Astin Bulletin (2026), **56**, pp. 770–795
doi:10.1017/asb.2026.10104

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Gradient boosted multi-population mortality modeling with high-frequency data

Ziting Miao, Han Li and Yuyu Chen

Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia
Corresponding author: Ziting Miao; Email: zitingm@student.unimelb.edu.au

Received: 14 July 2025 UTC; **Revised:** 28 March 2026 UTC; **Accepted:** 7 May 2026 UTC;
First published online: 19 June 2026 UTC

Keywords: Gradient boosting; stochastic mortality forecasting; clustering; high-frequency data; multi-population models; vector auto-regressive models

Abstract

High-frequency mortality data have attracted growing attention, but their use has largely been confined to specific applications rather than general modeling and forecasting. Such data pose new challenges to traditional mortality models due to pronounced seasonal patterns and short-term fluctuations. To address these challenges and produce more accurate forecasts with the high-frequency mortality data, this paper introduces a novel integration of gradient boosting techniques into traditional stochastic mortality models under a multi-population setting. Our key innovation lies in using the Li and Lee model as the weak learner within the gradient boosting framework, replacing conventional decision trees. Empirical studies are conducted using weekly mortality data from 30 countries (Human Mortality Database, 2015–2019). Empirical evidence highlights that the proposed methodology not only enhances model fit by accurately capturing underlying mortality trends and seasonal patterns but also achieves superior forecast accuracy, compared to the benchmark models. We also investigate a key challenge in multi-population mortality modeling: how to select appropriate subpopulations with sufficiently similar mortality experiences. A comprehensive clustering exercise is conducted based on mortality improvement rates and seasonal strength. The empirical results demonstrate that our proposed model maintains strong forecast accuracy across different clustering configurations, thereby reducing the need for extensive data preprocessing.

1. Introduction

Most existing research on mortality modeling uses annual data, where short-term mortality fluctuations are often overlooked. Nevertheless, modeling high-frequency data is increasingly important as it allows for capturing short-term mortality dynamics, seasonal fluctuations, and rapid responses to shocks such as epidemics or extreme weather events. While some recent studies have utilized monthly data to improve timely forecasts and support public health planning (see e.g., Léger and Rizzi, 2024; Li and Chen, 2024), the modeling of weekly mortality data remains relatively under-researched and has largely focused on the analysis of excess mortality using the Short-Term Mortality Fluctuations (STMF) data (STMF, 2021). This strand of literature highlights substantial discrepancies between excess deaths and officially reported COVID-19 fatalities, as well as the importance of accounting for temporal dynamics, demographic heterogeneity, and cross-country dependencies in mortality modeling and forecasting (see e.g., Kontis *et al*., 2020; Jdanov *et al*., 2021; Karlinsky and Kobak, 2021; Vanella *et al*., 2021; Hanebeck *et al*., 2025). More importantly, there is an urgent need for a new mortality modeling framework specifically designed for high-frequency mortality analysis – highlighting a crucial gap in current research.

The COVID-19 pandemic has accelerated the availability and use of weekly mortality data, which is originally collected to assess the effectiveness of national containment strategies. It enables timely and precise monitoring of recent mortality dynamics, with minimal reporting delays (Nepomuceno *et al*., 2022). However, weekly mortality data introduce added complexity, including stronger seasonal

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Actuarial Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

patterns and short-term fluctuations that challenge traditional stochastic mortality models, which are designed for lower-frequency data. To address this, we propose a modeling framework specifically designed for high-frequency settings by integrating multi-population stochastic mortality models with gradient boosting techniques. This approach uncovers seasonal patterns in both shared and country-specific mortality trends, while incorporating a highly flexible structure that accommodates heterogeneous mortality experiences across countries and age groups.

The literature on mortality modeling includes numerous efforts to develop stochastic models, both for single-population contexts (see e.g., Osmond, 1985; Lee and Carter, 1992; Jacobsen *et al*., 2002; Renshaw and Haberman, 2006; Hyndman and Ullah, 2007) and for multi-population settings (see e.g., Li and Lee, 2005; Russolillo *et al*., 2011; Hyndman *et al*., 2013; Danesi *et al*., 2015; Lam and Wang, 2023; Dimai, 2025). In this paper, we focus on multi-population mortality models for the following reasons. First, weather-related events such as heatwaves, cold spells, and regional storms typically span across national borders, affecting multiple neighboring countries simultaneously. Accordingly, multi-population modeling allows for the effective capture of joint mortality dynamics, offering valuable insight into the influence of common shocks across interconnected populations. Second, country-specific weekly mortality rates often exhibit considerable noise and variability, which may obscure underlying trends (Arik *et al*., 2026). This issue can be addressed by extracting the common trends across populations, which helps reveal smoother seasonal patterns by "borrowing" experience across countries, thereby generating more coherent future projections. Models based on weekly mortality data require a greater level of flexibility than traditional stochastic mortality models can provide, as they must simultaneously capture age-specific seasonal patterns (e.g., the elderly are more sensitive to temperature extremes), and geographically different seasonal trends (e.g., opposing Northern and Southern Hemisphere cycles).

Gradient boosting has emerged as a powerful tool in mortality modeling due to its flexibility and superior ability to uncover hidden patterns in the data. Deprez *et al.* (2017) incorporate a wide range of individual-level features, including gender, age, year, and cohort to better assess the impact of contributing factors on mortality. Additionally, they demonstrate how regression tree boosting can help detect more patterns and facilitate cause-of-death estimates. Tree-based ensemble methods are also shown to improve both model fitting and forecast performance (see e.g., Levantesi and Pizzorusso, 2019; Bjerre, 2022; Qiao *et al*., 2024). Traditional gradient boosting algorithms iteratively fit weak learners to residuals, gradually improving model performance until a stopping criterion is met. Common implementations include the Gradient Boosting Machine (Friedman, 2001), XGBoost (Chen and Guestrin, 2016), and LightGBM (Ke *et al*., 2017).

The integration of stochastic mortality models as weak learners within a gradient boosting framework is first explored by Li *et al*. (2025). They propose a gradient boosted Lee–Carter model for annual mortality data, where long-term trends dominate and seasonal effects are negligible. To forecast US state-level mortality rates, their approach incorporates spatial dependence via a penalty constraint that shrinks mortality rates according to a predefined adjacency-based matrix reflecting geographic proximity. However, this dependence structure is not applicable to general multi-population settings, particularly those involving countries across different hemispheres.

Instead of the shrinkage approach, we propose a multi-population gradient boosting framework with the Li and Lee (LL) model (Li and Lee, 2005) being the weak learner, which we refer to as the gradient boosted Li and Lee (GBLL) model, specifically designed to accommodate high-frequency mortality data. Through iterative applications of the LL model, the proposed method can capture multiple layers of trends and seasonal variations across all age groups and countries where these patterns are often under-represented in a single fitting. Final forecasts are obtained by aggregating the sequence of fitted LL models, weighted by an optimally tuned learning rate. Moreover, our proposed algorithm terminates once all underlying patterns have been captured, as indicated by residuals approaching white-noise behavior, or when the maximum in-sample fitting error across all horizons, countries, and age groups

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

falls below a specific threshold. A predefined maximum number of iterations is also imposed to prevent potential overfitting. This addresses a key limitation of the gradient boosted Lee–Carter model (Li *et al*., 2025), which determines its stopping criterion solely based on convergence between consecutive fits. Their approach may not be suitable for high-frequency data, where residual differences diminish rapidly between iterations. In contrast, our proposed framework employs a more comprehensive training procedure that preserves informative features across iterations, enhancing the capacity of our proposed model to extract meaningful trends and seasonal signals from high-frequency mortality data, which are often missed by simpler models (see Section 4.1 for more details).

In multi-population mortality modeling, the grouping strategy for subpopulations plays an important role in the identification of the shared mortality trends. Hyndman *et al*. (2013) suggest grouping subpopulations based on predefined characteristics such as sex or states within a country. Departing from this conventional approach, we employ data-driven clustering techniques to identify groups of countries that exhibit similar mortality dynamics, irrespective of economic or geographic classifications. While previous studies primarily group countries based on age-specific mortality patterns (see e.g., Hatzopoulos and Haberman, 2013; Andreopoulos *et al*., 2022; Boonen and Chen, 2026), or form groups using a combination of country and age characteristics (see e.g., Léger and Mazzuco, 2021; Debon *et al*., 2024), we propose to leverage seasonal trends and cyclical structures as the basis for grouping, reflecting the unique properties of high-frequency data. In particular, we highlight the contribution of using seasonal strength, which captures the magnitude of cyclical mortality patterns across populations. In addition to seasonal strength, we consider alternative clustering approaches based on trend slopes and combined features. This provides a broader framework to examine whether the proposed model exhibits improved predictive performance across different data inputs, offering insight into its stable performance under varying configurations.

We apply the proposed GBLL model to weekly mortality data from 30 countries, using the STMF series published by the Human Mortality Database for the period 2015–2019. To ensure comparability in the timing of mortality peaks and troughs, mortality rates for Southern Hemisphere countries (Australia and New Zealand) are inverted to align with the seasonal patterns of Northern Hemisphere countries. While a random walk with drift is typically used to forecast the time trend under annual data (see e.g., Lee and Carter, 1992; Girosi and King, 2007; Shapovalov *et al*., 2019), we incorporate additional Fourier regressors (Serfling, 1963) to account for seasonal cyclical patterns, on top of the time series models. The use of sine and cosine terms allows for the joint modeling of short-term seasonal variation and long-term decline in weekly mortality rates. Empirical results demonstrate that the proposed framework outperforms existing benchmark models in terms of both model fit and forecast accuracy. Moreover, the performance of the proposed model is not sensitive to the clustering selection criteria, which reduces the need for extensive data preprocessing or manual adjustments.

The contribution of this paper can be summarized as follows:

- We are among the first to model and forecast high-frequency mortality rates by introducing a flexible gradient boosting framework that employs the LL model as the weak learner within a multi-population context. To harmonise seasonal patterns across hemispheres, we propose a reciprocal transformation to observed mortality rates in Southern Hemisphere countries. This framework not only captures the underlying short-term seasonal cycles and declining trends but also enhances forecast accuracy.
- We conduct a clustering-based experiment to explore whether certain strategies for grouping countries yield improved performance and to examine the sensitivity of our proposed model to the input data. While certain clustering methods lead to minor improvements and others result in slight performance declines, empirical evidence suggests that our proposed model outperforms the baseline models across all clustering methods. This insensitivity of our framework is further demonstrated by its superior forecast performance, with only marginal differences observed across the three clustering approaches.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

The remainder of this paper is structured as follows. Section 2 describes and visualizes weekly mortality rates for 30 countries. Section 3 introduces the proposed GBLL model. Section 4 presents and compares the fitting and forecast performance of the proposed model with baseline models. Section 5 examines three clustering techniques to group countries with similar mortality experiences and evaluates their forecasting results. Finally, Section 6 concludes the paper and outlines directions for future research. The data and code are available in a GitHub repository at https://github.com/amymiao1019/GBLL. This paper is accompanied by a Supplementary Material with additional analysis regarding alternative parameter choices of the proposed model.

2. Weekly mortality data

2.1. Background and description

In response to the COVID-19 pandemic, the Human Mortality Database launched the STMF series (STMF, 2021) in May 2020, attracting growing interest in the analysis of excess mortality. Jdanov *et al*. (2021) develop an online tool to visualize weekly excess deaths across countries, using 2010–2019 as the reference period. Based on STMF data from 2016 to 2020, Islam *et al*. (2021) find that excess deaths in 29 high-income countries substantially exceed officially reported COVID-19 deaths, particularly among children under 15 years of age. This result is supported by Karlinsky and Kobak (2021), who underscore the need for timely and reliable mortality data to improve forecast accuracy. Kontis *et al*. (2020) focus on countries with populations over four million and incorporate seasonal patterns, temperature effects, long-term trends, and weekly dependencies into their model. Their findings reveal that individuals aged 65 years and older account for 94% of total excess deaths, both in absolute and relative terms. Vanella *et al*. (2021) apply the principal component analysis to mortality data from 19 European countries. They demonstrate that country-specific models often overlook shared common trends, reducing forecast accuracy. Their study underscores the importance of accounting for mortality correlations across countries, particularly given the transnational spread of pathogens. They also identify substantial cross-country variation in excess mortality, driven by differences in public health responses and demographic structures. Similarly, Arik *et al*. (2026) present a cross-country analysis of excess mortality during 2020–2023, showing broadly similar patterns across measures while highlighting that methodological choices can affect the magnitude and variability of the estimates.

While much of the literature has focused on explaining pandemic-related excess mortality, this paper aims to address the broader need for long-term mortality forecasts using high-frequency data. In this paper, we collect mortality data of 30 countries from the STMF series, excluding those with populations under two million due to the frequent occurrence of zero weekly deaths. Such data sparsity introduces substantial uncertainty and can result in unstable model estimates. We extract national-level mortality rates spanning from the second week of 2015 to the final week of 2019, comprising exactly 52 weeks per year. To address the issue of zero mortality values which result in undefined logarithmic transformations, we choose to exclude the youngest age group (0–14 years old) from our analysis. Consequently, we focus on the remaining four age groups: 15–64, 65–74, 75–84, and 85+ years.

