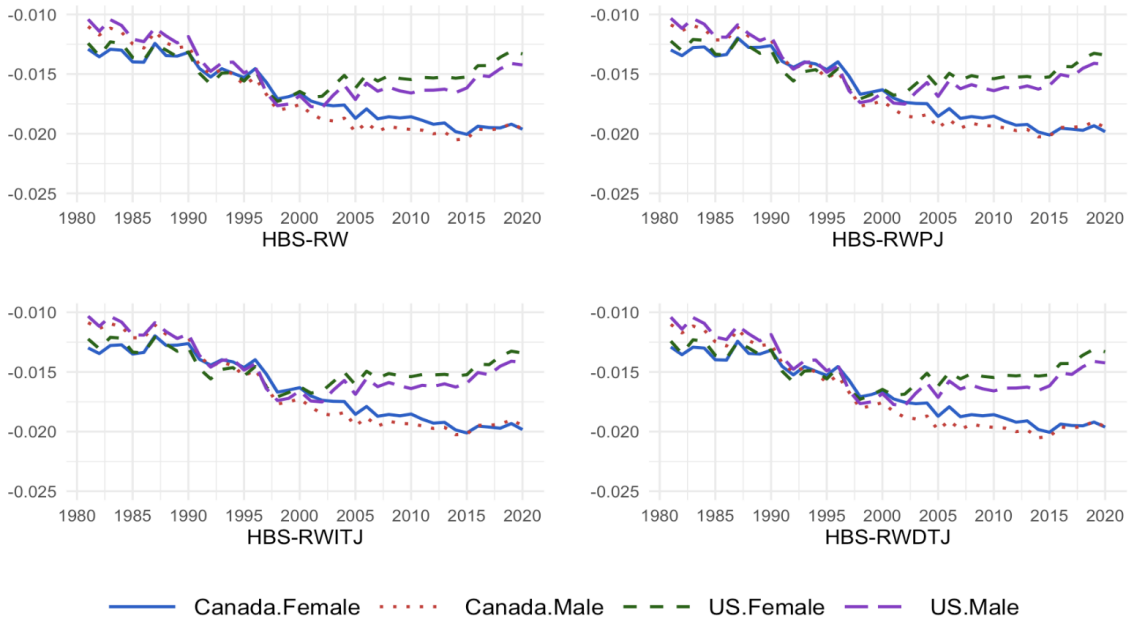

Figure 4.10: $\overline{\theta}^{HBS}_{i,t_U}$ for the North American group under the two-group method

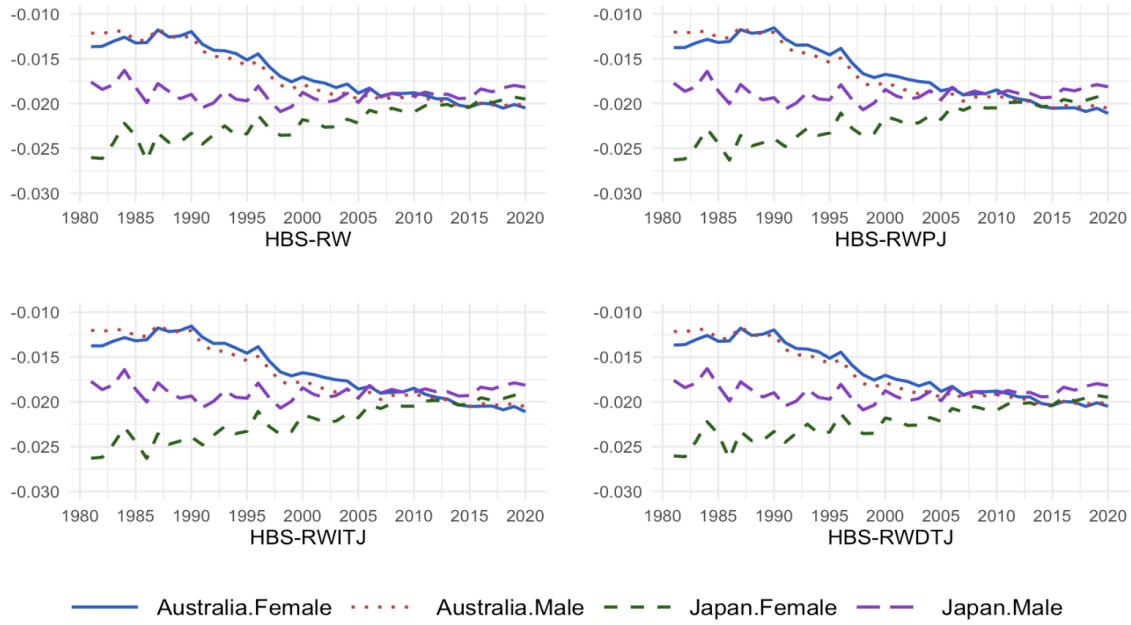

Figure 4.11: $\overline{\theta}^{HBS}_{i,t_U}$ for the Oceania-Asia group under the two-group method

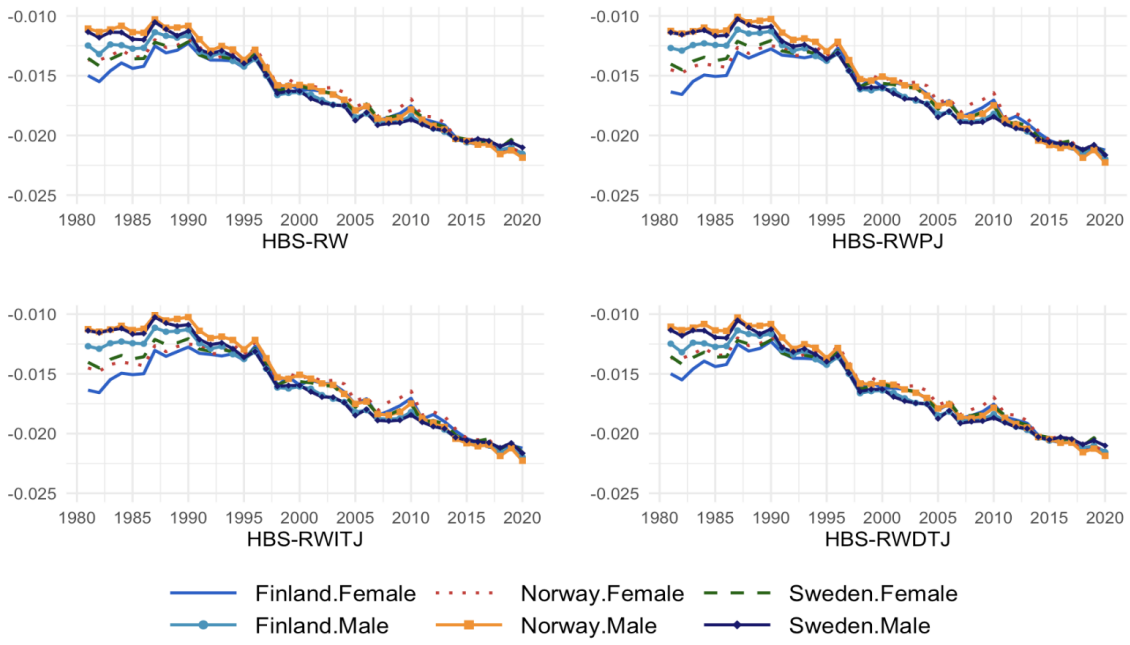

Figure 4.12: $\overline{\theta}_{i,t_U}^{HBS}$ for the North Europe group under the two-group method