2.2. Visualization

The weekly mortality rates across four age groups for all 30 countries over the period 2015–2019 are presented in Appendix A. However, the seasonal patterns are not easily discernible due to the high dimensions of the data. To facilitate a clearer examination of seasonal patterns, we present weekly mortality trends for four representative countries, Australia, England and Wales, Italy, and the USA, in Figure 1. The selection of these countries reflects both their geographic diversity, covering major regions such as Oceania, North America, and Europe, and the opportunity to compare contrasting seasonal patterns between the Northern and Southern Hemispheres. A recurring seasonal pattern is observed across all countries, with each cycle spanning approximately one year. Mortality rates tend to peak during

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

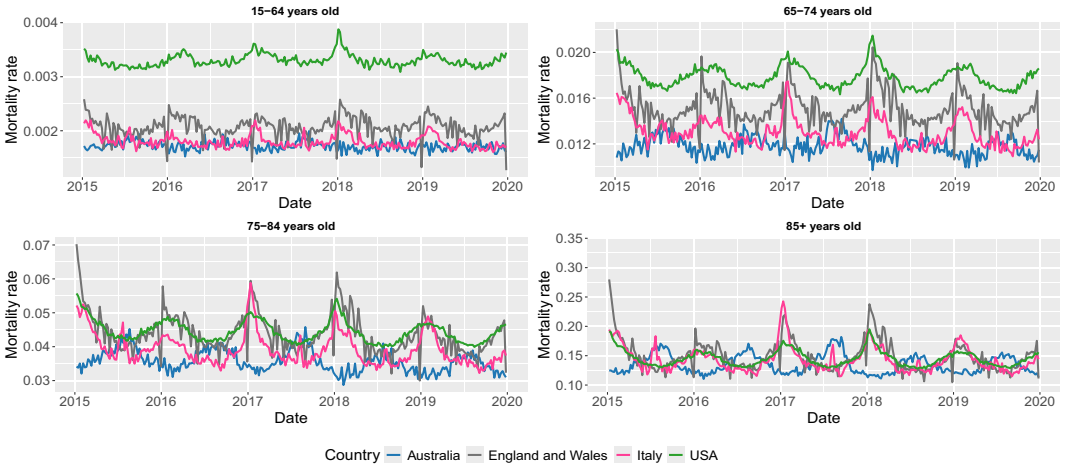


Figure 1. *2015–2019 mortality rates for the four representative countries by age groups.*

the winter and decline in the summer, reflecting the influence of climate factors. As age increases, the seasonal pattern becomes smoother and more pronounced, suggesting a stronger and more apparent relationship between weather conditions and mortality among elderly individuals. This observation aligns with findings from prior research on weather-related mortality (Huang *et al*., 2015). Historical evidence indicates that events such as influenza outbreaks, heatwaves, and extreme cold spells have contributed to short-term fluctuations in mortality rates (Jdanov *et al*., 2021), and the impacts are more pronounced on individuals aged 55 years and older (Li and Tang, 2022).

We then take a closer look at the seasonal timing differences between the Southern and Northern Hemispheres. In Australia, peak mortality rates occur around the middle of the year, corresponding to the winter in the Southern Hemisphere. In contrast, Italy, England and Wales, as well as the USA, exhibit peak mortality at the start of the year, reflecting their winter seasons in the Northern Hemisphere. Notably, beyond the typical U-shaped annual mortality pattern, a W-shaped seasonal trajectory appears in certain countries during specific years. This pattern reflects two distinct mortality peaks within a year: a primary and more pronounced peak in winter, commonly associated with cold waves, and a secondary, smaller summer peak likely driven by heatwaves. The W-shaped seasonal pattern is particularly evident among older age groups (75–84 and 85+ years) and is especially pronounced in Italy and England and Wales.

3. Methodology

In this section, we propose a gradient boosting framework, which integrates the multi-population stochastic mortality model. Section 3.1 presents a brief description of the LL model, which serves as the weak learner within the gradient boosting framework. In Section 3.2, we first introduce the gradient boosting framework, outlining its key components and applications, and then present the proposed GBLL model in detail.

Throughout the paper, the observed mortality rate for country j, age group x, and time t, denoted by $\tilde{m}^j_{x,t}$, is defined as $\tilde{m}^j_{x,t} = \tilde{D}^j_{x,t}/\tilde{E}^j_{x,t}$, where $\tilde{D}^j_{x,t}$ denotes the observed number of deaths and $\tilde{E}^j_{x,t}$ is the corresponding exposure for country j, age group x, and time t. Further details on the construction of these quantities can be found in STMF (2021). We use $m^j_{x,t}$ to denote the modeled mortality rate and $\hat{m}^j_{x,t}$ to denote the predictor of the mortality rates. A model parameter with a hat is used to denote an estimated value of the parameter (e.g., $\hat{\theta}$ denotes the predicted value of a generic model parameter θ). Bold symbols represent matrices, and lowercase letters refer to their elements (e.g., $x_{i,j}$ denotes the element in row i and column j of the matrix $\boldsymbol{X}$).

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

3.1. LL model

Li and Lee (2005) propose the multi-population version of the Lee–Carter model to simultaneously estimate and forecast the mortality rates of a group of populations with similar economic conditions and close-knit connections. The Lee–Carter model proposed by Lee and Carter (1992) serves as the foundation for the estimation of the LL model, and its definition and estimation algorithm are summarized in Appendix B. Li and Lee (2005) develop an augmented common factor model: for age group $x \in \{1, ..., N\}$, time $t \in \{1, ..., T\}$, and country $j \in \{1, ..., J\}$,

$$\log(m_{x,t}^j) = A_x^j + b_x^p \kappa_t^p + b_x^j \kappa_t^j + \varepsilon_{x,t}^j,$$

where $m_{x,t}^j$ is the mortality rate for age group x and country j at time t, A_x^j is the age-specific mean of log mortality rate over time for country j, and $\varepsilon_{x,t}^j$ is the corresponding error term for age group x, time t, and country j. The common trend is captured by b_x^p and κ_t^p, obtained by applying the Lee–Carter method for the combined group including all the populations to avoid long-term divergence in mean mortality forecasts. In addition, b_x^j and κ_t^j represent the shorter-term excessive rate of mortality change in country j over the common trend.

The original estimation method proposed by Li and Lee (2005) first extracts the age-specific mean of log mortality rates A_x^j. Then the common trend $b_x^p \kappa_t^p$ across all subpopulations is computed using the singular value decomposition (SVD) on the pooled, mean-adjusted log mortality rates $\log(m_{x,t}^j) - A_x^j$. Population-specific residuals are then analyzed via the SVD to obtain additional components $b_x^j \kappa_t^j$ to capture the deviations from the common trend. However, to ensure more coherent and simultaneous modeling across populations with other benchmark models, we follow the product-ratio functional method proposed by Hyndman *et al*. (2013) with the order chosen to be 1, to estimate the LL model. The detailed estimation is outlined as follows. First, let $\boldsymbol{Y}^j$ denote the $T \times N$ (time × age group) matrix of the observed log mortality rates for country j, whose (t, x)-th element is given by $y_{x,t}^j = \log(\tilde{m}_{x,t}^j)$. We then re-express the observed mortality rate as $\tilde{m}_{x,t}^j = \tilde{p}_{x,t} \tilde{r}_{x,t}^j$ where $\tilde{p}_{x,t} = (\tilde{m}_{x,t}^1 \times ... \times \tilde{m}_{x,t}^j)^{1/J}$ is the product term, calculated as the geometric mean of the observed mortality rate across all countries, and $\tilde{r}_{x,t}^j = \tilde{m}_{x,t}^j / \tilde{p}_{x,t}$ represents the ratio term for country j. We use $\boldsymbol{Y}^j$ to fit separate Lee–Carter models to the product and ratio terms, to get the common factors $\hat{\boldsymbol{\kappa}}_t^p, \hat{\boldsymbol{a}}_x^p, \hat{\boldsymbol{b}}_x^p$ and country-specific factors $\hat{\boldsymbol{\kappa}}_t^j, \hat{\boldsymbol{a}}_x^j, \hat{\boldsymbol{b}}_x^j$, respectively; see Appendix B for the model specification a_x, b_x, and κ_t of the Lee–Carter model. We then compute the overall age-specific mean as $\hat{\boldsymbol{A}}_x^j = \hat{\boldsymbol{a}}_x^p + \hat{\boldsymbol{a}}_x^j$ and the final predicted log mortality rate is given by

$$\hat{y}_{x,t}^j = \log(\hat{m}_{x,t}^j) = \hat{A}_x^j + \hat{b}_x^p \hat{\kappa}_t^p + \hat{b}_x^j \hat{\kappa}_t^j,$$

for age group x, time t, and country j. The complete estimation procedure is summarized in Algorithm 1.

3.2. GBLL model

Given concerns about the capacity of existing stochastic mortality models to accurately capture seasonality of mortality rates, we adopt a gradient boosting framework to iteratively model different layers of mortality trends. Gradient boosting is an additive modeling technique that iteratively improves predictive accuracy by correcting residual errors from previous fits (Friedman, 2001). Due to its flexibility and predictive power, the gradient boosting method can capture complex, nonlinear relationships and enhance both model fitting and forecast accuracy. Consequently, this data-driven approach has become increasingly popular in mortality rate modeling (see e.g., Blanes-Selva *et al*., 2021; Asgari *et al*., 2022; Qiu *et al*., 2022; Neshat *et al*., 2026).

There are two main components in the gradient boosting framework, namely the weak learner and the associated learning rate. Traditionally, a weak learner refers to a simple model with limited predictive power that performs slightly better than random guessing, such as a shallow decision tree. More recently, boosting has been used with other weak learners, including generalized linear models (Bühlmann and Hothorn, 2007), generalized additive models (Tutz and Binder, 2006), and copulas (Hans *et al*., 2023). By sequentially aggregating each fitted weak learner to the final model with a learning rate that controls

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Algorithm 1: Estimation of the LL model (Product-Ratio Method with order 1)

Input: $\boldsymbol{Y}^j$ for $j \in \{1, ..., J\}$ (all with size $T \times N$)
Output: $\hat{y}^j_{x,t}$ for $x \in \{1, ..., N\}$, $t \in \{1, ..., T\}$ and $j \in \{1, ..., J\}$
$\hat{\boldsymbol{\kappa}}^p_t = (\hat{\kappa}^p_1, ..., \hat{\kappa}^p_T)$, $\hat{\boldsymbol{a}}^p_x = (\hat{a}^p_1, ..., \hat{a}^p_N)$ **and** $\hat{\boldsymbol{b}}^p_x = (\hat{b}^p_1, ..., \hat{b}^p_N)$
$\hat{\boldsymbol{\kappa}}^j_t = (\hat{\kappa}^j_1, ..., \hat{\kappa}^j_T)$, $\hat{\boldsymbol{a}}^j_x = (\hat{a}^j_1, ..., \hat{a}^j_N)$ **and** $\hat{\boldsymbol{b}}^j_x = (\hat{b}^j_1, ..., \hat{b}^j_N)$
Procedure: LL($\boldsymbol{Y}^j$)
1: $\tilde{p}_{x,t} \leftarrow (\tilde{m}^1_{x,t} \times ... \times \tilde{m}^j_{x,t})^{\frac{1}{J}}$, $\forall x, t$
2: Denote $\boldsymbol{P}$ by $\log(\tilde{p}_{x,t})$ in row t and column x
3: $\tilde{r}^j_{x,t} \leftarrow \frac{\tilde{m}^j_{x,t}}{\tilde{p}_{x,t}}$, $\forall x, t$ and j
4: Denote $\boldsymbol{R}^j$ by $\log(\tilde{r}^j_{x,t})$ in row t and column x
5: $\hat{\boldsymbol{\kappa}}^p_t, \hat{\boldsymbol{a}}^p_x, \hat{\boldsymbol{b}}^p_x \leftarrow \mathrm{LC}(\mathbf{P})$, $\forall x, t$
6: $\hat{\boldsymbol{\kappa}}^j_t, \hat{\boldsymbol{a}}^j_x, \hat{\boldsymbol{b}}^j_x \leftarrow \mathrm{LC}(\boldsymbol{R}^j)$, $\forall x, t$ and j
7: $\hat{\boldsymbol{A}}^j_x = \hat{\boldsymbol{a}}^p_x + \hat{\boldsymbol{a}}^j_x$, $\forall x, j$
8: $\hat{y}^j_{x,t} = \log(\hat{m}^j_{x,t}) \leftarrow \hat{A}^j_x + \hat{b}^p_x \hat{\kappa}^p_t + \hat{b}^j_x \hat{\kappa}^j_t$, $\forall x, t$ and j

the contribution of each weak learner to the final model, it results in a stronger and more comprehensive model with enhanced forecast accuracy.

In this paper, we propose to employ the LL model as the weak learner to leverage domain expertise in mortality modeling, enhancing the interpretability of the overall model. Gradient boosting seeks to construct an ensemble of weak learners defined as $\sum_{g=1}^{l} \gamma_g \hat{\boldsymbol{y}}^j_{x,g}$, where $\hat{\boldsymbol{y}}^j_{x,g}$ denotes the g-th predicted log mortality rate from the LL model for country j and age group x, γ_g represents the associated coefficient, and l is the number of iterations. This procedure involves fitting each weak learner to the residuals and solving a one-dimensional optimization problem to determine its contribution.

To estimate the GBLL model, we first let $\boldsymbol{Y}^j = (\boldsymbol{y}^j_1, ..., \boldsymbol{y}^j_N)$ denote the input mortality matrix with dimension $T \times N$ (time $\times$ age group) for country j. In the first iteration, the parameter set $\{\hat{\boldsymbol{\kappa}}^p_{t,1}, \hat{\boldsymbol{a}}^p_{x,1}, \hat{\boldsymbol{b}}^p_{x,1}, \hat{\boldsymbol{\kappa}}^j_{t,1}, \hat{\boldsymbol{a}}^j_{x,1}, \hat{\boldsymbol{b}}^j_{x,1}\}$ is obtained by fitting the LL model to $\boldsymbol{Y}^j$. Let $\hat{\boldsymbol{Y}}^j_1 = (\hat{\boldsymbol{y}}^j_{1,1}, ..., \hat{\boldsymbol{y}}^j_{N,1})$ denote the fitted value of the first iteration. The learning rate γ_1 is then determined by minimizing the following quadratic loss function:

$$L_1(\gamma) = \sum_{j=1}^{J} \left\| \boldsymbol{Y}^j - \gamma \hat{\boldsymbol{Y}}^j_1 \right\|_F^2,$$

where $\|\cdot\|_F$ represents the Frobenius norm that measures the overall magnitude of a matrix, that is, L_1 can be equivalently written as:

$$L_1(\gamma) = \sum_{j=1}^{J} \sum_{x=1}^{N} (\boldsymbol{y}^j_x - \gamma \hat{\boldsymbol{y}}^j_{x,1})'(\boldsymbol{y}^j_x - \gamma \hat{\boldsymbol{y}}^j_{x,1}).$$

In the $(g+1)$-th iteration for $g \geq 1$, the parameter set $\{\hat{\boldsymbol{\kappa}}^p_{t,g+1}, \hat{\boldsymbol{a}}^p_{x,g+1}, \hat{\boldsymbol{b}}^p_{x,g+1}, \hat{\boldsymbol{\kappa}}^j_{t,g+1}, \hat{\boldsymbol{a}}^j_{x,g+1}, \hat{\boldsymbol{b}}^j_{x,g+1}\}$ and the fitted value $\hat{\boldsymbol{Y}}^j_{g+1} = (\hat{\boldsymbol{y}}^j_{1,g+1}, \ldots, \hat{\boldsymbol{y}}^j_{N,g+1})$ are obtained by fitting the LL model to residual $\boldsymbol{E}^j_g = (\boldsymbol{e}^j_{1,g}, ..., \boldsymbol{e}^j_{N,g})$ from the g-th iteration, defined by:

$$\boldsymbol{E}^j_g = \begin{cases} \boldsymbol{Y}^j - \gamma_1 \hat{\boldsymbol{Y}}^j_1 = (\boldsymbol{e}^j_{1,1}, ..., \boldsymbol{e}^j_{N,1}) = (\boldsymbol{y}^j_1 - \gamma_1 \hat{\boldsymbol{y}}^j_{1,1}, ..., \boldsymbol{y}^j_N - \gamma_1 \hat{\boldsymbol{y}}^j_{N,1}), & \text{if } g = 1, \\ \boldsymbol{E}^j_{g-1} - \gamma_g \hat{\boldsymbol{Y}}^j_g = (\boldsymbol{e}^j_{1,g}, ..., \boldsymbol{e}^j_{N,g}) = (\boldsymbol{e}^j_{1,g-1} - \gamma_g \hat{\boldsymbol{y}}^j_{1,g}, ..., \boldsymbol{e}^j_{N,g-1} - \gamma_g \hat{\boldsymbol{y}}^j_{N,g}), & \text{if } g \geq 2. \end{cases}$$

The learning rate γ_g for $g \geq 2$ is determined by minimizing the following quadratic loss function:

$$L_g(\gamma) = \sum_{j=1}^{J} \left\| \boldsymbol{E}^j_{g-1} - \gamma \hat{\boldsymbol{Y}}^j_g \right\|_F^2.$$

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

The algorithm terminates when (a) all 120 residual series (30 countries, 4 age groups each) have Ljung–Box test p-values of at least 0.05 (Ljung and Box, 1978), or (b) the maximum in-sample fitting MAPE across all age groups, countries and training periods falls below a small threshold. For condition (a), the p-value $p^j_{x,g}$ for each age group x, country j and iteration g is given by $1 - F_{\chi^2_{52}}\left(T(T+2)\sum_{k=1}^{52}(\rho^{(j)}_{k,x,g})^2/(T-k)\right)$, where $F_{\chi^2_{52}}(\cdot)$ represents the cumulative distribution function (CDF) of a chi-squared distribution with 52 degrees of freedom, T is the whole training period, lag is chosen at 52, and $\rho^{(j)}_{k,x,g}$ is the sample autocorrelation at lag k for the residual $\boldsymbol{e}^j_{x,g}$ for country j, age group x, and the g-th iteration. For condition (b), the maximum in-sample fitting MAPE across all age groups $x \in \{1, ..., N\}$, countries $j \in \{1, ..., J\}$, and training periods $t \in \{1, ..., T\}$ is defined as (see Section 4.2 for more details on the definition of MAPE):

$$\max_{x,j,t}\left(\text{MAPE}_{x,j,t}\right) = \max_{x,j,t}\left(\frac{|\hat{m}^j_{x,t} - \tilde{m}^j_{x,t}|}{\tilde{m}^j_{x,t}}\right).$$

The algorithm stops if the maximum in-sample fitting MAPE is less than 0.001. This constraint prevents unnecessary iterations once a sufficiently accurate fit is reached and avoids reducing the fitting error to values that may reflect overfitting rather than meaningful improvement. The 0.001 threshold aligns with common optimization tolerances, generally ranging from 10^{-6} to 10^{-3} (Nocedal and Wright, 2006). It suggests that the residuals contain no remaining systematic patterns and the model adequately captures the underlying data structure. We set the maximum number of boosting iterations to $G = 50$. A detailed summary of this procedure is provided in Algorithm 2. The fitted log mortality rates are obtained as follows: for age group $x \in \{1, ..., N\}$, time $t \in \{1, ..., T\}$, and country $j \in \{1, ..., J\}$,

$$\log(\hat{m}^j_{x,t}) = \sum_{g=1}^{l} \gamma_g \left(\hat{a}^p_{x,g} + \hat{b}^p_{x,g}\hat{\kappa}^p_{t,g} + \hat{a}^j_{x,g} + \hat{b}^j_{x,g}\hat{\kappa}^j_{t,g}\right).$$

To ensure consistency in seasonal mortality patterns across countries, particularly between the hemispheres, we address the phase reversal in seasonal cycles observed in weekly mortality data. Specifically, winter-related mortality peaks occur in opposite calendar months across hemispheres. During model estimation, we therefore apply a temporary monotonic reciprocal transformation (i.e., $\tilde{m}^{j*}_{x,t} = 1/\tilde{m}^j_{x,t}$ where $\tilde{m}^{j*}_{x,t}$ and $\tilde{m}^j_{x,t}$ are the transformed and raw observed mortality rates, respectively) to the Southern Hemisphere mortality rates (Australia and New Zealand) to align the phase of their seasonal patterns with those of Northern Hemisphere countries. This transformation is used solely for estimation purposes to facilitate the identification of a common seasonal component within the multi-population framework. After forecasting, all series are transformed back to their original scale (i.e., $\hat{m}^j_{x,t} = 1/\hat{m}^{j*}_{x,t}$ where $\hat{m}^j_{x,t}$ is the fitted value of the mortality rate and $\hat{m}^{j*}_{x,t}$ is the fitted value of the transformed mortality rate), ensuring that the resulting mortality forecasts remain fully interpretable.

Forecasting mortality rates primarily relies on projecting both the common and country-specific time trend component κ_t. In classical mortality modeling, κ_t is often assumed to follow a random walk with drift. For high-frequency weekly mortality data, a natural extension is to employ a seasonal-ARIMA (SARIMA) model to account for annual seasonality. However, SARIMA models require seasonal differencing at lag 52, which effectively removes a full year of observations from the already short five-year sample (260 weeks), potentially leading to unstable parameter estimates and unreliable forecasts. Moreover, SARIMA with a long seasonal period is computationally intensive, especially in a multi-population gradient boosting framework. To overcome this limitation, we adopt a hybrid forecasting approach in which seasonal patterns are modeled deterministically using Fourier regressors (Serfling, 1963), while the residual stochastic dynamics are captured by a non-seasonal ARIMA(p,d,q) process. This strategy allows us to retain the full sample for estimation while flexibly capturing both annual and biannual seasonal cycles. Specifically, κ_t is modeled as

$$\kappa_t = \beta_1 \sin\left(2\pi w(t)\right) + \beta_2 \cos\left(2\pi w(t)\right) + \beta_3 \sin\left(4\pi w(t)\right) + \beta_4 \cos\left(4\pi w(t)\right) + \text{ARIMA}(p, d, q),$$

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Algorithm 2: Estimation of the GBLL Model

Input: $\boldsymbol{Y}^j$ for $j \in \{1, ..., J\}$ (all with size $T \times N$)

Output: $\hat{\boldsymbol{\kappa}}^p_{t,l}, \hat{\boldsymbol{a}}^p_{x,l}, \hat{\boldsymbol{b}}^p_{x,l}$ and γ_l for $x \in \{1, ..., N\}$, $t \in \{1, ..., T\}$ and $l \in \{1, ..., G\}$

$\hat{\boldsymbol{\kappa}}^j_{t,l}, \hat{\boldsymbol{a}}^j_{x,l}, \hat{\boldsymbol{b}}^j_{x,l}$ for $x \in \{1, ..., N\}$, $t \in \{1, ..., T\}$, $l \in \{1, ..., G\}$ and $j \in \{1, ..., J\}$

Procedure: GBLL($\boldsymbol{Y}^j$)

$\hat{\boldsymbol{y}}^j_{x,1}, \hat{\boldsymbol{\kappa}}^p_{t,1}, \hat{\boldsymbol{a}}^p_{x,1}, \hat{\boldsymbol{b}}^p_{x,1}, \hat{\boldsymbol{\kappa}}^j_{t,1}, \hat{\boldsymbol{a}}^j_{x,1}, \hat{\boldsymbol{b}}^j_{x,1} \leftarrow \text{LL}(\boldsymbol{Y}^j)$, $\forall x, t$ and j

$L_1 \leftarrow \sum_{j=1}^{J} \sum_{x=1}^{N} (\boldsymbol{y}^j_x - \gamma \hat{\boldsymbol{y}}^j_{x,1})'(\boldsymbol{y}^j_x - \gamma \hat{\boldsymbol{y}}^j_{x,1})$

$\gamma_1 = \arg\min_{\gamma} L_1$

$\boldsymbol{e}^j_{x,1} \leftarrow \boldsymbol{y}^j_x - \gamma_1 \hat{\boldsymbol{y}}^j_{x,1}$, $\forall x, j$

$\boldsymbol{E}^j_1 \leftarrow (\boldsymbol{e}^j_{1,1}, ..., \boldsymbol{e}^j_{N,1})$, $\forall j$

$g \leftarrow 1$

while $g \leq G$ **do**

$p^j_{x,g} \leftarrow 1 - F_{\chi^2_{52}}\left(T(T+2)\sum_{k=1}^{52} \frac{(\rho^{(j)}_{k,x,g})^2}{T-k}\right)$, the p–value of $\boldsymbol{e}^j_{x,g}$ under the Ljung–Box test for country j and age group x

if $g = G$ **or** $\min_{x,j} p^j_{x,g} \geq 0.05$ **or** $\max_{x,j,t} \frac{|\hat{m}^j_{x,t} - \tilde{m}^j_{x,t}|}{\tilde{m}^j_{x,t}} < 0.001$ **then**

$l \leftarrow g$

break

else

$\hat{\boldsymbol{y}}^j_{x,g+1}, \hat{\boldsymbol{\kappa}}^p_{t,g+1}, \hat{\boldsymbol{a}}^p_{x,g+1}, \hat{\boldsymbol{b}}^p_{x,g+1}, \hat{\boldsymbol{\kappa}}^j_{t,g+1}, \hat{\boldsymbol{a}}^j_{x,g+1}, \hat{\boldsymbol{b}}^j_{x,g+1} \leftarrow \text{LL}(\boldsymbol{E}^j_g)$, $\forall x, t$ and j

$L_{g+1} \leftarrow \sum_{j=1}^{J} \sum_{x=1}^{N} (\boldsymbol{e}^j_{x,g} - \gamma \hat{\boldsymbol{y}}^j_{x,g+1})'(\boldsymbol{e}^j_{x,g} - \gamma \hat{\boldsymbol{y}}^j_{x,g+1})$

$\gamma_{g+1} = \arg\min_{\gamma} L_{g+1}$

$\boldsymbol{e}^j_{x,g+1} \leftarrow \boldsymbol{e}^j_{x,g} - \gamma_{g+1} \hat{\boldsymbol{y}}^j_{x,g+1}$, $\forall x, j$

$\boldsymbol{E}^j_{g+1} \leftarrow (\boldsymbol{e}^j_{1,g+1}, ..., \boldsymbol{e}^j_{N,g+1})$, $\forall j$

$g \leftarrow g + 1$

where $w(t) \in [0, 1)$ denotes the fractional part of the calendar year corresponding to time t. There are four Fourier regressors, including both sine and cosine functions to capture both the annual and biannual cycles. The statistical significance of the Fourier components is assessed using a Wald test on the corresponding regression coefficients (Wald, 1943). Specifically, for each Fourier term, the null hypothesis that its coefficient is equal to zero is tested using the associated z-statistic $z_i = \hat{\beta}_i / \text{SE}(\hat{\beta}_i)$, for $i = 1, 2, 3, 4$, along with the corresponding p-value obtained from the `coeftest` function in `R`. Under

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

the large-sample approximation, the test statistic follows $Z \sim \mathcal{N}(0, 1)$. A Fourier regressor is deemed statistically significant at the 5% level if its p-value is less than 0.05. If none of the Fourier components satisfy this criterion, the forecasting procedure relies solely on the ARIMA(p,d,q) model. Therefore, standing on the final observed time T, the final h-step-ahead forecasts of the log mortality rate is given by: for age group $x \in \{1, ..., N\}, j \in \{1, ..., J\}$,

$$\log(\hat{m}^j_{x,T+h}) = \sum_{g=1}^{l} \gamma_g \left(\hat{a}^p_{x,g} + \hat{b}^p_{x,g} \hat{\kappa}^p_{T+h|T,g} + \hat{a}^j_{x,g} + \hat{b}^j_{x,g} \hat{\kappa}^j_{T+h|T,g} \right).$$

There are alternative choices for each component of the GBLL model, including the loss function, stopping criterion, models used to fit and forecast the time trend κ_t, and seasonal adjustments for countries in the Southern Hemisphere. Detailed descriptions of these alternatives are provided in Section A of the Supplementary Material. Nevertheless, our main conclusions remain unchanged, and the proposed GBLL model is shown to be insensitive to the choice of individual components.