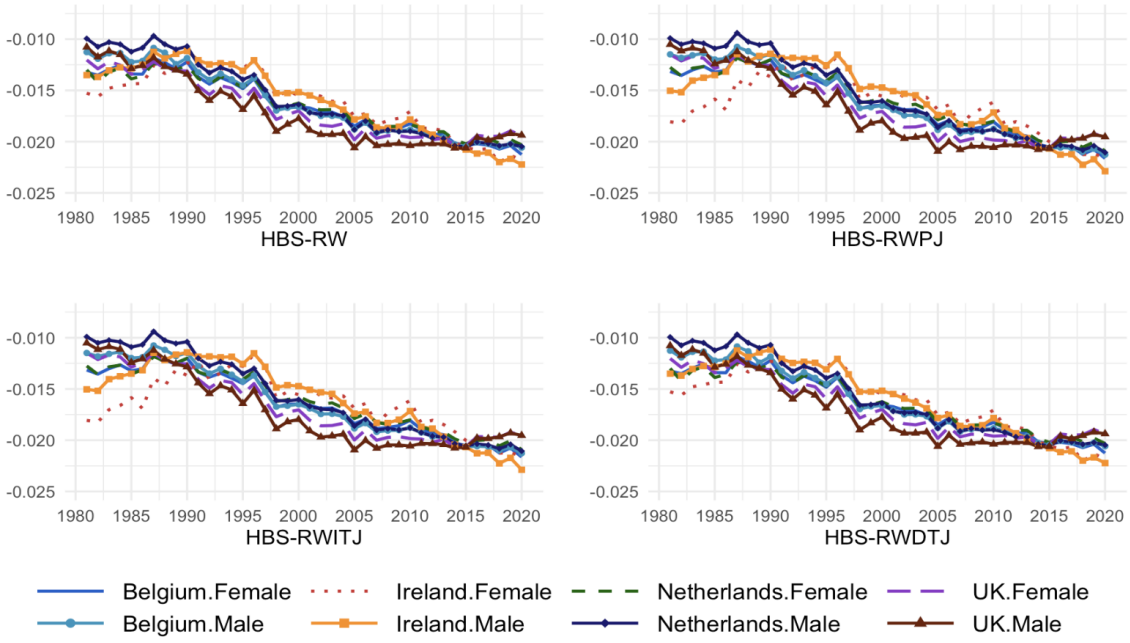

Figure 4.13: $\overline{\theta}_{i,t_U}^{HBS}$ for the West Europe group under the two-group method

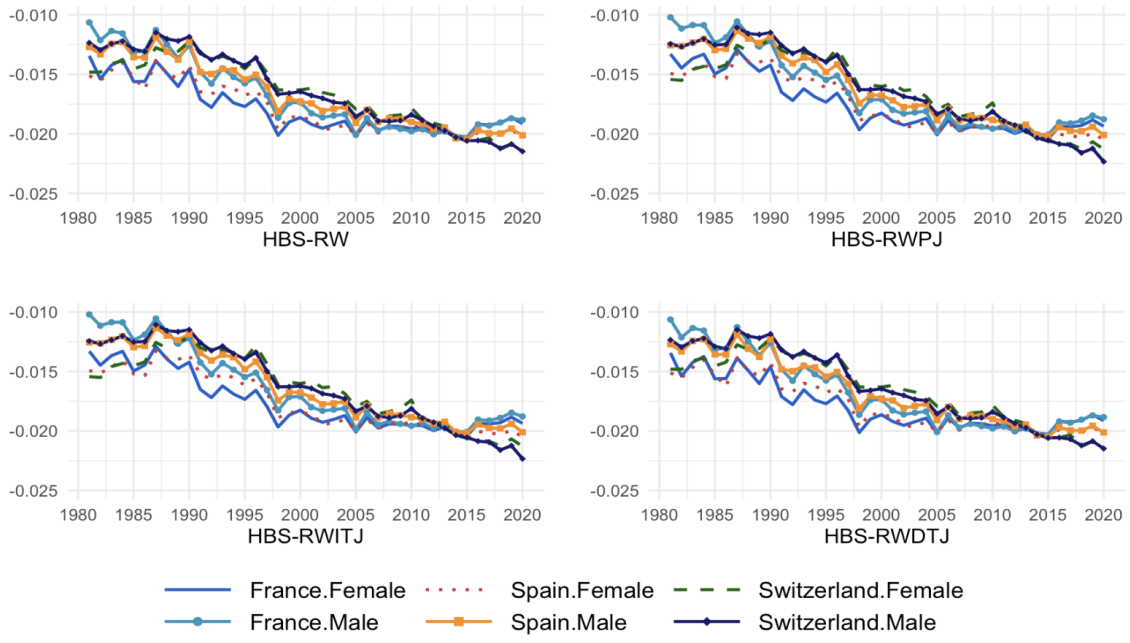

Figure 4.14: $\overline{\theta}_{i,t_U}^{HBS}$ for the South Europe group under the two-group method

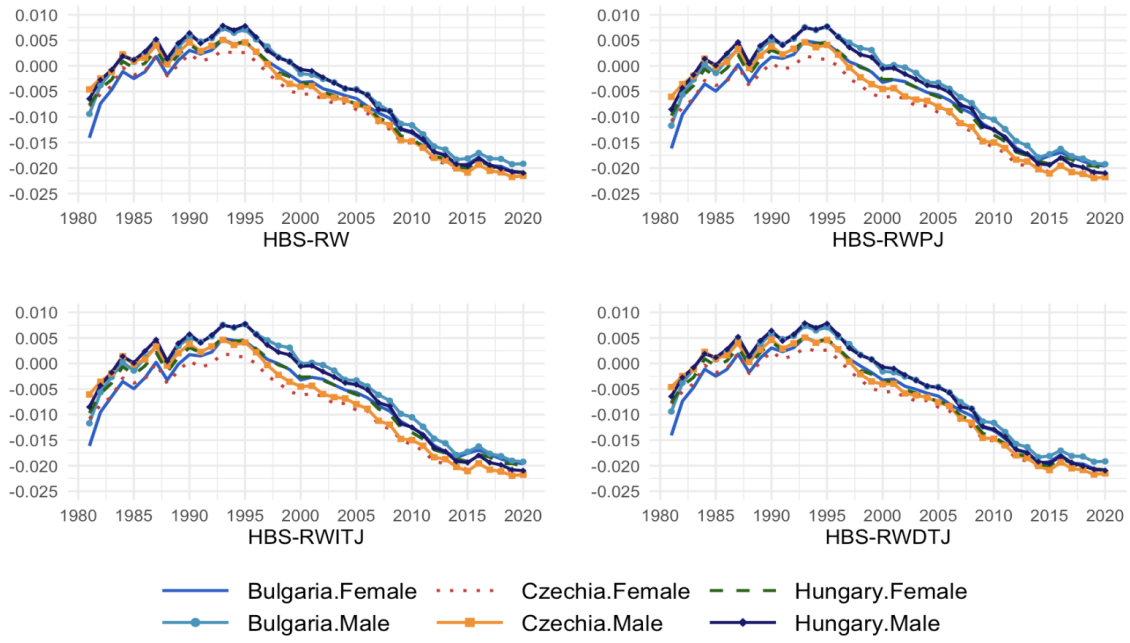

Figure 4.15: $\overline{\theta}_{i,t_U}^{HBS}$ for the East Europe group under the two-group method