4. Empirical results

This section evaluates both the in-sample fitting (Section 4.1) and out-of-sample forecast performance (Section 4.2) of the proposed GBLL model, in comparison with the baseline LL model (see Section 3.1), the Hyndman–Booth–Yasmeen (HBY) model, the Vector Auto-Regressive (VAR) model, and the multi-population Global Vector Auto-Regressive (GVAR) model. An overview of the HBY model is given in Appendix C, and Appendix D contains a detailed analysis based on the VAR and GVAR model.

To mitigate the impact of stochastic variability in the model fitting and forecast process, we adopt a 10-fold expanding window approach. In this setup, the starting point of the estimation window is fixed at the initial point of the dataset, the second week of January 2015. The endpoint is incrementally advanced from the final week of March 2018 to the final week of December 2018, increasing by one month at each step. This results in a progressively larger training window over successive iterations. The forecast horizon is held constant at 52 weeks, and forecast accuracy is computed monthly across all age groups and all countries. By reducing the estimation bias through repeated sampling, this approach yields more reliable and stable results.

4.1. Fitting performance

Empirically, the GBLL algorithm terminates due to condition (b), typically after around eight iterations. The maximum in-sample fitting error across all horizons, countries, and age groups is below 0.001, indicating that the model has captured all underlying significant patterns and meaningful variations in the data. This adequacy can be further assessed by examining the number of residual series exhibiting white-noise behavior across age groups and countries using the Ljung–Box test at a 5% significance level (Ljung and Box, 1978).

For each expanding window, we count the number of residual series identified as white noise (out of 120, corresponding to 30 countries and 4 age groups). The average number of white-noise residuals across the 10 expanding windows show that the LL and HBY models produce 54 and 15 white-noise residuals, respectively, while the proposed GBLL model (with eight iterations) produces 73. This indicates that the GBLL model captures nearly all significant and meaningful patterns in the mortality data, as also reflected by the small maximum in-sample fitting error (below 0.001). In contrast, the HBY model fails to capture some structures, particularly for the 65–74 and 75–84 years of age groups, while the LL model performs better but still leaves more residual structure unexplained. The higher number of white-noise residuals in the GBLL model, together with the relatively small fitting error, demonstrates its superior ability to extract the underlying trends and seasonality across countries and age groups. It is worth noting that our comparison focuses on the HBY model with order 6; alternative specifications may perform differently, but exploring the full range of configurations is beyond the scope of this paper.

Looking closely at those residuals that are not white noise, there remain either seasonal cyclical patterns or clear downward/upward sloping trends or both. This means that, given the complexity and

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

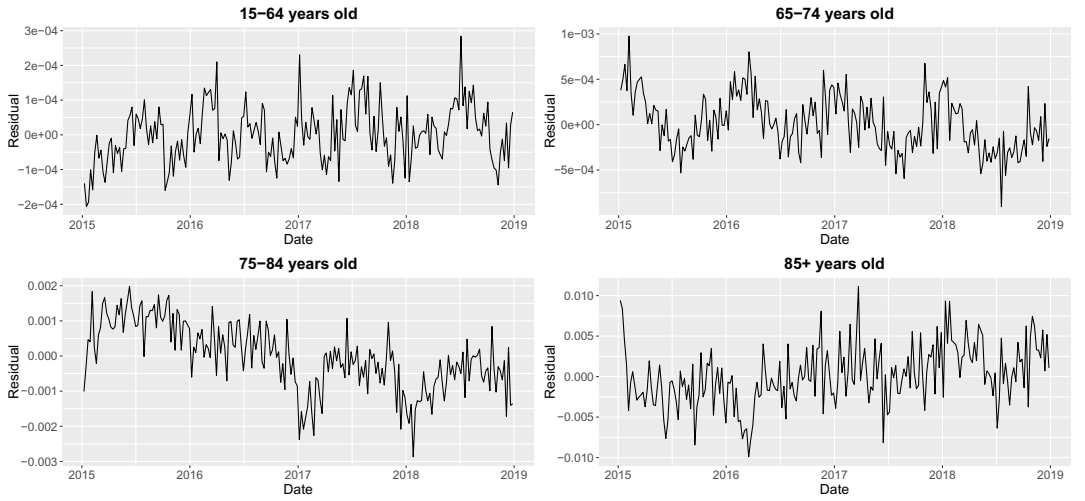


Figure 2. *The residual plot of Canada under the LL model.*

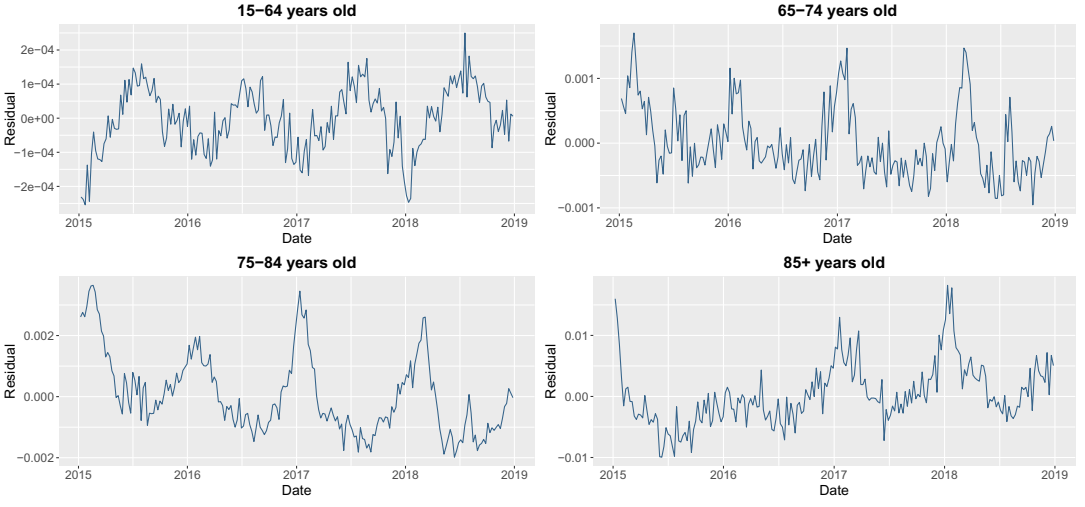


Figure 3. *The residual plot of Canada under the HBY model.*

high dimensions of the dataset, it is not suitable to use the existing methods, as they fail to capture all the patterns. For instance, in the case of Canada, residual plots from both the LL and HBY models reveal remaining seasonal cycles and trends. Figure 2 shows clear seasonal cycles for the 65–74 years old and a downward sloping trend for the 75–84 years old, under the LL model. Similarly, Figure 3 gives the residual plot of Canada under the HBY model, and there are clear seasonal cyclical patterns for all the four age groups and a slightly decreasing trend for the 65–74 years old. Notably, the age group of 15–64 years exhibits an inverse seasonal pattern. This may be due to the model imposing seasonal structure to accommodate for other age groups, even though the raw data show little seasonal variation. Additionally, the residuals for individuals aged 85 years and over are approximately around ± 0.01, clearly indicating that both the LL and HBY models struggle to capture the underlying patterns for this age group. In contrast, the residuals across all four age groups under the GBLL model exhibit white-noise behaviors, indicating improved model adequacy. Additional visualizations of in-sample fits for the GBLL model and the benchmark models for selected countries are provided in Section C.1 of the Supplementary Material.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

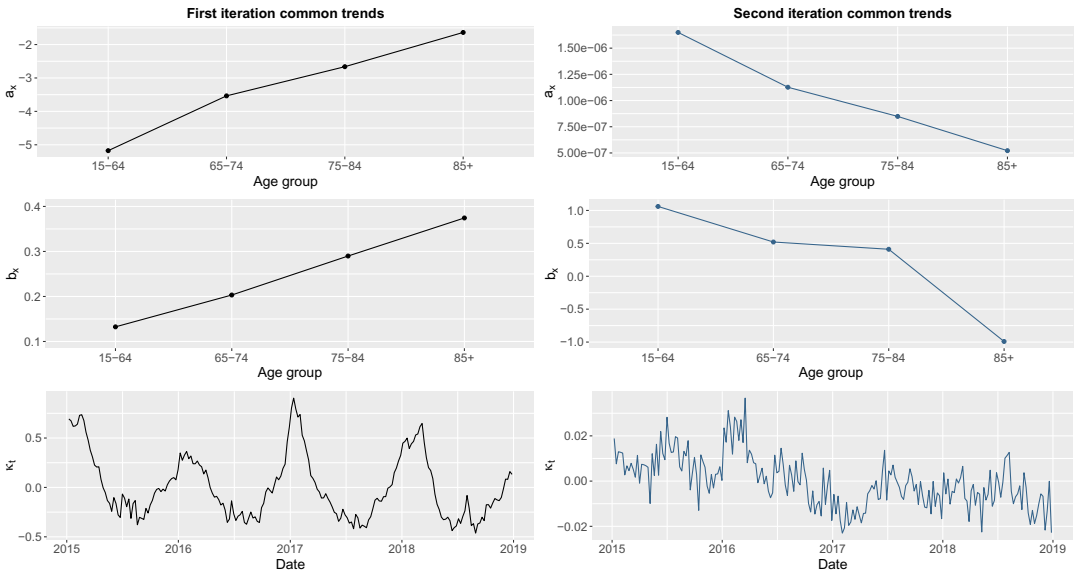


Figure 4. *Components of the common trends under the GBLL model: first iteration (left) and second iteration (right).*

Finally, Figure 4 illustrates the contribution of each component at different stages of the model fitting process. We present the decomposition of the GBLL model for the first two iterations here, focusing on the common trends. The country-specific components breakdown are provided in Appendix E. The first column (left) displays, from top to bottom, the estimated age-specific intercepts ($\hat{a}_x$), age loadings ($\hat{b}_x$), and the time-varying index ($\hat{\kappa}_t$) from the first iteration. The second column (right) presents the corresponding components from the second iteration.

- For the age-specific intercepts $\hat{a}_x$, the first iteration captures a generally increasing trend. It highlights the vulnerability of older people, resulting in higher average mortality rates. In the second iteration, a predominantly decreasing pattern is observed. Importantly, the magnitude of these adjustments is on the order of 10^{-6}, suggesting that the primary age-related mortality pattern is effectively captured in the first iteration.
- For the age loadings $\hat{b}_x$, the first iteration exhibits an upward trend, indicating the greater exposure experienced by elderly individuals to higher mortality rates. In the second iteration, a declining trend is observed. This may be related to the original smoother and more pronounced seasonal patterns typically exhibited by older age groups. As these dominant effects are captured in the first iteration, only a small portion of the residual variation remains to be explained. Additionally, some spikes exist in the 15–64 and 85+ age groups, indicating that this iteration primarily captures residual effects concentrated in the extreme age cohorts.
- For the common time trend component $\hat{\kappa}_t$, the first iteration effectively captures the dominant seasonal cycles. The second iteration provides a refinement by identifying smaller-scale cyclical patterns with short-term fluctuations and a mild downward trend, indicative of gradual improvements in mortality rates over time. A prominent W-shaped seasonal pattern is evident in the first iteration across all years, arising from the aggregation of mortality dynamics across all countries. This aggregation enhances structural features that may be weakly present but commonly shared across populations, thereby making the W-shape (with dual seasonal peaks) more pronounced. Additionally, an M-shaped pattern is discernible, particularly around early 2015 and 2018. This may be attributed to the reciprocal transformation applied to the Southern Hemisphere countries, where a secondary mortality peak, likely associated with heatwaves, typically occurs at the beginning of the calendar year.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

4.2. Forecast performance with expanding windows

To evaluate the forecast performance, we adopt the Mean Absolute Percentage Error (MAPE) as the error measure to quantify the average magnitude of the percentage deviation between forecasted and actual values. MAPE is scale-independent, allowing direct comparison of forecast accuracy across countries with different mortality levels. It has been widely used in the mortality forecasting literature as a standard metric for evaluating predictive performance (see e.g., McNown and Rogers, 1992; Neves *et al*., 2017; Li *et al*., 2019; Tsai and Cheng, 2021; Qiao *et al*., 2024). We first define the set $H = \{4, 9, 13, 17, 22, 26, 30, 35, 39, 43, 48, 52\}$, which corresponds to the cumulative number of weeks until the end of each month, thereby facilitating the conversion from weekly to monthly forecast horizons. For month $h \in \{1, ..., 12\}$, expanding window $r \in \{0, ..., 9\}$, the MAPE for h-step-ahead forecasts across all age groups, countries, and expanding windows is given by:

$$\text{MAPE}_h = \frac{1}{4 \times H_{(h)} \times 30 \times 10} \sum_{x=1}^{4} \sum_{u=1}^{H_{(h)}} \sum_{j=1}^{30} \sum_{r=0}^{9} \frac{|\hat{m}^j_{x,169+u+H_{(r)}} - \tilde{m}^j_{x,169+u+H_{(r)}}|}{\tilde{m}^j_{x,169+u+H_{(r)}}},$$

where $H_{(i)}$ represents the i-th order statistic of the set H with $H_{(0)} = 0$ and 169 is the number of weeks in the first training sample. An additional scale-invariant error metric, the Mean Absolute Scaled Error (MASE), is also employed and leads to comparable conclusions to those obtained using MAPE. Full results and discussions are provided in Section B of the Supplementary Material.