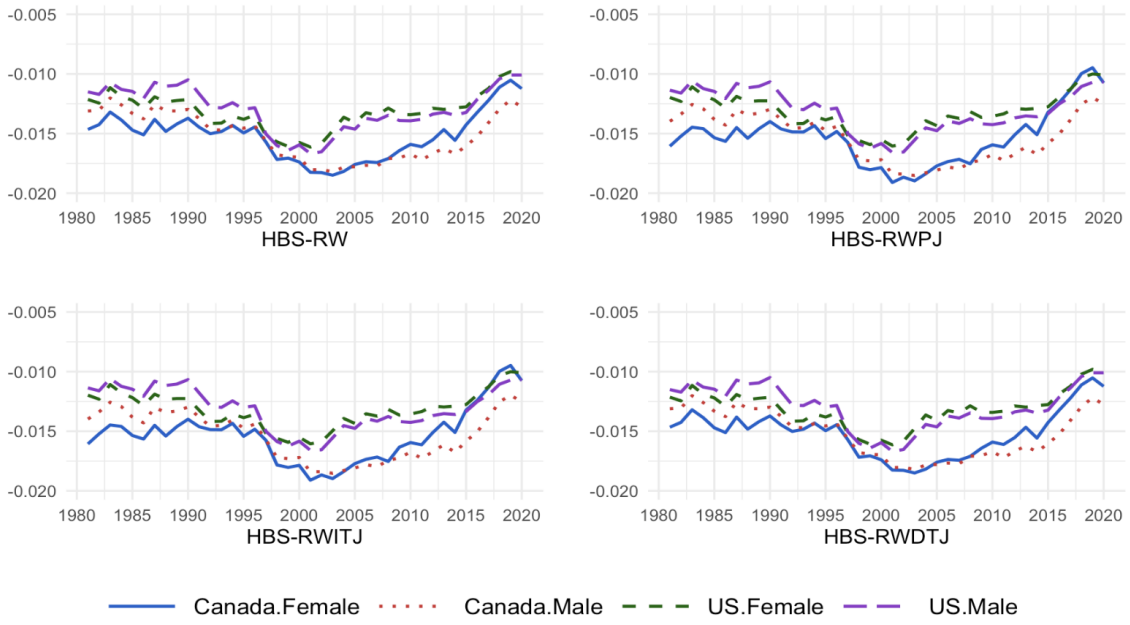

Figure 4.16: $\overline{\theta}_{i,t_U}^{HBS}$ for the North American group under the three-group method

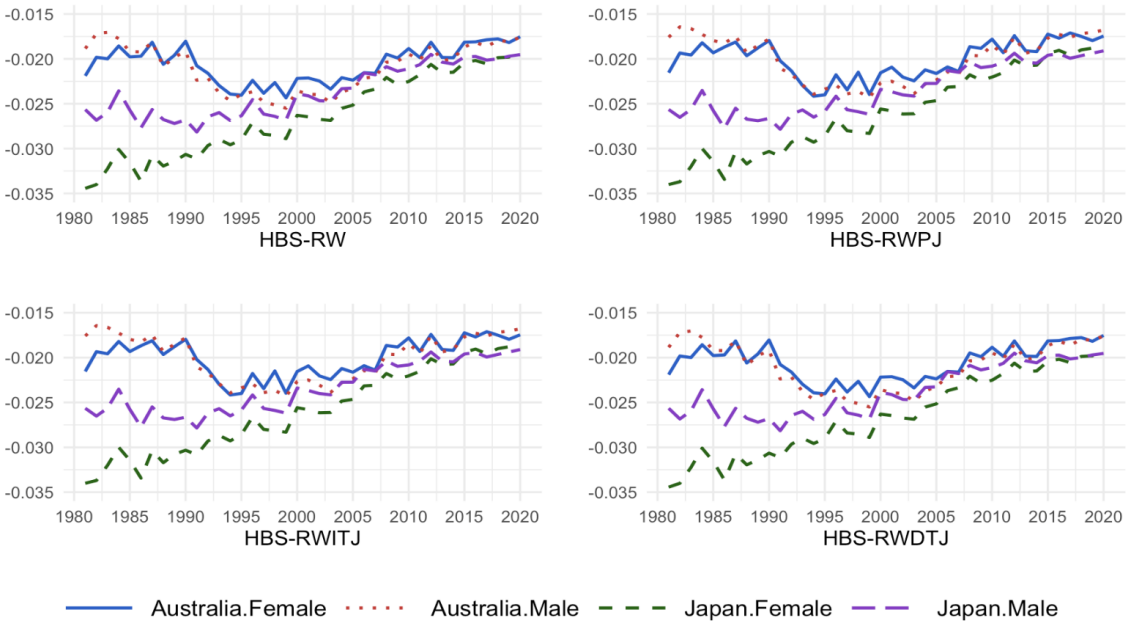

Figure 4.17: $\overline{\theta}_{i,t_U}^{HBS}$ for the Oceania-Asia group under the three-group method

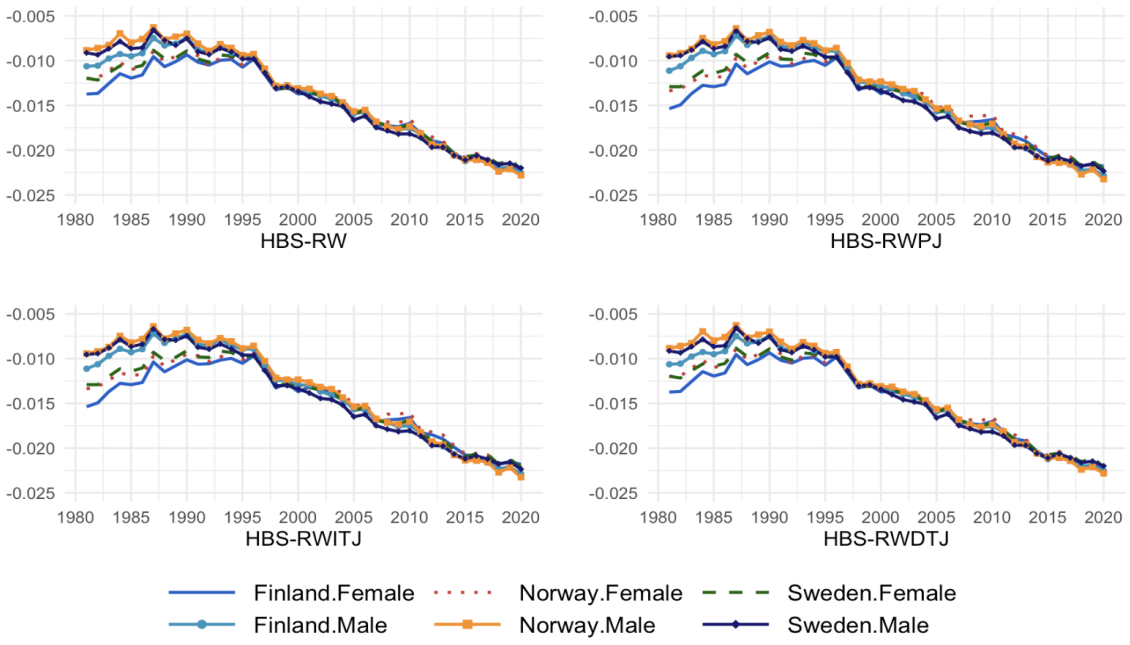

Figure 4.18: $\overline{\theta}^{HBS}_{i,t_U}$ for the North Europe group under the three-group method

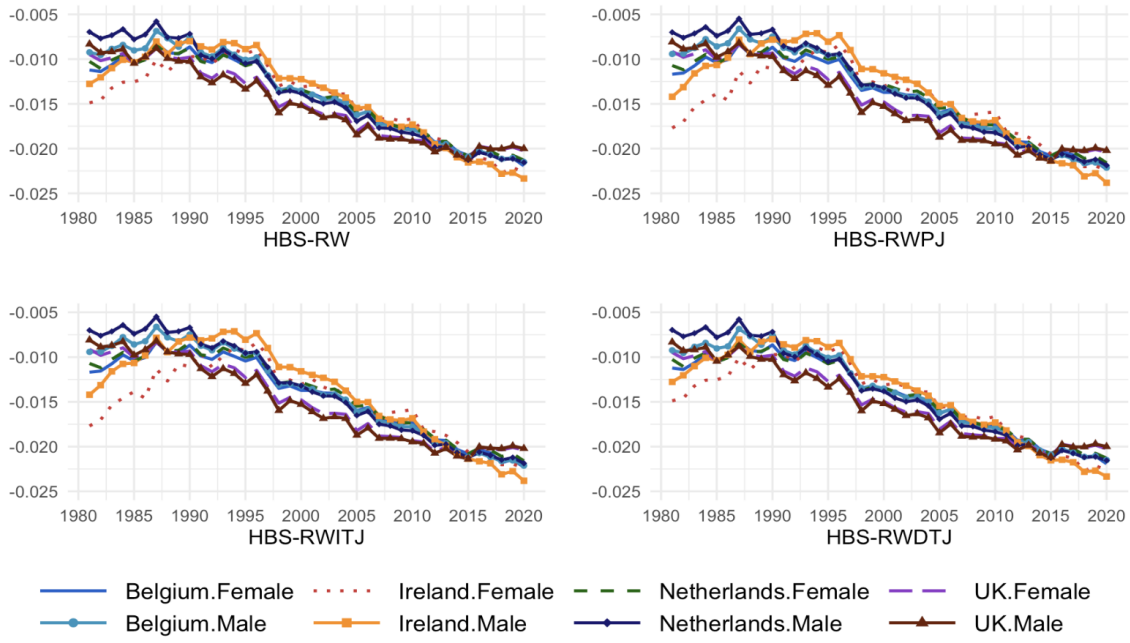

Figure 4.19: $\overline{\theta}^{HBS}_{i,t_U}$ for the West Europe group under the three-group method

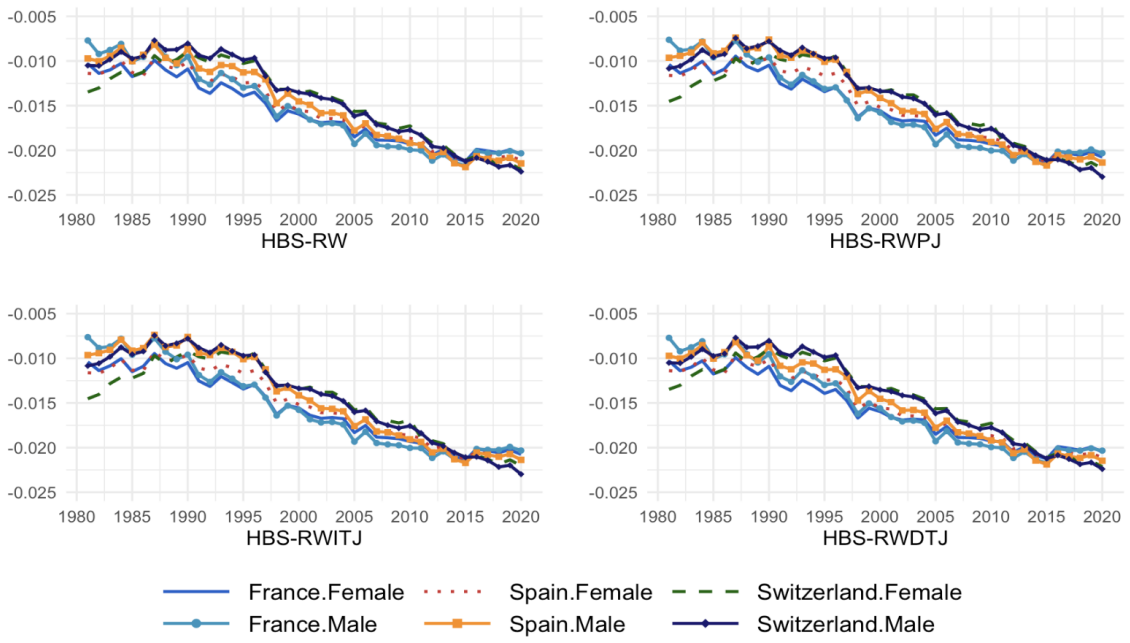

Figure 4.20: $\overline{\theta}^{HBS}_{i,t_U}$ for the South Europe group under the three-group method

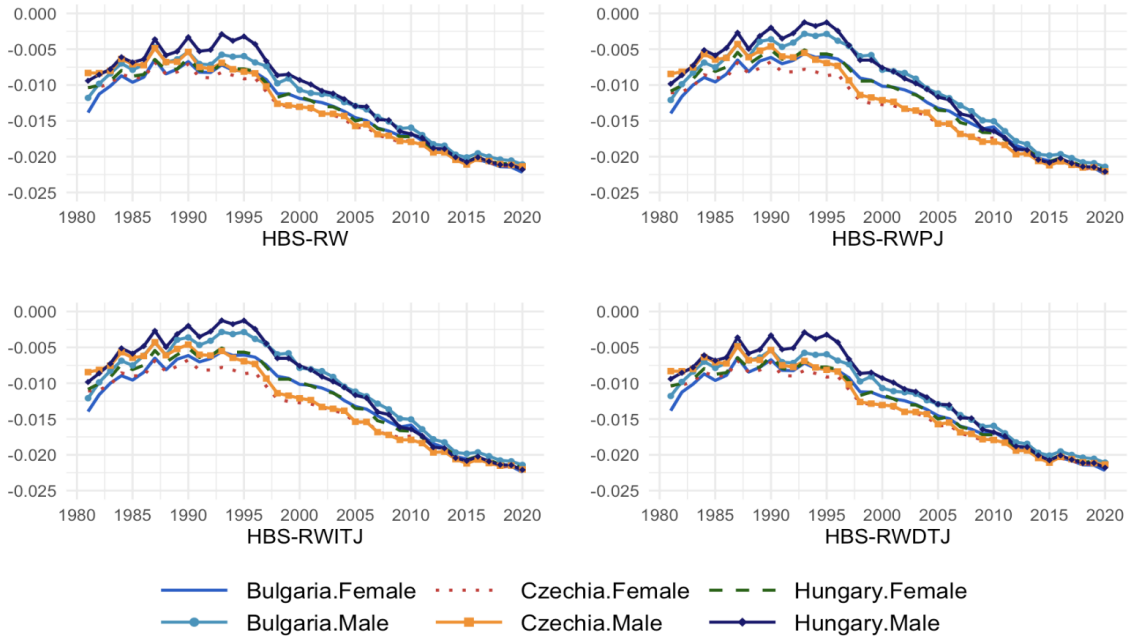

Figure 4.21: $\overline{\theta}^{HBS}_{i,t_U}$ for the East Europe group under the three-group method

Chapter 5

Conclusions

This project incorporates jump components into the random walk with drift model and applies the hierarchical Bayesian theory to modeling of mortality rates for single populations and multiple populations. The studied four models, based on the random walk with drift framework, can be categorized into no jump, permanent jump, independent and transitory jump effects, and dependent and transitory jump effects. These models are applied to the data from the US, the UK, and Japan, representing the Americas, Europe, and Asia, respectively. To validate the predictive performance, these models are compared with the classical Lee-Carter model (for a single population) and its three extensions (JK, CI, and ACF) for multiple populations.

Subsequently, we expand our dataset by adding data from 14 additional countries, bringing the total to 17. We then conduct a comprehensive performance evaluation of the hierarchical jump models using three grouping analyses. Tables 4.2 to 4.5 demonstrate that the hierarchical models with jumps typically outperform the model without jump components and the Lee-Carter-based models. Among them, the HBS-RWITJ model generally produces the lowest $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values in most cases, while the HBS-RWPJ and HBS-RWDTJ models achieve the lowest $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values in a few instances. Our findings indicate that, in most cases, the models incorporating jump components produce the lowest $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values. However, for some specific female populations, the model without jump components may surpass those with jump components.

Finally, applying three grouping analyses to observe the averages of expected improvement rates per year in the logarithm of central death rate for all 17 countries, we find that mortality rate improvements in developing countries tend to be faster than in developed countries over the same time period. Additionally, except for the US, the expected improvement rates for all other countries eventually converge to an average about 2%.