Table 1. *Mean MAPE of out-of-sample forecasts across 30 countries (*$\times 100$ *scale).*

h	LL	HBY	GBLL	VAR	GVAR
1	5.436	5.985	**5.278**	8.107	10.637
2	5.779	6.512	**5.620**	8.707	11.063
3	5.746	6.537	**5.582**	8.991	11.540
4	5.774	6.547	**5.602**	9.259	12.017
5	5.846	6.553	**5.671**	9.578	12.620
6	5.889	6.563	**5.718**	9.783	13.057
7	5.923	6.610	**5.753**	9.958	13.446
8	5.998	6.698	**5.829**	10.134	13.847
9	6.048	6.752	**5.878**	10.239	14.095
10	6.085	6.758	**5.914**	10.290	14.259
11	6.096	6.745	**5.922**	10.288	14.306
12	6.091	6.726	**5.911**	10.265	14.305

In addition to the factor-based LL and HBY models, we include further comparisons of forecast performance using VAR models. The specifications of both the single-population VAR and the multi-population GVAR models are provided in Appendix D. Table 1 presents the mean MAPE of out-of-sample forecasts over the forecast horizon h from 1 to 12 months (scaled by a factor of 100). For each horizon, the lowest mean MAPE, indicating the best-performing model, is highlighted in bold. The results indicate that, based on the STMF dataset, the proposed GBLL model outperforms the two benchmark factor models and the two VAR models across all forecast horizons. This superior performance is achieved by the iterative structure of the gradient boosting framework, which enables the model to sequentially identify and incorporate multiple layers of complex patterns. The comparatively weaker performance of VAR and GVAR models can be attributed to their generic time series nature and the differencing required to ensure stationarity. In contrast to factor models (LL, HBY, and GBLL), which are specifically designed to capture and forecast long-run mortality trends, VAR-type models are typically estimated on differenced series and are therefore better suited to analyzing short-run dynamics, such as spillover effects and shock transmission as characterized by impulse response functions. Moreover, the relatively short time span of mortality data exacerbates the curse of dimensionality in VAR and GVAR

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

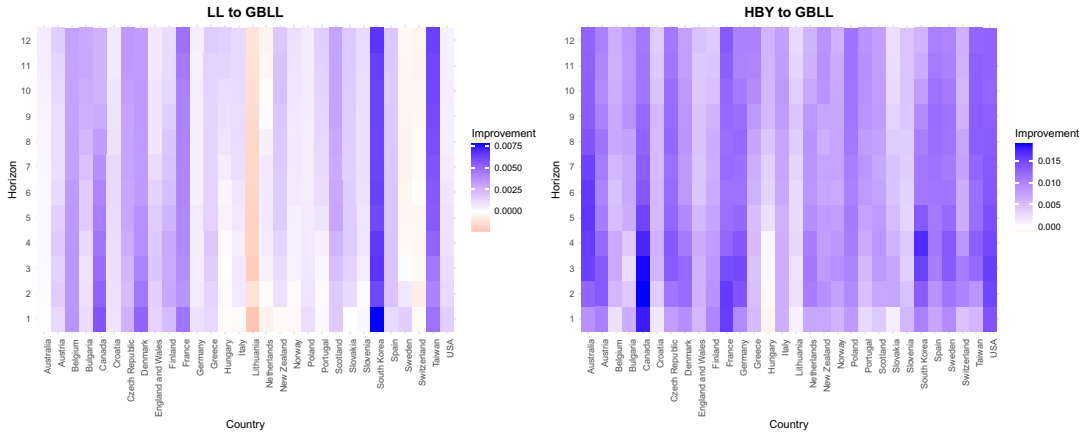


Figure 5. *Improvement in MAPE from LL to GBLL (left) and from HBY to GBLL (right).*

models, leading to increased estimation uncertainty. Accordingly, subsequent detailed comparisons of model performance will focus only on the class of factor-based models.

To gain a better understanding of the enhanced forecast accuracy achieved by the proposed GBLL model, Figure 5 presents the improvement in MAPE across countries and forecast horizons. Compared to the standard LL model, the GBLL model enhances forecast accuracy. Substantial gains are observed in countries such as South Korea and Taiwan, while Lithuania is slightly better under the LL model. Relative to the HBY model, GBLL again shows superior performance in nearly all cases, except for Hungary in the first two forecast months. Overall, the GBLL model achieves the highest accuracy across the three models evaluated. Additional visualizations of out-of-sample forecasts for the GBLL model and the benchmark models for selected countries are provided in Section C.2 of the Supplementary Material. Interestingly, the LL model outperforms the HBY model despite its simpler structure. A likely explanation is that the use of only four age groups, as opposed to single-year-of-age data for which the HBY model is designed, reduces the number of principal components needed to explain the variation in data. Consequently, the additional principal components may contribute little beyond the first. Therefore, given the complexity and heterogeneity of mortality data across 30 countries, simpler or progressive approaches such as the gradient boosting framework, may be better suited to capturing shared and country-specific patterns than more complex one-step models. We have conducted an additional analysis to evaluate forecast performance across age groups, with a detailed discussion provided in Section D of the Supplementary Material. Overall, the proposed GBLL model demonstrates superior forecast accuracy for most age groups, particularly among younger cohorts, while the simpler LL model performs comparably for the oldest age group, likely due to its smoother mortality patterns.

In addition to evaluating point forecast accuracy, we assess each model's ability to quantify predictive uncertainty via marginal calibration of 95% prediction intervals (Czado *et al*., 2009). For each model, 100 forecast paths for κ_t are simulated, and prediction intervals for $\hat{m}^j_{x,t}$ are constructed from the empirical 2.5th and 97.5th percentiles, denoted $\hat{p}^j_{2.5,x,t}$ and $\hat{p}^j_{97.5,x,t}$. Marginal calibration evaluates whether the nominal probability of the interval aligns with the empirical coverage, that is, the proportion of observed outcomes that fall within the interval. Formally, for week $h_w \in \{1, ..., 52\}$, expanding window $r \in \{0, ..., 9\}$, age group $x \in \{1, ..., 4\}$, and country $j \in \{1, ..., 30\}$, the marginal calibration $f_{[\hat{p}^j_{2.5,x,t}, \hat{p}^j_{97.5,x,t}]}$ is given by:

$$f_{[\hat{p}^j_{2.5,x,t}, \hat{p}^j_{97.5,x,t}]} = \frac{1}{52 \times 10 \times 4 \times 30} \sum_{h_w=1}^{52} \sum_{r=0}^{9} \sum_{x=1}^{4} \sum_{j=1}^{30} \mathbb{1}\left(\hat{p}^j_{2.5,x,169+h_w+H_{(r)}} \leq \tilde{m}^j_{x,169+h_w+H_{(r)}} \leq \hat{p}^j_{97.5,x,169+h_w+H_{(r)}}\right),$$

where $\mathbb{1}$ denotes the indicator function. Mean empirical marginal calibration rates show that the LL, HBY, and GBLL models achieve 90%, 94%, and 96% coverage, respectively. These results indicate that the GBLL model is the most properly calibrated, producing prediction intervals that reliably capture

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

realized mortality rates and providing improved uncertainty quantification alongside superior point forecast performance.

5. An experiment on clustering

Clustering is an unsupervised learning technique that partitions data into distinct groups based on similarity, aiming to ensure that observations within the same cluster are more alike than those in different clusters (Rokach and Maimon, 2005). Clustering can offer an alternative way to group data based on characteristics such as geographic, socioeconomic, or seasonal factors and can therefore be used to assess the sensitivity of the model to the underlying data structure. Evaluating model performance across these clusters provides insight into the extent to which the model responds to structural heterogeneity. Limited variation in performance indicates stability, whereas substantial discrepancies may signal sensitivity to differences across groups. Ensuring homogeneity within clusters enhances coherence in subpopulation forecasts, though achieving this is more challenging in cross-country comparisons due to diverse environmental, social, and political backgrounds.

Clustering techniques are increasingly employed in mortality rate modeling to enhance the interpretability and accuracy of forecasts. Hatzopoulos and Haberman (2013) aim to construct common mortality trends across multiple populations with similar mortality experiences, enabling more coherent forecasts. They apply fuzzy c-means clustering to the main time effects, which serve as a summary of mortality dynamics. Similarly, Léger and Mazzuco (2021) adopt a functional data analysis framework to mix countries and ages in the modeling process. Tsai and Cheng (2021) investigate age-specific mortality trends and propose that grouping ages with similar temporal patterns can improve forecast performance.

5.1. Setup

Clustering typically involves four key components: the data input, clustering algorithm, distance measure, and the optimal number of clusters (Yin *et al*., 2024). In this paper, we first extract the time trend component κ_t for each country using the Lee–Carter model. We then consider three different variations of the raw time series as the data input to investigate population similarity, while keeping the remaining three clustering components fixed.

Clustering algorithms can be broadly classified into several categories, including partitional, hierarchical, density-based, grid-based, and model-based approaches (Ghosal *et al*., 2020). In this study, we adopt the K-means algorithm, a representative method within the partitional clustering framework, which minimizes within-cluster sum of squares through an iterative process of centroid updating and point re-assignment (Lloyd, 1982). The accompanying method to find the optimal number of clusters is the elbow method, which plots the inertia (within-cluster sum of squares) against a range of values for the number of clusters (Thorndike, 1953). Inertia is calculated as:

$$\text{Inertia} = \sum_{q=1}^{Q} \sum_{\mathbf{x} \in C_q} (\mathbf{x} - \boldsymbol{\mu}_q)'(\mathbf{x} - \boldsymbol{\mu}_q),$$

where Q is the total number of clusters, $\mathbf{x}$ is an observation for each country assigned to cluster C_q, and $\boldsymbol{\mu}_q$ is the centroid of cluster C_q. The final optimal number is identified at the "elbow" point of the curve, where the rate of decrease in inertia slows significantly, indicating diminishing returns in clustering quality with additional clusters. Finally, the distance measure used to define similarity between data points is the Euclidean distance metric. It calculates the straight-line distance between two points in an n-dimensional space.

5.1.1. Method 1: Clustering based on raw time trend series

We begin by analyzing the raw time series of the country-specific κ_t as the data input. This approach identifies three distinct clusters, as detailed in Table 2. Figure 6 presents the mean value of the time trends for each cluster. Notably, Cluster 3, which comprises only two countries from the Southern Hemisphere,

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Table 2. *List of countries in each cluster under Method 1.*

Cluster	Number of countries	Country name
1	11	Canada, Denmark, England and Wales, Finland, Netherlands, Norway, Scotland, South Korea, Sweden, Taiwan, USA
2	17	Austria, Belgium, Bulgaria, Croatia, Czech Republic, France, Germany, Greece, Hungary, Italy, Lithuania, Poland, Portugal, Slovakia, Slovenia, Spain, Switzerland
3	2	Australia, New Zealand

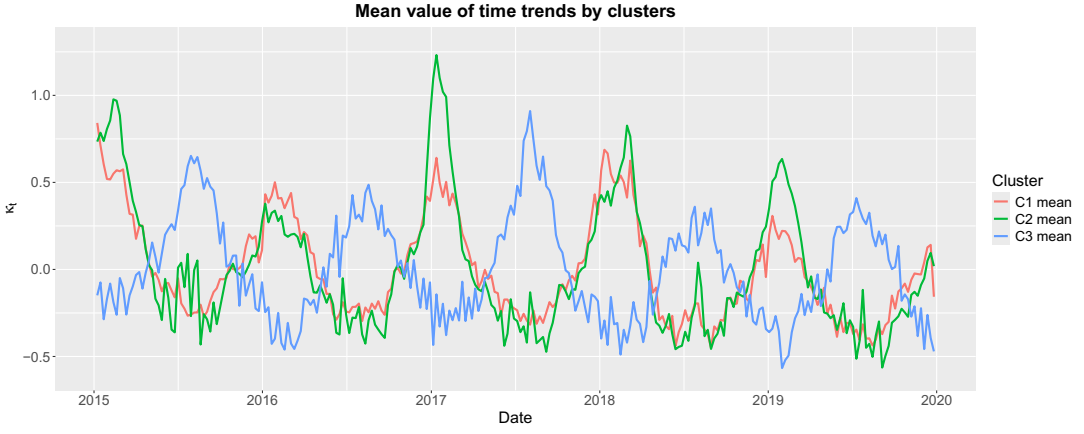


Figure 6. *Mean value of the time trends in each cluster.*

exhibits seasonal patterns that are opposite to those of the other clusters. Cluster 2 consists exclusively of European countries and displays more pronounced seasonal peaks relative to Cluster 1, highlighting its higher mortality rates in winter.