The hierarchical Bayesian models inherit the advantages of the hierarchical Bayesian framework. It aims to address potential inconsistencies in the Lee-Carter model resulting from its two-stage model fitting procedures. Simultaneously, it introduces the co-movement across ages and multiple populations within a highly flexible framework. This flexible approach allows the inclusion of mortality data for any combination of ages and populations, and further assigns weights (the posterior probabilities) to corresponding slope estimates from the random walk with drift models, with or without jump components. The weighted average, called the Bayesian estimate, is then used to generate mortality forecasts for all study ages and populations through a linear function. The introduction of jump models enables the consideration of the impact of extreme events on mortality rates. Our project supports adding trajectory jump effects into the hierarchical Bayesian model, specifically the HBS-RWITJ or HBS-RWDTJ model, both of which achieve the lowest $\text{AMAPE}^{[25,84]}_{[t_U+1,2020],i}$ values in most conducted tests. From a practical perspective, we favor the HBS-RWITJ model due to its computational efficiency. The model estimation process for the HBS-RWITJ model is simpler compared to the HBS-RWDTJ model, where the likelihood involves a series of iterative processes, resulting in slower computational efficiency. In conclusion, our proposed model builds upon the hierarchical Bayesian framework, inheriting its capacity to model dependencies among mortality rates for multiple populations. Notably, we address the limitation of not accounting for adverse mortality risk. Empirical results demonstrate that the models with jump components effectively enhance predictive performances in most cases.

Bibliography

Alexander, M., Zagheni, E., & Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography*, *54*(6), 2025–2041.

Arató, N. M., Dryden, I. L., & Taylor, C. C. (2006). Hierarchical Bayesian modelling of spatial age-dependent mortality. *Computational Statistics and Data Analysis*, *51*(2), 1347–1363.

Becker, G. S., Philipson, T. J., & Soares, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *American Economic Review*, *95*(1), 277–291.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., & Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, *41*(1), 29–59.

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

Chen, H., & Cox, S. H. (2009). Modeling mortality with jumps: Applications to mortality securitization. *Journal of Risk and Insurance*, *76*(3), 727–751.

Cox, S. H., Lin, Y., & Wang, S. (2006). Multivariate exponential tilting and pricing implications for mortality securitization. *Journal of Risk and Insurance*, *73*(4), 719–736.

Czado, C., Delwarde, A., & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, *36*(3), 260–284.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*(419), 659–671.

Leng, X., & Peng, L. (2016). Inference pitfalls in Lee-Carter model for forecasting mortality. *Insurance: Mathematics and Economics*, *70*, 58–65.

Li, H., De Waegenaere, A., & Melenberg, B. (2015). The choice of sample size for mortality forecasting: A Bayesian learning approach. *Insurance: Mathematics and Economics*, *63*, 153–168.

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

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

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

Lin, T., & Tsai, C. C.-L. (2022). Hierarchical Bayesian modeling of multi-country mortality rates. *Scandinavian Actuarial Journal*, *2022*(5), 375–398.

Lin, Y., & Cox, S. H. (2008). Securitization of catastrophe mortality risks. *Insurance: Mathematics and Economics*, *42*(2), 628–637.

Liu, Y., & Li, J. S.-H. (2015). The age pattern of transitory mortality jumps and its impact on the pricing of catastrophic mortality bonds. *Insurance: Mathematics and Economics*, *64*, 135–150.

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

Pedroza, C. (2006). A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics*, *7*(4), 530–550.

Renshaw, A., & Haberman, S. (2003). Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *52*(1), 119–137.

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

Tsai, C. C.-L., & Lin, T. (2017). Incorporating the Bühlmann credibility into mortality models to improve forecasting performances. *Scandinavian Actuarial Journal*, *2017*(5), 419–440.

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

Wong, J. S., Forster, J. J., & Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, *83*, 206–221.

Zhou, R., Li, J. S.-H., & Tan, K. S. (2013). Pricing standardized mortality securitizations: A two-population model with transitory jump effects. *Journal of Risk and Insurance*, *80*(3), 733–774.

Appendix A

Lee-Carter (LC) based model

A.1 Classical Lee-Carter (LC) model

Lee and Carter (1992) introduced the Lee-Carter model for a single population i, with the logarithm of the central death rate, $\ln(m_{x,t,i})$, for age x in year t and population i defined as follows:

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

where $\alpha_{x,i}$ is the average age-specific mortality factor, $k_{t,i}$ is the time-varying index, $\beta_{x,i}$ represents the age-specific reaction to $k_{t,i}$, and the model errors $\varepsilon_{x,t,i}$ for all t are independent and identically distributed, capturing age-specific effects that are not accounted for in the model.

The imposed two constraints, $\sum_{t=t_L}^{t_U} k_{t,i} = 0$ and $\sum_{x=x_L}^{x_U} \beta_{x,i} = 1$, lead to the estimates of $\alpha_{x,i}$ and $k_{t,i}$ as $\hat{\alpha}_{x,i} = \sum_{t=t_L}^{t_U} \ln(m_{x,t,i})/(t_U - t_L + 1)$ and $\hat{k}_{t,i} = \sum_{x=x_L}^{x_U} [\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$, respectively. Furthermore, $\hat{\beta}_{x,i}$ is achieved by regressing the residual $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ on the time-varying index for mortality prediction, $\hat{k}_{t,i}$, without including a constant term; and $\{\hat{k}_{t,i} : t = t_L, t_L + 1, \ldots, t_U\}$ is assumed to follow a random walk with a drift model: $\hat{k}_{t,i} = \hat{k}_{t-1,i} + \theta_i^{LC} + \epsilon_{t,i}$, where the time-varying index errors $\epsilon_{t,i}$ for $t = t_L + 1, \ldots, t_U$ are independent and identically distributed, and are also independent of the model errors. The drift parameter θ_i^{LC} for population i can be derived as

$$\hat{\theta}_i^{LC} = \frac{1}{t_U - t_L} \sum_{t=t_L+1}^{t_U} (\hat{k}_{t,i} - \hat{k}_{t-1,i}) = \frac{\hat{k}_{t_U,i} - \hat{k}_{t_L,i}}{t_U - t_L} = \sum_{x=x_L}^{x_U} \frac{\ln(m_{x,t_U,i}) - \ln(m_{x,t_L,i})}{t_U - t_L},$$

that is, the time-varying index for year $t_U + \tau$ is projected as $\hat{k}_{t_U+\tau,i} = \hat{k}_{t_U,i} + \tau \cdot \hat{\theta}_i^{LC}$, $\tau = 1, 2, \ldots$. Therefore, the logarithm of the projected central death rate for age x in year $t_U + \tau$ and population i can be expressed as a linear function as follows:

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

with intercept $\ln(\hat{m}^{LC}_{x,t_U,i}) = \hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \cdot \hat{k}_{t_U,i}$ and slope $(\hat{\beta}_{x,i} \cdot \hat{\theta}^{LC}_i)$. To assess the impact of the slope on predicting mortality rates, the intercept within the Lee-Carter model is adjusted. The adjustment replaces the fitted value, $\ln(\hat{m}^{LC}_{x,t_U,i})$, with the observed value, $\ln(m_{x,t_U,i})$, effectively preventing jump-off bias. The adjusted LC model is denoted by the LCr model.

A.2 Joint-k (JK) Lee-Carter model

Carter and Lee (1992) proposed a joint-k model designed to consider the co-movement among mortality rates across multiple populations. The logarithm of central death rates, $\ln(m_{x,t,i})$, representing age x in year t and population i, is formulated by incorporating the time-varying index $k_{t,i} = k_t$ for $i = 1, \ldots, I$ into the classical Lee-Carter model.

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

where $\alpha_{x,i}$ is the average age-specific mortality factor, $k_{t,i}$ represents the common time-varying index in year t, $\beta_{x,i}$ is the age-specific reaction to $k_{t,i}$, and the model errors $\varepsilon_{x,t,i}$ for all t are independent and identically distributed, capturing age-specific effects that are not accounted for in the model.

The two constraints, $\sum_{t=t_L}^{t_U} k_t = 0$ and $\sum_{i=1}^{I} \sum_{x=x_L}^{x_U} \beta_{x,i} = 1$, give the estimates of $\alpha_{x,i}$ and k_t as $\hat{\alpha}_{x,i} = \sum_{t=t_L}^{t_U} \ln(m_{x,t,i})/(t_U - t_L + 1)$ and $\hat{k}_t = \sum_{i=1}^{I} \sum_{x=x_L}^{x_U} [\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$, respectively. Furthermore, $\hat{\beta}_{x,i}$ is achieved by regressing the residual $[\ln(m_{x,t,i}) - \hat{\alpha}_{x,i}]$ on the common time-varying index, $\hat{k}_t$, without including a constant term, and $\{\hat{k}_t : t = t_L, t_L+1, \ldots, t_U\}$ is assumed to follow a random walk with a drift model: $\hat{k}_t = \hat{k}_{t-1} + \theta^{JK} + \epsilon_t$, where the time-varying index errors $\epsilon_{t,i}$ for $t = t_L+1, \ldots, t_U$ are independent and identically distributed, and are also independent of the model errors. The drift parameter θ^{JK} can be derived as:

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

Therefore, the logarithm of the projected central death rate for age x in year $t_U + \tau$ and population i under the joint-k (JK) Lee-Carter model can be expressed as a linear function as follows:

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

with intercept $\ln \hat{m}^{JK}_{x,t_U,i} = \hat{\alpha}_{x,i} + \hat{\beta}_{x,i} \cdot \hat{k}_{t_U}$ and slope $(\hat{\beta}_{x,i} \cdot \hat{\theta}^{JK})$. The intercept in the JK model is adjusted, same as the LC model, from the fitted value, $\ln \hat{m}^{JK}_{x,t_U,i}$, to the observed value, $\ln(m_{x,t_U,i})$. The adjusted JK model is denoted as the JKr model.

A.3 Co-integrated (CI) Lee-Carter model

Li and Hardy (2011) introduced the co-integrated model that suggests a linear relationship between the time-varying indexes of population i ($i \geq 2$) and population 1 (the reference

population). This setting requires the re-estimation of the time-varying index for population i ($i \geq 2$).

The classical Lee-Carter model is first applied to each population i, yielding the estimates $\hat{\alpha}_{x,\,i}$, $\hat{\beta}_{x,\,i}$, and $\hat{k}_{t,\,i}$ for $i = 1, \ldots, I$. Subsequently, a linear relationship with an error term $e_{t,\,i}$ between $\hat{k}_{t,\,1}$ and $\hat{k}_{t,\,i}$ is assumed, expressed as $\hat{k}_{t,\,i} = a_i + b_i \cdot \hat{k}_{t,\,1} + e_{t,\,i}$. The re-estimation of $k_{t,\,i}$ is performed through simple linear regression as $\hat{\hat{k}}_{t,\,i} = \hat{a}_i + \hat{b}_i \cdot \hat{k}_{t,\,1}$ with $(\hat{a}_1,\, \hat{b}_1) = (0,\, 1)$. The re-calculated estimate for the drift of the time-varying index for population i, $\hat{\theta}_i^{CI}$, is then given by

$$\hat{\theta}_i^{CI} = \frac{\hat{\hat{k}}_{t_U,\,i} - \hat{\hat{k}}_{t_L,\,i}}{t_U - t_L} = \hat{b}_i \cdot \frac{\hat{k}_{t_U,\,1} - \hat{k}_{t_L,1}}{t_U - t_L} = \hat{b}_i \cdot \hat{\theta}_1^{LC}, \quad i = 1, 2, \ldots, I.$$

Therefore, the logarithm of the projected central death rate for age x in year $t_U + \tau$ and population i under the co-integrated (CI) Lee-Carter model can be expressed as a linear function as follows:

$$\ln(\hat{m}^{CI}_{x,\,t_U+\tau,i}) = \hat{\alpha}_{x,\,i} + \hat{\beta}_{x,\,i} \cdot (\hat{k}_{t_U,\,i} + \tau \cdot \hat{\hat{\theta}}_i^{CI}) = \ln(\hat{m}^{LC}_{x,\,t_U,\,i}) + (\hat{\beta}_{x,\,i} \cdot \hat{\hat{\theta}}_i^{CI}) \cdot \tau, \quad \tau = 1, 2, \ldots,$$

with intercept of $\ln(\hat{m}^{LC}_{x,\,t_U,\,i})$ and a slope of $(\hat{\beta}_{x,\,i} \cdot \hat{\theta}_i^{CI})$. The intercept in the CI model is adjusted, same as the LC model, from the fitted value, $\ln(\hat{m}^{LC}_{x,\,t_U,\,i})$, to the observed value, $\ln(m_{x,\,t_U,\,i})$. The adjusted CI model is denoted by the CIr model.

A.4 Augmented common factor (ACF) Lee-Carter model

Li and Lee (2005) proposed an augmented common factor that considers both commonalities in historical experience and individual trend difference. The classical Lee-Carter model is firstly adjusted to a common factor model by setting $\beta_{x,\,i} = \beta_x$ and $k_{t,\,i} = k_t$ for all populations as

$$\ln(m_{x,\,t,\,i}) = \alpha_{x,\,i} + \beta_x \cdot k_t + \varepsilon_{x,\,t,\,i}, \quad i = 1, \ldots, I, \quad x = x_L, \ldots, x_U, \quad t = t_L, \ldots, t_U.$$

The imposed two constraints, $\sum_{t=t_L}^{t_U} k_t = 0$ and $\sum_{i=1}^{I} \sum_{x=x_L}^{x_U} w_i \cdot \beta_x = 1$, where w_i is set to $1/I$ in this project, representing the weight for population i with the condition $\sum_{i=1}^{I} w_i = 1$, lead to the estimates of $\alpha_{x,\,i}$ and $k_{t,\,i}$ as $\hat{\alpha}_{x,\,i} = \sum_{t=t_L}^{t_U} \ln(m_{x,\,t,\,i})/(t_U - t_L + 1)$ and $\hat{k}_t = \sum_{i=1}^{I} \sum_{x=x_L}^{x_U} w_i \cdot [\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i}]$, respectively. Furthermore, $\hat{\beta}_{x,\,i}$ is achieved by regressing the residual $\sum_{i=1}^{I} w_i \cdot [\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i}]$ on the time-varying index for mortality prediction, $\hat{k}_t$, without including a constant term.

Secondly, an additional factor $\beta'_{x,\,i} \cdot k'_{t,\,i}$ is added to the common factor model by Li and Lee (2005) to incorporate the individual difference in trend. As a result, the augmented common factor model is

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

with an additional constraint, $\sum_{x=x_L}^{x_U} \beta'_{x,\,i} = 1$, which implies that $\hat{k}'_{t,\,i}$ can be determined as $\hat{k}'_{t,\,i} = \sum_{x=x_L}^{x_U} \Big[\ln\left(m_{x,\,t,\,i}\right) - \hat{\alpha}_{x,\,i} - \hat{\beta}_x \cdot \hat{k}_t\Big]$, and $\hat{\beta}'_{x,\,i}$ for each age x can be achieved by regressing $[\ln(m_{x,\,t,\,i}) - \hat{\alpha}_{x,\,i} - \hat{\beta}_x \cdot \hat{k}_t]$ on the time-varying index for mortality prediction, $\hat{k}'_{t,\,i}$, without the constant term. Also, $\hat{k}_t$ and $\hat{k}'_{t,\,i}$ follow random walks with drifts θ^{CF} and θ'^{ACF}_i, respectively. The models are defined as $\hat{k}_t = \hat{k}_{t-1} + \theta^{CF} + \epsilon_t$ and $\hat{k}'_{t,\,i} = \hat{k}'_{t-1,\,i} + \theta'^{ACF}_i + \epsilon_{t,\,i}$, where each of the time-varying index errors ϵ_t and $\epsilon_{t,\,i}$ for $t = t_L + 1, \ldots, t_U$, are independent and identically distributed. Moreover, all three error terms, $\varepsilon_{x,\,t,\,i}$, ϵ_t, and $\epsilon_{t,\,i}$, are also to be independent. The estimates of both drift parameters θ^{CF} and θ'^{ACF}_i can be obtained as $\hat{\theta}^{CF} = (\hat{k}_{t_U} - \hat{k}_{t_L})/(t_U - t_L)$ and $\hat{\theta}'^{ACF}_i = (\hat{k}'_{t_U,\,i} - \hat{k}'_{t_L,i})/(t_U - t_L)$. Therefore, the logarithm of the predicted central death rates for age x in year $t_U + \tau$ and population i under the augmented common factor (ACF) Lee-Carter model can be expressed as a linear function as follows:

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

with intercept $\ln(\hat{m}^{ACF}_{x,\,t_U,\,i}) = \hat{\alpha}_{x,\,i} + \hat{\beta}_x \cdot \hat{k}_{t_U} + \hat{\beta}'_{x,\,i} \cdot \hat{k}'_{t_U,\,i}$ and slope $(\hat{\beta}_x \cdot \hat{\theta}^{CF} + \hat{\beta}'_{x,\,i} \cdot \hat{\theta}'^{ACF}_i)$. The intercept in the ACF model is adjusted, same as the JK and CI models, from the fitted value, $\ln(\hat{m}^{ACF}_{x,\,t_U,\,i})$, to the observed value, $\ln(m_{x,\,t_U,\,i})$. The adjusted ACF model is denoted by the $ACFr$ model.