It is noteworthy that Cluster 2 exhibits a pronounced spike during the winter of 2017, which can be attributed to an extreme cold-weather event, affecting most of Europe in January. Temperatures remain significantly below seasonal averages for several days, particularly impacting southern and southeastern Europe. Greece, for instance, records its highest January mortality rate in 2017 compared to any other January between 2013 and 2018, with individuals aged over 80 years being most affected (Kostopoulou, 2023). Italy also experiences abnormal snowfall due to cold air moving across the relatively warmer Adriatic Sea, picking up moisture and generating intense precipitation (Murtas and Russo, 2019). Demirtaş (2022) further investigates the broader effects of severe snowstorms and cold temperatures across southeastern Europe. A similar pattern is observed in 2019, with another cold wave, triggered by a Sudden Stratospheric Warming (SSW) event. It results in widespread snowfall and freezing temperatures across central and southern Europe, contributing to increased mortality and significant societal disruptions (Knight *et al*., 2021).

5.1.2. Method 2: Clustering based on estimated trend slopes

Building on the previous analysis of the raw time series, we now focus on the trend reflecting mortality improvement over the years. We first apply Seasonal and Trend decomposition using LOESS (STL) to the country-specific κ_t series (Cleveland *et al*., 1990). This method uses locally estimated scatterplot smoothing (LOESS) to decompose each time series into three distinct components: a long-term trend,

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Table 3. *List of countries in each cluster under Method 2.*

Cluster	Number of countries	Country name
1	5	Croatia, Lithuania, Slovakia, South Korea, Sweden
2	12	Australia, Austria, Belgium, Czech Republic, Hungary, Italy, Norway, Scotland, Slovenia, Spain, Switzerland, Taiwan
3	13	Bulgaria, Canada, Denmark, England and Wales, Finland, France, Germany, Greece, Netherlands, New Zealand, Poland, Portugal, USA

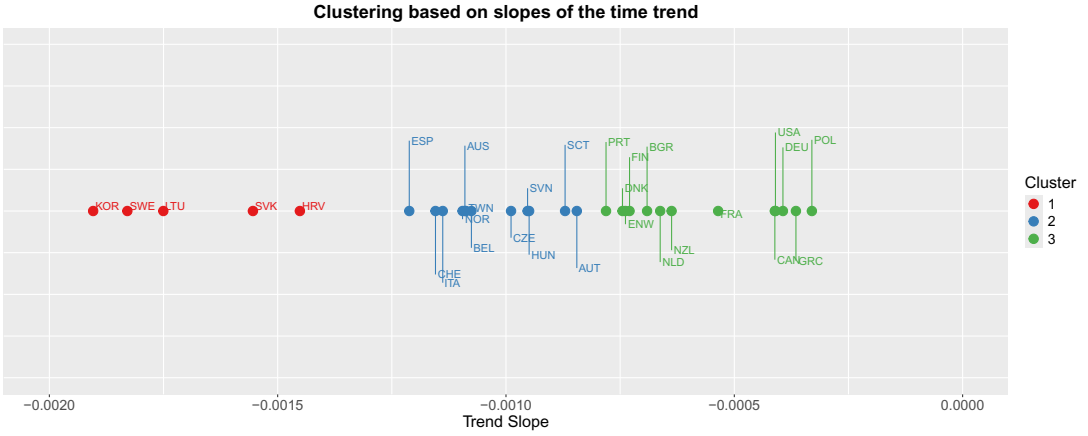


Figure 7. *Trend slopes by clusters.*

a recurring seasonal pattern, and a residual representing irregular fluctuations. To classify countries based on their trend dynamics, we compute the slope of each decomposed long-term trend using linear regression. This procedure identifies three distinct clusters, as detailed in Table 3. Figure 7 displays the distribution of trend slopes by clusters. Countries in Cluster 1 exhibit the most substantial mortality improvement, whereas Cluster 2 countries have moderate improvement and Cluster 3 countries are more stable over the five-year period.

All countries have negative trend slopes, indicating the mortality improvement across five years. The rapid mortality improvement observed in Cluster 1 may be attributed to factors such as healthcare system reforms and advances in medical technology and treatment (see e.g., OECD, 2021; Pekarcikova, 2024). Given that countries like Croatia, Lithuania, and Slovakia have historically exhibited relatively higher mortality rates, these advancements suggest a particularly substantial improvement in health outcomes for these populations, resulting in higher trend slopes in absolute values. In Clusters 2 and 3, some geographical patterns are evident, including the grouping of Canada and the USA, as well as several neighboring European countries. These similarities likely reflect shared characteristics such as public health infrastructure, climatic conditions, and socioeconomic contexts, which may contribute to comparable rates of mortality improvement over the period from 2015 to 2019.

5.1.3. Method 3: Clustering based on estimated trend slopes and seasonal strength

Finally, building upon the trend slopes employed in Method 2, we additionally incorporate the measure of seasonal strength derived from the aforementioned STL decomposition. These two features, trend and seasonality, are essential for characterizing the underlying dynamics of mortality evolution. To ensure comparability between the two variables, min-max scaling is applied. The resulting three clusters are

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Table 4. *List of countries in each cluster under Method 3.*

Cluster	Number of countries	Country name
1	5	Croatia, Lithuania, Slovakia, South Korea, Sweden
2	15	Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, Germany, Greece, Hungary, Norway, Poland, Scotland, Slovenia, Switzerland, Taiwan
3	10	Australia, Canada, England and Wales, France, Italy, Netherlands, New Zealand, Portugal, Spain, USA

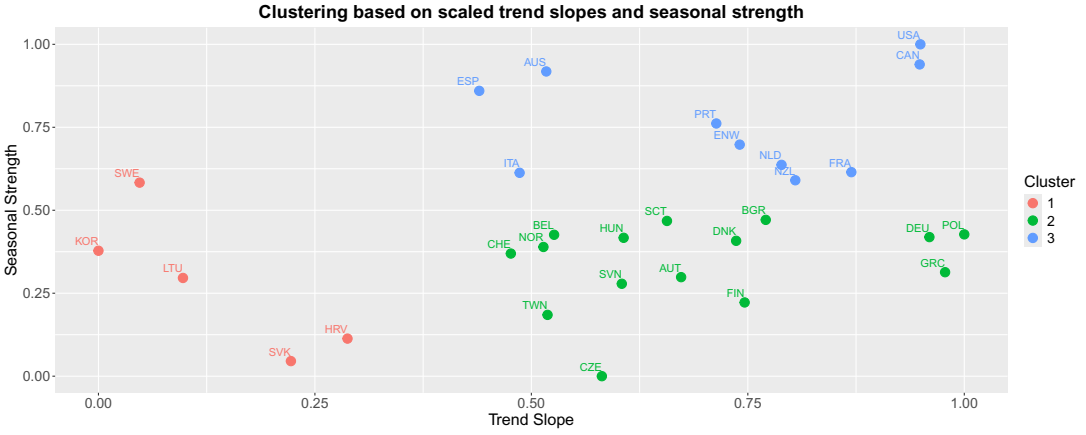


Figure 8. *Clusters based on min-max scaled trend slopes and seasonal strength.*

detailed in Table 4. Figure 8 presents a scatterplot, illustrating the clustering outcome. Cluster 1 is mainly determined by the trend slopes, where the distinction between Cluster 2 and Cluster 3 is primarily driven by differences in seasonal strength. Cluster 1 comprises countries exhibiting a higher rate of mortality improvement, as indicated by more negative trend slopes. Cluster 2 includes countries with relatively stable mortality patterns and weaker seasonality. Cluster 3 consists of countries with similar rates of mortality improvement to Cluster 2, exhibiting stronger seasonal variations.

We investigate the countries in Cluster 3 with respect to their pronounced seasonal mortality patterns. These stronger seasonal fluctuations are potentially attributable to cold-weather effects, which contribute to elevated mortality rates during winter months. Furthermore, all countries in Cluster 3 have aging populations, wherein older adults exhibit increased vulnerability to seasonal health variations (Bonnet *et al*., 2021). Falagas *et al*. (2009) analyze monthly mortality data to examine seasonal patterns across Mediterranean countries, including France, Italy, Spain, and Portugal, as well as Australia, Canada, and the USA. Their findings indicate that the lowest average daily mortality occurs in September for countries in the Northern Hemisphere and in March for Australia and New Zealand. These more pronounced seasonal variations are further attributed to environmental factors, such as temperature and daylight duration.

5.2. Forecast performance with expanding windows

To mitigate randomness in the sampling procedure, we again employ a 10-fold expanding window approach. Table 5 summarizes the forecast performance across the clustering methods described above in Section 5.1. Among the three clustering approaches, Method 1, which uses raw time trend series, yields the poorest performance, likely due to its higher dimensionality and associated noise. In contrast,

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Table 5. *Mean MAPE of out-of-sample forecasts across 30 countries with clustering (*$\times 100$ *scale).*

	Method 1			Method 2			Method 3			Whole
h	LL	HBY	GBLL	LL	HBY	GBLL	LL	HBY	GBLL	GBLL
1	5.459	6.246	**5.309**	5.357	5.924	**5.199**	5.453	5.867	**5.299**	5.278
2	5.734	6.458	**5.586**	5.647	6.220	**5.481**	5.746	6.057	**5.577**	5.620
3	5.774	6.476	**5.625**	5.715	6.242	**5.540**	5.783	6.106	**5.605**	5.582
4	5.818	6.504	**5.658**	5.766	6.251	**5.588**	5.841	6.136	**5.659**	5.602
5	5.908	6.500	**5.741**	5.816	6.322	**5.632**	5.902	6.151	**5.713**	5.671
6	5.954	6.553	**5.785**	5.861	6.401	**5.680**	5.940	6.176	**5.747**	5.718
7	5.997	6.576	**5.828**	5.882	6.455	**5.703**	5.963	6.183	**5.772**	5.753
8	6.050	6.640	**5.879**	5.924	6.494	**5.747**	6.021	6.230	**5.832**	5.829
9	6.091	6.689	**5.917**	5.955	6.522	**5.779**	6.061	6.281	**5.875**	5.878
10	6.112	6.708	**5.935**	5.968	6.536	**5.790**	6.077	6.308	**5.887**	5.914
11	6.110	6.689	**5.924**	5.959	6.503	**5.777**	6.072	6.300	**5.876**	5.922
12	6.104	6.665	**5.913**	5.962	6.493	**5.778**	6.077	6.311	**5.875**	5.911

Method 2, which relies solely on the slopes of country-specific time trends, achieves the best forecast accuracy. This result suggests that utilizing a more distinct and parsimonious clustering criterion may facilitate the grouping of structurally similar countries, thereby enhancing predictive performance. In this experiment, regardless of the clustering strategy employed, based on the available data, our proposed GBLL model outperforms both the LL and HBY models, with the latter exhibiting the weakest forecast accuracy.

To further assess the impact of clustering, we compare the results with the case without clustering, as presented in Table 1. The overall performance remains comparable; however, Method 2 is the only clustering strategy that improves forecast accuracy across all models and forecast horizons relative to the baseline case without clustering. The varying performance may be attributed to the heterogeneity in the rate of mortality improvement and the degree of seasonality across individual countries, particularly given that weekly mortality data tend to exhibit higher-frequency fluctuations and varying levels of noise. These findings suggest that, within this dataset of 30 countries, trend slopes serve as the most effective factor for grouping countries with similar mortality experience together. Additionally, we include a reference column reporting the forecast MAPE of the GBLL model without any clustering. Notably, this unclustered version of the GBLL model still outperforms all variants of the LL and HBY models under each of the three clustering approaches and across all forecast horizons. Based on the available data, the empirical evidence highlights that the GBLL framework achieves superior forecast performance without the need for clustering, surpassing baseline models that rely on clustering to enhance coherence and forecast accuracy.

Focusing on the HBY model, which emphasizes the need for grouping countries with similar characteristics to enhance forecast performance, the results lend support to this premise. All three clustering methods improve forecast accuracy compared to the unclustered baseline, with the exception of the first month under Method 1. However, identifying an optimal clustering strategy that yields substantial improvement on all models over the baseline case without clustering remains challenging. This may be attributed to the limited span of data with only five years, during which there is minimal divergence in mortality trends, resulting in relatively homogeneous raw time trend series across countries. These observations underscore the importance of a modeling framework that is both adaptive and insensitive, regardless of clustering choices. The LL model demonstrates relatively similar performance across various clustering methods and may be regarded as a stable, though less flexible, approach. However, this stability comes at the cost of lower predictive accuracy. Conversely, the HBY model exhibits substantial sensitivity to country groupings and continues to deliver the weakest forecast performance overall. In

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

contrast, the proposed GBLL model distinguishes itself by maintaining high forecast accuracy across all clustering strategies. Its performance is largely unaffected by the social, economic, or environmental heterogeneity of the countries, highlighting its insensitivity and superior forecasting capability.