Appendix B

Density function and log-likelihood function

B.1 Random walk with permanent jump effects

If $N_{x,\,t+1,\,i} = 0$, random variable $Y_{x,\,t,\,i}$ is normally distributed with mean $(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i})$ and variance $\sigma^2_{x,\,i}$; whereas, if $N_{x,\,t+1,\,i} = 1$, random variable $Y_{x,\,t,\,i}$ is normally distributed with mean $(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i})$ and variance $(\sigma^2_{x,\,i} + s^2_{x,\,i})$. The probability density function of $Y_{x,\,t,\,i}$ is given by

$$\begin{aligned} f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}) &= f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t+1,\,i} = 0) \cdot \Pr(N_{x,\,t+1,\,i} = 0) \\ &\quad + f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t+1,\,i} = 1) \cdot \Pr(N_{x,\,t+1,\,i} = 1) \\ &= \frac{1}{\sqrt{2\,\pi\,\sigma^2_{x,\,i}}} \exp\left\{-\frac{[y_{x,\,t,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i})]^2}{2\,\sigma^2_{x,\,i}}\right\} \cdot (1 - \eta_{x,\,i}) \\ &\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}\right\} \cdot \eta_{x,\,i}. \end{aligned} \tag{B.1}$$

Assuming we have observations $y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U,\,i}$, then the log-likelihood function can be expressed as follows:

$$\begin{aligned} &\sum_{t=t_L+1}^{t_U} \ln[f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i})] \\ &= \sum_{t=t_L+1}^{t_U} \ln\left\{\frac{1}{\sqrt{2\,\pi\,\sigma^2_{x,\,i}}} \exp\left(-\frac{[y_{x,\,t,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i})]^2}{2\,\sigma^2_{x,\,i}}\right) \cdot (1 - \eta_{x,\,i})\right. \\ &\quad \left. + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}} \exp\left(-\frac{[y_{x,\,t,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}\right) \cdot \eta_{x,\,i}\right\}. \end{aligned}$$

B.2 Random walk with independent transitory jump effects

The framework of RWITJ assumes an independent data structure. The derivation of the density and likelihood of the RWITJ model can be referenced in the scenarios outlined in Table 3.1.

$$
\begin{aligned}
f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}) &= f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=0)\cdot \Pr(N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=0) \\
&\quad + f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=0)\cdot \Pr(N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=0) \\
&\quad + f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=1)\cdot \Pr(N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=1) \\
&\quad + f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i}|N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=1)\cdot \Pr(N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=1) \\
&= \frac{1}{\sqrt{2\,\pi\,\sigma_{x,\,i}^2}}\exp\left\{-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2)]^2}{2\,\sigma_{x,\,i}^2}\right\}\cdot(1-\eta_{x,\,i})^2 \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}}\exp\left\{-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2-\mu_{x,\,i})]^2}{2\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}\right\}\cdot\eta_{x,\,i}\,(1-\eta_{x,\,i}) \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}}\exp\left\{-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2+\mu_{x,\,i})]^2}{2\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}\right\}\cdot(1-\eta_{x,\,i})\,\eta_{x,\,i} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+2\,s_{x,\,i}^2)}}\exp\left\{-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2)]^2}{2\,(\sigma_{x,\,i}^2+2\,s_{x,\,i}^2)}\right\}\cdot\eta_{x,\,i}^2. \qquad \text{(B.2)}
\end{aligned}
$$

Assuming we have observations $y_{x,\,t_L+1,\,i},\ldots,y_{x,\,t_U,\,i}$, then the log-likelihood function can be expressed as follows:

$$
\begin{aligned}
&\sum_{t=t_L+1}^{t_U}\ln[f_{Y_{x,\,t,\,i}}(y_{x,\,t,\,i})] \\
&= \sum_{t=t_L+1}^{t_U}\ln\left\{\frac{1}{\sqrt{2\,\pi\,\sigma_{x,\,i}^2}}\exp\left(-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2)]^2}{2\,\sigma_{x,\,i}^2}\right)\cdot(1-\eta_{x,\,i})^2\right. \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}}\exp\left(-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2-\mu_{x,\,i})]^2}{2\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}\right)\cdot\eta_{x,\,i}\,(1-\eta_{x,\,i}) \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}}\exp\left(-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2+\mu_{x,\,i})]^2}{2\,(\sigma_{x,\,i}^2+s_{x,\,i}^2)}\right)\cdot(1-\eta_{x,\,i})\,\eta_{x,\,i} \\
&\quad \left. + \frac{1}{\sqrt{2\,\pi\,(\sigma_{x,\,i}^2+2\,s_{x,\,i}^2)}}\exp\left(-\frac{[y_{x,\,t,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma_{x,\,i}^2)]^2}{2\,(\sigma_{x,\,i}^2+2\,s_{x,\,i}^2)}\right)\cdot\eta_{x,\,i}^2\right\}.
\end{aligned}
$$

B.3 Random walk with dependent transitory jump effects

The framework of RWDTJ takes into account dependent data structure. As two consecutive periods are correlated, the conditional maximum likelihood estimation (CMLE) should be employed for parameter estimation. The joint density can be derived as follows:

$$
\begin{aligned}
&f(y_{x,\,t_L+1,\,i}, y_{x,\,t_L+2,\,i}, \ldots, y_{x,\,t_U,\,i}) \\
&= f(y_{x,\,t_U,\,i}|y_{x,\,t_L+1,\,i}, y_{x,\,t_L+2,\,i}, \ldots, y_{x,\,t_U-1,\,i})\, f(y_{x,\,t_L+1,\,i}, y_{x,\,t_L+2,\,i}, \ldots, y_{x,\,t_U-1,\,i}) \\
&= f(y_{x,\,t_U,\,i}|y_{x,\,t_U-1,\,i})\, f(y_{x,\,t_U-1,\,i}|y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U-2,\,i})\, f(y_{x,\,t_L+1,\,i}, \ldots, y_{x,\,t_U-2,\,i}) \\
&\qquad\qquad \vdots \\
&= f(y_{x,\,t_U,\,i}|y_{x,\,t_U-1,\,i})\, f(y_{x,\,t_U-1,\,i}|y_{x,\,t_U-2,\,i}) \cdots f(y_{x,\,t_L+2,\,i}|y_{x,\,t_L+1,\,i})\, f(y_{x,\,t_L+1,\,i}) \\
&= \prod_{t=t_L+1}^{t_U-1} f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i}) \cdot f(y_{x,\,t_L+1,\,i}). \qquad \text{(B.3)}
\end{aligned}
$$