6. Conclusion

This paper proposes a multi-population mortality model, the GBLL model that integrates gradient boosting techniques to iteratively capture the heterogeneous seasonal patterns among all age groups and countries. Specifically, we apply this model to weekly mortality rates across 30 countries for the period 2015–2019. Empirical findings demonstrate that the GBLL model outperforms the benchmark models, including the original LL model and the HBY model of order 6, in both in-sample fitting and out-of-sample forecasts. In addition, we conduct an experiment on clustering to group countries with similar mortality dynamics. Empirical results indicate that the performance varies depending on the clustering method employed, and clustering does not universally lead to better outcomes. Nevertheless, the proposed GBLL model exhibits insensitivity to the choice of country groupings, maintaining stable and superior forecast accuracy across different clustering arrangements. This adaptability provides evidence of the flexibility of the GBLL model in handling heterogeneous data without the need for extensive preprocessing or manual adjustment.

The proposed model provides several promising directions for future research. First, the reciprocal transformation of mortality rates for Southern Hemisphere countries may be leveraged in mortality risk hedging strategies, given the inverse seasonal patterns and the offset in the timing of mortality peaks and troughs across hemispheres. Second, the gradient boosting framework may be extended to other stochastic mortality models, such as the Cairns–Blake–Dowd model (Cairns *et al*., 2006) and its variants. Further extensions could incorporate finer age stratification, such as 5-year or 10-year age intervals, as well as additional dimensions, including sex, causes of death, or regional subgroups, to provide more insights into granular dynamics and improve the precision of mortality modeling. However, this would depend on the availability of such detailed data. Finally, the proposed framework is well suited to within-year mortality forecasting, enabling early-season assessments of expected winter mortality and facilitating timely monitoring of unusually elevated mortality levels linked to extreme events, including influenza seasons and pandemics.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/asb.2026.10104

Acknowledgements. The authors thank Rob J. Hyndman of Monash University, Australia, Li Li of the University of Science and Technology Beijing, China, and Hanlin Shang of Macquarie University, Australia, for their helpful comments on an earlier version of this paper. Ziting Miao is supported by the Australian Government Research Training Program Scholarship and the Henry Buck Scholarship from the University of Melbourne.

Competing interests. The authors declare no conflicts of interest or competing interests in this paper.

References

Andreopoulos P., Bersimis F.G. and Tragaki A. (2022) A different approach to current developments in the twenty-first century–grouping European countries in terms of mortality. In *Quantitative Methods in Demography: Methods and Related Applications in the Covid-19 Era*. Springer, 373–385.

Arik A., Klein A. and Li H. (2026) On measuring COVID-19 excess mortality: Insights and challenges. *Insurance: Mathematics and Economics* **126**, 103199.

Asgari P., Miri M.M. and Asgari F. (2022) The comparison of selected machine learning techniques and correlation matrix in ICU mortality risk prediction. *Informatics in Medicine Unlocked* **31**, 100995.

Bjerre D.S. (2022) Tree-based machine learning methods for modeling and forecasting mortality. *ASTIN Bulletin: The Journal of the IAA* **52**(3), 765–787.

Blanes-Selva V., Ruiz-Garca V., Tortajada S., Bened J.-M., Valdivieso B. and Garca-Gómez J.M. (2021) Design of 1-year mortality forecast at hospital admission: A machine learning approach. *Health Informatics Journal* **27**(1), 1–13.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Bonnet C., Cambois E., Fontaine R., Dutreuilh C. and van Hoorn Alkema B. (2021) Population ageing in high-longevity countries: Demographic dynamics and socio-economic challenges. *Population*, **76**(2), 217–310.
Boonen T.J. and Chen Y. (2026) Low-rank tensor autoregressive models for mortality modelling. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 1–31. https://doi.org/10.1093/jrsssa/qnag048.
Brouhns N., Denuit M. and Vermunt J.K. (2002) A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* **31**(3), 373–393.
Bühlmann P. and Hothorn T. (2007) Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* **22**(4), 477–505.
Cairns A.J., Blake D. and Dowd K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, **73**(4), 687–718.
Chen T. and Guestrin C. (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
Cleveland R.B., Cleveland W.S., McRae J.E. and Terpenning I. (1990) STL: A seasonal-trend decomposition. *Journal of Official Statistics* 6(1), 3–73.
Czado C., Gneiting T. and Held L. (2009) Predictive model assessment for count data. *Biometrics* **65**(4), 1254–1261.
Danesi I.L., Haberman S. and Millossovich P. (2015) Forecasting mortality in subpopulations using Lee–Carter type models: A comparison. *Insurance: Mathematics and Economics* **62**, 151–161.
Debon A., Haberman S. and Piscopo G. (2024) Multipopulation mortality analysis: Bringing out the unobservable with latent clustering. *Quality & Quantity* **58**(6), 5107–5123.
Demirtas M. (2022) The anomalously cold January 2017 in the South-Eastern Europe in a warming climate. *International Journal of Climatology* **42**(11), 6018–6026.
Deprez P., Shevchenko P.V. and Wüthrich M.V. (2017) Machine learning techniques for mortality modeling. *European Actuarial Journal* **7**, 337–352.
Dimai M. (2025) Multi-population mortality modeling with economic, environmental and lifestyle variables. *Quality & Quantity* **59**, 153–205.
Di Mauro F. and Pesaran M.H. (2013) *The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis*. Oxford: Oxford University Press.
Falagas M.E., Karageorgopoulos D.E., Moraitis L.I., Vouloumanou E.K., Roussos N., Peppas G. and Rafailidis P.I. (2009) Seasonality of mortality: The September phenomenon in Mediterranean countries. *Canadian Medical Association Journal* **181**(8), 484–486.
Friedman J.H. (2001) Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **29**(5), 1189–1232.
Ghosal A., Nandy A., Das A.K., Goswami S. and Panday M. (2020). A short review on different clustering techniques and their applications. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*, 69–83.
Girosi F. and King G. (2007) Understanding the Lee–Carter mortality forecasting method. *Technical report, RAND Corporation.*
Guibert Q., Lopez O. and Piette P. (2019) Forecasting mortality rate improvements with a high-dimensional VAR. *Insurance: Mathematics and Economics* **88**, 255–272.
Hanebeck A., Li H. and Czado C. (2025) Dependence modelling across major causes of death via time-varying copula state space models. *Journal of the Royal Statistical Society Series C: Applied Statistics* **75**, 407–430.
Hans N., Klein N., Faschingbauer F., Schneider M. and Mayr A. (2023) Boosting distributional copula regression. *Biometrics* **79**(3), 2298–2310.
Hatemi-J A. (2003) A new method to choose optimal lag order in stable and unstable VAR models. *Applied Economics Letters* **10**(3), 135–137.
Hatzopoulos P. and Haberman S. (2013) Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics* **52**(2), 320–337.
Huang C., Chu C., Wang X. and Barnett A.G. (2015) Unusually cold and dry winters increase mortality in Australia. *Environmental Research* **136**, 1–7.
Hyndman R.J., Booth H. and Yasmeen F. (2013) Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography* **50**(1), 261–283.
Hyndman R.J. and Ullah M.S. (2007) Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**(10), 4942–4956.
Islam N., Shkolnikov V.M., Acosta R.J., Klimkin I., Kawachi I., Irizarry R.A., Alicandro G., Khunti K., Yates T., Jdanov D.A., White M., Lewington S. and Lacey B. (2021) Excess deaths associated with covid-19 pandemic in 2020: Age and sex disaggregated time series analysis in 29 high income countries. *BMJ* 373, n1137.
Ivanov V. and Kilian L. (2005) A practitioner's guide to lag order selection for VAR impulse response analysis. *Studies in Nonlinear Dynamics & Econometrics* **9**(1), Article 2.
Jacobsen R., Keiding N. and Lynge E. (2002) Long term mortality trends behind low life expectancy of Danish women. *Journal of Epidemiology & Community Health* **56**(3), 205–208.
Jdanov D.A., Galarza A.A., Shkolnikov V.M., Jasilionis D., Németh L., Leon D.A., Boe C. and Barbieri M. (2021) The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. *Scientific Data* **8**(1), 235.
Karlinsky A. and Kobak D. (2021) The world mortality dataset: Tracking excess mortality across countries during the covid-19 pandemic. *eLife* **10**, e69336.
Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q. and Liu T. (2017) Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 3149–3157.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Knight J., Scaife A., Bett P.E., Collier T., Dunstone N., Gordon M., Hardiman S., Hermanson L., Ineson S., Kay G., McLean P., Pilling C., Smith D., Stringer N., Thornton H. and Walker B. (2021) Predictability of european winters 2017/2018 and 2018/2019: Contrasting influences from the tropics and stratosphere. *Atmospheric Science Letters* 22(1), e1009.

Kontis V., Bennett J.E., Rashid T., Parks R.M., Pearson-Stuttard J., Guillot M., Asaria P., Zhou B., Battaglini M., Corsetti G., McKee M., Di Cesare M., Mathers C.D. and Ezzati M. (2020) Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature Medicine* 26(12), 1919–1928.

Kostopoulou E. (2023) Analysis of the January 2017 cold spell in Greece and its implications on human health. *Environmental Sciences Proceedings* **26**(1), 195.

Lam K.K. and Wang B. (2023) Multipopulation mortality modelling and forecasting: The weighted multivariate functional principal component approaches. *Journal of Applied Statistics* **50**(15), 3177–3198.

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

Léger A.E. and Mazzuco S. (2021) What can we learn from the functional clustering of mortality data? An application to the human mortality database. *European Journal of Population* **37**, 769–798.

Léger A.E. and Rizzi S. (2024) Month-to-month all-cause mortality forecasting: A method allowing for changes in seasonal patterns. *American Journal of Epidemiology* **193**(6), 898–907.

Levantesi S. and Pizzorusso V. (2019) Application of machine learning to mortality modeling and forecasting. *Risks* **7**(1), 26.

Li H. and Chen H. (2024) Hierarchical mortality forecasting with EVT tails: An application to solvency capital requirement. *International Journal of Forecasting* **40**(2), 549–563.

Li H., Li H., Lu Y. and Panagiotelis A. (2019) A forecast reconciliation approach to cause-of-death mortality modeling. *Insurance: Mathematics and Economics* **86**, 122–133.

Li H. and Tang Q. (2022) Joint extremes in temperature and mortality: A bivariate POT approach. *North American Actuarial Journal* **26**(1), 43–63.

Li L., Li H. and Panagiotelis A. (2025) Boosting domain-specific models with shrinkage: An application in mortality forecasting. *International Journal of Forecasting* **41**(1), 191–207.

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

Ljung G.M. and Box G.E. (1978) On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297–303.

Lloyd S. (1982) Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28**(2), 129–137.

McNown R. and Rogers A. (1992) Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting* **8**(3), 413–432.

Murtas R. and Russo A.G. (2019) Effects of pollution, low temperature and influenza syndrome on the excess mortality risk in winter 2016–2017. *BMC Public Health* **19**, 1–9.

Nepomuceno, M.R., Klimkin I., Jdanov D.A., Alustiza-Galarza A. and Shkolnikov, V.M. (2022). Sensitivity analysis of excess mortality due to the COVID-19 pandemic. *Population and Development Review* **48**(2), 279–302.

Neshat M., Phipps M., Jha N., Khojasteh D., Tong M. and Gandomi A.H. (2026). Effective predictive modelling for emergency department visits and evaluating exogenous variables impact: Using explainable meta-learning gradient boosting. *ACM Transactions on Computing for Healthcare* **7**(1), 1–33.

Neves C., Fernandes C. and Hoeltgebaum H. (2017) Five different distributions for the Lee–Carter model of mortality forecasting: A comparison using GAS models. *Insurance: Mathematics and Economics* **75**, 48–57.

Nocedal J. and Wright S.J. (2006) *Numerical Optimization*. New York: Springer.

OECD (2021) *State of Health in the EU Croatia: Country Health Profile 2021*. Paris: OECD Publishing.

Osmond C. (1985) Using age, period and cohort models to estimate future mortality rates. *International Journal of Epidemiology* **14**(1), 124–129.

Pekarcikova J. (2024) Cancer mortality attributable to air pollution in Slovakia. *European Journal of Public Health* **34**(Supplement_3), ckae144.1429.

Pesaran M.H., Schuermann T. and Weiner S.M. (2004) Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics* **22**(2), 129–162.

Qiao Y., Wang C. and Zhu W. (2024) Machine learning in long-term mortality forecasting. *The Geneva Papers on Risk and Insurance – Issues and Practice* **49**(2), 340–362.

Qiu W., Chen H., Dincer A.B., Lundberg S., Kaeberlein M. and Lee S.-I. (2022) Interpretable machine learning prediction of all-cause mortality. *Communications Medicine* **2**(1), 125.

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

Rokach L. and Maimon O. (2005) Clustering methods. In *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer, 321–352.

Russolillo M., Giordano G. and Haberman S. (2011). Extending the Lee–Carter model: A three-way decomposition. *Scandinavian Actuarial Journal* **2011**(2), 96–117.

Serfling R.E. (1963) Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports* **78**(6), 494.

Shapovalov V., Landsman Z. and Makov, U. (2019) Bayesian log-bilinear mortality projection with a random walk with drift. Available at SSRN 3375920.

Sims C.A. (1980) Macroeconomics and reality. *Econometrica* **48**(1), 1–48.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

STMF (2021) *Short-Term Mortality Fluctuations (STMF) Data Series*. *Human Mortality Database*. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at https://www.mortality.org/Data/STMF (accessed 18 October 2024).

Thorndike R.L. (1953) Who belongs in the family? *Psychometrika* **18**(4), 267–276.

Tsai C.C.-L. and Cheng E.S. (2021) Incorporating statistical clustering methods into mortality models to improve forecasting performances. *Insurance: Mathematics and Economics* **99**, 42–62.

Tutz G. and Binder H. (2006) Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics* **62**(4), 961–971.