Referring to Equation (3.5), the system exhibits a Markov property: when $N_{x,\,t+1,\,i} = 0$, $Y_{x,\,t,\,i}$ and $Y_{x,\,t+1,\,i}$ are not affected by the jump component in period $t+1$, resulting in $Y_{x,\,t,\,i}$ being independent of $Y_{x,\,t+1,\,i}$. Conversely, when $N_{x,\,t+1,\,i} = 1$, both $Y_{x,\,t,\,i}$ and $Y_{x,\,t+1,\,i}$ are simultaneously driven by the jump component in period $t+1$, requiring further analyses for different combinations of $\{N_{x,\,t,\,i},\, N_{x,\,t+1,\,i},\, N_{x,\,t+2,\,i}\}$. For clarity, $f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i})$ for $t = t_L + 1, \ldots, t_U - 1$ and $f(y_{x,\,t_L+1,\,i})$ in Equation (B.3), with the corresponding scenarios outlined in Tables 3.2 and 3.1, respectively, can be derived as follows:

$$
\begin{aligned}
&f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i}) \\
&= f(y_{x,\,t+1,\,i}|N_{x,\,t+1,\,i} = 0, N_{x,\,t+2,\,i} = 0) \cdot \Pr(N_{x,\,t+1,\,i} = 0, N_{x,\,t+2,\,i} = 0) \\
&\quad + f(y_{x,\,t+1,\,i}|N_{x,\,t+1,\,i} = 0, N_{x,\,t+2,\,i} = 1) \cdot \Pr(N_{x,\,t+1,\,i} = 0, N_{x,\,t+2,\,i} = 1) \\
&\quad + f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i},\, N_{x,\,t,\,i} = 0, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 0) \cdot \Pr(N_{x,\,t,\,i} = 0, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 0) \\
&\quad + f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i},\, N_{x,\,t,\,i} = 1, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 0) \cdot \Pr(N_{x,\,t,\,i} = 1, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 0) \\
&\quad + f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i},\, N_{x,\,t,\,i} = 0, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 1) \cdot \Pr(N_{x,\,t,\,i} = 0, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 1) \\
&\quad + f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i},\, N_{x,\,t,\,i} = 1, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 1) \cdot \Pr(N_{x,\,t,\,i} = 1, N_{x,\,t+1,\,i} = 1, N_{x,\,t+2,\,i} = 1) \\
&= \frac{1}{\sqrt{2\,\pi\,\sigma^2_{x,\,i}}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i})]^2}{2\,\sigma^2_{x,\,i}}\right\} \cdot P_{n\,n} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i} + s^2_{x,\,i})}\right\} \cdot P_{n\,y} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(2\,\sigma^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (-y_{x,\,t,\,i} + 2\,(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}))]^2}{2\,(2\,\sigma^2_{x,\,i})}\right\} \cdot P_{n\,y\,n} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(2\,\sigma^2_{x,\,i} + s^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (-y_{x,\,t,\,i} + 2\,(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} - \mu_{x,\,i}))]^2}{2\,(2\,\sigma^2_{x,\,i} + s^2_{x,\,i})}\right\} \cdot P_{y\,y\,n} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(2\,\sigma^2_{x,\,i} + s^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (-y_{x,\,t,\,i} + 2\,(\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i} + \mu_{x,\,i}))]^2}{2\,(2\,\sigma^2_{x,\,i} + s^2_{x,\,i})}\right\} \cdot P_{n\,y\,y} \\
&\quad + \frac{1}{\sqrt{2\,\pi\,(2\,\sigma^2_{x,\,i} + 2\,s^2_{x,\,i})}} \exp\left\{-\frac{[y_{x,\,t+1,\,i} - (-y_{x,\,t,\,i} + (\alpha_{x,\,i} - \frac{1}{2}\sigma^2_{x,\,i}))]^2}{2\,(2\,\sigma^2_{x,\,i} + 2\,s^2_{x,\,i})}\right\} \cdot P_{y\,y\,y},
\end{aligned}
$$

and

$$
\begin{aligned}
f(y_{x,\,t_L+1,\,i}) &= f(y_{x,\,t_L+1,\,i}|N_{x,\,t_L+1,\,i}=0, N_{x,\,t_L+2,\,i}=0)\cdot \Pr(N_{x,\,t_L+1,\,i}=0, N_{x,\,t_L+2,\,i}=0)\\
&\quad + f(y_{x,\,t_L+1,\,i}|N_{x,\,t_L+1,\,i}=0, N_{x,\,t_L+2,\,i}=1)\cdot \Pr(N_{x,\,t_L+1,\,i}=0, N_{x,\,t_L+2,\,i}=1)\\
&\quad + f(y_{x,\,t_L+1,\,i}|N_{x,\,t_L+1,\,i}=1, N_{x,\,t_L+2,\,i}=0)\cdot \Pr(N_{x,\,t_L+1,\,i}=1, N_{x,\,t_L+2,\,i}=0)\\
&\quad + f(y_{x,\,t_L+1,\,i}|N_{x,\,t_L+1,\,i}=1, N_{x,\,t_L+2,\,i}=1)\cdot \Pr(N_{x,\,t_L+1,\,i}=1, N_{x,\,t_L+2,\,i}=1)\\
&= \frac{1}{\sqrt{2\,\pi\,\sigma^2_{x,\,i}}}\exp\left\{-\frac{[y_{x,\,t_L+1,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma^2_{x,\,i})]^2}{2\,\sigma^2_{x,\,i}}\right\}\cdot P_{n\,n}\\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i}+s^2_{x,\,i})}}\exp\left\{-\frac{[y_{x,\,t_L+1,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma^2_{x,\,i}+\mu_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i}+s^2_{x,\,i})}\right\}\cdot P_{n\,y}\\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i}+s^2_{x,\,i})}}\exp\left\{-\frac{[y_{x,\,t_L+1,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma^2_{x,\,i}-\mu_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i}+s^2_{x,\,i})}\right\}\cdot P_{y\,n}\\
&\quad + \frac{1}{\sqrt{2\,\pi\,(\sigma^2_{x,\,i}+2\,s^2_{x,\,i})}}\exp\left\{-\frac{[y_{x,\,t_L+1,\,i}-(\alpha_{x,\,i}-\frac{1}{2}\sigma^2_{x,\,i})]^2}{2\,(\sigma^2_{x,\,i}+2\,s^2_{x,\,i})}\right\}\cdot P_{y\,y},
\end{aligned}
$$

where

$$
\begin{aligned}
P_{n\,n} &= \Pr(N_{x,\,t+1,\,i}=0, N_{x,\,t+2,\,i}=0),\\
P_{n\,y} &= \Pr(N_{x,\,t+1,\,i}=0, N_{x,\,t+2,\,i}=1),\\
P_{y\,n} &= \Pr(N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=0),\\
P_{y\,y} &= \Pr(N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=1),\\
P_{n\,y\,n} &= \Pr(N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=0),\\
P_{y\,y\,n} &= \Pr(N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=0),\\
P_{n\,y\,y} &= \Pr(N_{x,\,t,\,i}=0, N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=1),\\
P_{y\,y\,y} &= \Pr(N_{x,\,t,\,i}=1, N_{x,\,t+1,\,i}=1, N_{x,\,t+2,\,i}=1).
\end{aligned}
$$

Note that the derivation of $f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i})$ aligns with the previously mentioned fact that when $N_{x,\,t+1,\,i}=0$, $Y_{x,\,t,\,i}$ and $Y_{x,\,t+1,\,i}$ are independent each other. Therefore, there is no need to involve $y_{x,\,t+1,\,i}$ in the first two terms.

Next, assuming we have observations $y_{x,\,t_L+1,\,i},\ldots,y_{x,\,t_U,\,i}$, then the log-likelihood function can be calculated by taking the logarithm on Equation (B.3) as follows:

$$
\begin{aligned}
\ln[f(y_{x,\,t_L+1,\,i},\ldots,y_{x,\,t_U,\,i})] &= \ln[f(y_{x,\,t_U,\,i}|y_{x,\,t_U-1,\,i})\,\cdots f(y_{x,\,t_L+2,\,i}|y_{x,\,t_L+1,\,i})\,f(y_{x,\,t_L+1,\,i})]\\
&= \sum_{t=t_L+1}^{t_U-1}\ln[f(y_{x,\,t+1,\,i}|y_{x,\,t,\,i})] + \log[f(y_{x,\,t_L+1,\,i})].
\end{aligned}
$$