Vanella P., Basellini U. and Lange B. (2021) Assessing excess mortality in times of pandemics based on principal component analysis of weekly mortality data–the case of COVID-19. *Genus* **77**, 1–36.

Wald A. (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54**(3), 426–482.

Yin H., Aryani A., Petrie S., Nambissan A., Astudillo A. and Cao S. (2024) A rapid review of clustering algorithms. arXiv preprint arXiv:2401.07389.

Appendices

A. Mortality data for 30 countries

While Figure A1 does not clearly reveal country-specific mortality trends, the overarching patterns discussed in Section 2.2 remain similar.



Figure A1. *2015–2019 mortality rate for 30 countries by age groups.*

B. The Lee–Carter model

The Lee–Carter model (Lee and Carter, 1992) is the first single-population stochastic mortality model and serves as the foundation for the estimation of the LL model. This model separates the historical mortality rate into age-specific and time-varying components. For age group $x \in \{1, ..., N\}$ and time $t \in \{1, ..., T\}$, the method is formulated as follows:

$$\log(m_{x,t}) = a_x + b_x\kappa_t + \varepsilon_{x,t},$$

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

Algorithm B.1 Estimation of the Lee–Carter model

Input: $\boldsymbol{Y}$ (a $T \times N$ matrix)
Output: $\hat{y}_{x,t}$ for $x \in \{1, ..., N\}$, $t \in \{1, ..., T\}$, $\hat{\boldsymbol{\kappa}}_t = (\hat{\kappa}_1, \hat{\kappa}_2, ..., \hat{\kappa}_T)$, $\hat{\boldsymbol{a}}_x = (\hat{a}_1, \hat{a}_2, ..., \hat{a}_N)$
and $\hat{\boldsymbol{b}}_x = (\hat{b}_1, \hat{b}_2, ..., \hat{b}_N)$
Procedure: LC($\boldsymbol{Y}$)
1: $\hat{a}_x \leftarrow \frac{1}{T}\sum_{t=1}^{T} \log(\tilde{m}_{x,t}), \forall x$
2: $\bar{y}_{x,t} \leftarrow \log(\tilde{m}_{x,t}) - \hat{a}_x$
3: Form $\bar{\boldsymbol{Y}} = (\bar{\boldsymbol{y}}_{1,t}, \bar{\boldsymbol{y}}_{2,t}, ..., \bar{\boldsymbol{y}}_{N,t})$
4: Do the Singular Value Decomposition (SVD) on $\bar{\boldsymbol{Y}}$
5: $\hat{\boldsymbol{\kappa}}_t \leftarrow d \times \boldsymbol{u} \times \sum_{x=1}^{N} v_x$, where d is the first singular value of $\bar{\boldsymbol{Y}}$, $\boldsymbol{u}$ is the first column of the left singular vectors of $\bar{\boldsymbol{Y}}$ and $\boldsymbol{v}$ is the first column of the right singular vectors of $\bar{\boldsymbol{Y}}$
6: $\hat{\boldsymbol{b}}_x \leftarrow \frac{\boldsymbol{v}}{\sum_{x=1}^{N} v_x}, \forall x$
7: $\hat{y}_{x,t} = \log(\hat{m}_{x,t}) \leftarrow \hat{a}_x + \hat{b}_x\hat{\kappa}_t, \forall x, t$

where $m_{x,t}$ is the mortality rate for age group x at time t, a_x represents the age-specific mean of log mortality rate over time, κ_t is the mortality improvement over time shared by all age groups, b_x is the corresponding age-specific loading, and $\varepsilon_{x,t}$ is an error term. The model contains two constraints to ensure the uniqueness of parameter estimates:

$$\sum_{x=1}^{N} b_x = 1 \text{ and } \sum_{t=1}^{T} \kappa_t = 0.$$

The estimation of the Lee–Carter model is via SVD. First let $\boldsymbol{Y}$ denote a $T \times N$ (time $\times$ age group) matrix with $\log(\tilde{m}_{x,t})$ in row t and column x. The age-specific intercept $\hat{a}_x$ is calculated by taking the sample mean of the logarithmic mortality rate for each age group x over T periods. Then, subtract the age-specific intercept from each row of $\boldsymbol{Y}$ and denote the resulting matrix as $\bar{\boldsymbol{Y}}$. The SVD of $\bar{\boldsymbol{Y}}$ gives the estimated age-specific coefficients $\hat{b}_x$ and the estimated time component $\hat{\kappa}_t$. The completed version of the estimation procedure is summarized as follows in Algorithm B.1. A wide range of extensions of the Lee–Carter model have been developed and published (see e.g., Brouhns *et al*., 2002; Cairns *et al*., 2006; Renshaw and Haberman, 2006; Hyndman and Ullah, 2007).

C. The HBY model

Hyndman *et al*. (2013) propose a generalized LL model using the product-ratio method and apply principal component analysis to both product and ratio functions to forecast subpopulation mortality coherently. The aggregated model for each country $j \in \{1, ..., J\}$ is given by: for age group $x \in \{1, ..., N\}$ and time $t \in \{1, ..., T\}$,

$$\log(m^j_{x,t}) = \mu_{x,j} + \sum_{r=1}^{R} \beta_{t,r}\phi_{x,r} + \sum_{u=1}^{U} \gamma^j_{t,u}\psi^j_{x,u} + z^j_{x,t},$$

where $\mu_{x,j}$ is the mean of log mortality rate for age group x and country j over time, and $\{\phi_{x,r}\}$ and $\{\psi^j_{x,u}\}$ are two sets of orthonormal basis functions for age group x, country j, index $r \in \{1, ..., R\}$, and $u \in \{1, ..., U\}$. In addition, $\beta_{t,r}$ represents the common time trend and $\gamma^j_{t,u}$ is the country-specific time trend for time t, country j, index $r \in \{1, ..., R\}$, and $u \in \{1, ..., U\}$, respectively. Finally, $z^j_{x,t}$ is the error term. The HBY model can be implemented via the R package `demography`. Following Hyndman *et al*. (2013), we employ the HBY model with order 6 as a baseline for comparative performance evaluation. While determining the optimal order of the HBY model for high-frequency mortality data is of importance, we leave it for future investigation.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

D. The VAR models

The VAR model is one of the most flexible and easy-to-use models for the analysis of multivariate time series (Sims, 1980). It treats all series of the weekly mortality data jointly endogenous, allowing each age group or population to depend on its own lags and those of others. This structure captures both temporal dynamics and cross-group dependencies, which is crucial for multi-population mortality forecasting. However, the VAR framework assumes stationarity, requiring appropriate data preprocessing including normal and seasonal differencing, given the noisy weekly mortality rate with annual cyclical patterns.

First, we consider the single-population VAR model. Based on the observation from Figure 1 and Figure A1, the mortality rate series are typically non-stationary. Therefore, we first take the seasonal difference for each country j if necessary, denoted by $z^j_{x,t} = y^j_{x,t} - y^j_{x,t-52}$. If the differenced series is still not stationary, we will follow Guibert *et al*. (2019) to take the first difference of $z^j_{x,t}$, denoted by $v^j_{x,t} = z^j_{x,t} - z^j_{x,t-1}$ to remove the linear trend. The series represents the log-mortality improvement rates. The VAR(p) model for $\mathbf{v}^j_t = (v^j_{1,t}, ..., v^j_{N,t})$ is given by: for country $j \in \{1, ..., J\}$,

$$\mathbf{v}^j_t = \mathbf{c}^j + \sum_{i=1}^{p} A^j_i \mathbf{v}^j_{t-i} + \boldsymbol{\varepsilon}^j_t,$$

where $\mathbf{c}^j \in \mathbb{R}^x$ is a vector of intercept terms, p is the order, $A^j_i \in \mathbb{R}^{x\times x}$ for $i = 1, .., p$ are autoregressive coefficient matrices, and ε^j_t is the error term. The lag order p of the VAR model is selected using the Hannan–Quinn (HQ) information criterion, which balances goodness-of-fit and model parsimony through a penalty term proportional to $2k \log(\log N)$, where k is the total number estimated parameters and N is the total number of observations. The HQ criterion balances the overfitting nature of the Akaike information criterion (AIC) and the underfitting of the Bayesian information criterion (BIC) (see e.g., Hatemi-J, 2003; Ivanov and Kilian, 2005). The VAR model can be implemented via the R package `vars`, and the out-of-sample forecast is carried out via the standard `predict` function.

The GVAR model (Pesaran *et al*., 2004) extends the standard VAR frameworks by jointly modeling multiple interconnected time series through country-specific VARX models augmented with weighted foreign variables. Cross-country dependence is captured via these foreign aggregates rather than through a fully unrestricted multivariate system, allowing for scalable estimation and coherent global dynamics. The GVAR model consists of country-specific VARX models of the form:

$$\mathbf{v}^j_t = \mathbf{c}^j + \sum_{p=1}^{P^j} A^j_p \mathbf{v}^j_{t-p} + \sum_{q=0}^{Q^j} B^j_q \mathbf{v}^{j*}_{t-q} + \boldsymbol{\varepsilon}^j_t,$$

where $B^j_q \in \mathbb{R}^{x\times x}$ for $q = 1, .., Q^j$ are autoregressive coefficient matrices for the foreign variables $\mathbf{v}^{j*}_{t-q}$. The foreign variables $\mathbf{v}^{j*}_t$ entering the country-specific models are defined as weighted averages of other countries' variables:

$$\mathbf{v}^{j*}_t = \sum_{j\neq i} w_{ij} \mathbf{v}^i_t,$$

where $w_{ij} = 1/29$ for $j \neq i$ and $w_{ij} = 0$ otherwise, since we have 30 countries in the modeling process. The GVAR model can be implemented via the R package `GVARX`. However, the order can only be chosen between 1 and 2. The out-of-sample forecast is carried out following the approach detailed in Section 2.2 of Di Mauro and Pesaran (2013).

E. The country-specific trends under the GBLL model

In addition to the common trends discussed in Section 4.1, we present a detailed decomposition of the country-specific components under the GBLL model for the first two iterations in Figure E1.

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press

- For the age-specific intercepts $\hat{a}_x$, Australia and New Zealand exhibit an opposite pattern relative to the Northern Hemisphere countries. This is due to the use of the reciprocal transformation of mortality rates. As with the common components, the primary effects are largely captured in the first iteration, as evidenced by the larger magnitudes.
- For the age loadings $\hat{b}_x$ in the first iteration, Slovenia displays elevated mortality sensitivity among individuals aged 65–74 years. Denmark shows comparatively higher exposure in the 15–64 years of age group, despite having relatively favorable outcomes for the 65–74 years of cohort compared to most other countries. Additionally, higher exposure is observed across all age groups except for those aged 85 years and above in Netherlands. In the second iteration, the country-specific components reveal particularly pronounced deviations for Belgium and Finland, suggesting substantial excess loadings relative to other countries.
- For the time-varying index $\hat{\kappa}_t$, the clear W-shaped seasonal pattern evident in the common trend is often obscured at the individual country level, largely due to local noise and country-specific variability. The first iteration captures the majority of temporal dynamics, as indicated by the larger magnitudes of the components and a lower degree of white-noise behavior relative to the second iteration. Nonetheless, several countries exhibit unexpected behaviors for the second iteration. For example, Denmark shows a sharp downward spike in January 2017, while Netherlands experiences a pronounced upward spike in March 2018.

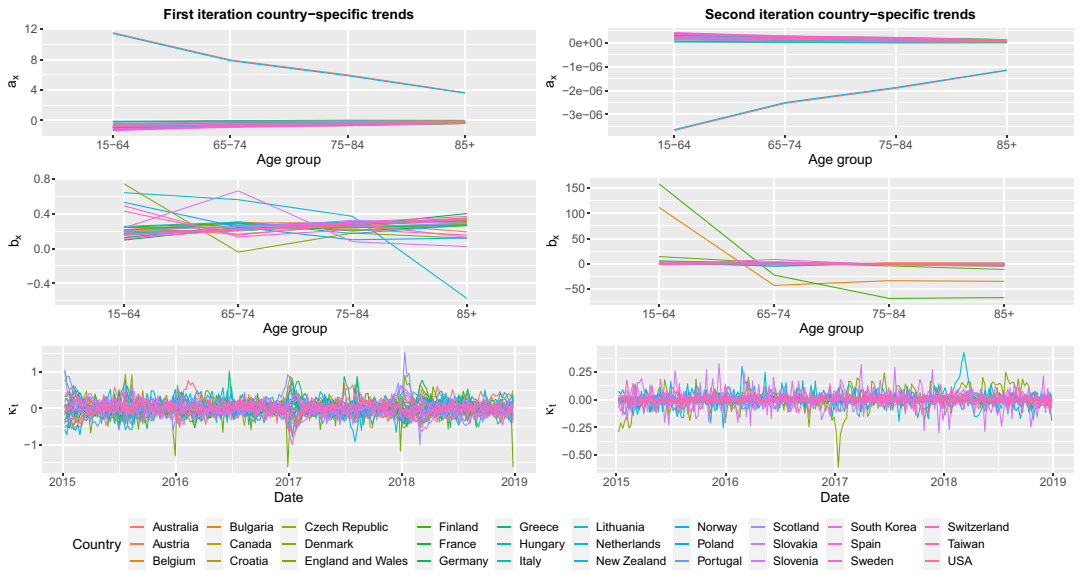


Figure E1. *Components of the country-specific trends under the GBLL model: first iteration (left) and second iteration (right).*

https://doi.org/10.1017/asb.2026.10104 Published online by Cambridge University Press