

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

Ziting Miao, Han Li, Yuyu Chen

Centre for Actuarial Studies, Department of Economics,
The University of Melbourne, Australia

**L21: The Twenty First International Longevity Risk and
Capital Markets Solutions Conference, Rome, Italy**

September 8 – 9, 2026

- **Research gap:** Higher frequency mortality data is rarely used in mortality forecasting due to:
 - Recent availability of data (first published in May 2020),
 - Existing models do not accommodate for seasonal cycles.
- **Existing literature:** COVID-19 related excess mortality. (see, e.g., [Kontis et al., 2020](#); [Jdanov et al., 2021](#); [Islam et al., 2021](#)).
- **Data usefulness:**
 - To better understand the seasonal patterns (**only detectable in weekly data**), e.g., extra short-term fluctuations, elevated winter mortality, excess deaths due to heatwaves, significant effects on elder adults,
 - To provide more accurate forecasts in a timely manner.

- **Data source:** National-level weekly mortality data from the Short-Term Mortality Fluctuations (STMF) series, HMD ([STMF, 2021](#)).
- **Time:** Second week of 2015 to the final week of 2019.
- **Age groups:** 15–64, 65–74, 75–84, and 85+ years old.
- **Country list:** We use 30 countries in this analysis.



Figure 1: 2015–2019 Mortality rate for the four representative countries by age groups

- We propose a modeling framework to provide timely forecasts with high-frequency data under the multi-population setting.
- This model is:
 - **Adaptive** to seasonal patterns,
 - **Insightful** on age and time trend components,
 - **Flexible** in parameter selection,
 - **Insensitive** to the choices of countries.

Multi-population mortality model: The Li and Lee (LL) model

Li and Lee (2005) develop an augmented common factor model: For age group $x \in \{1, \dots, N\}$, time $t \in \{1, \dots, T\}$, and country $j \in \{1, \dots, J\}$,

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

where

- A_x^j : age-specific mean of mortality rate over time for country j ,
- b_x^p : age-specific loading for the common time trend κ_t^p ,
- κ_t^p : common time trend of mortality rates,
- b_x^j : age-specific loading for country-specific time trend κ_t^j ,
- κ_t^j : country-specific mortality trend for country j ,
- $\varepsilon_{x,t}^j$: errors with zero mean.

Multi-population mortality model: The Hyndman–Booth–Yasmeen (HBY) model

Hyndman et al. (2013) propose a generalised version of the Li and Lee model by performing the **principal component analysis** on both the product and ratio functions. The aggregated model for each country $j \in \{1, \dots, J\}$ is given by: For $x \in \{1, \dots, N\}$ and $t \in \{1, \dots, T\}$,

$$\log(m_{x,t}^j) = \mu_{x,j} + \sum_{r=1}^R \beta_{t,r} \times \phi_{x,r} + \sum_{u=1}^U \gamma_{t,u}^j \times \psi_{x,u}^j + z_{x,t}^j,$$

where

- $\mu_{x,j}$: the group mean, the level of log mortality rate at age x for country j over time t ,
- $\{\phi_{x,r}\}$ and $\{\psi_{x,u}^j\}$ are two sets of orthonormal basis functions,
- $\beta_{t,r}$ represents the coefficients for the common trend for $r \in \{1, \dots, R\}$,
- $\gamma_{t,u}^j$ is the country-specific coefficient for $u \in \{1, \dots, U\}$,
- $z_{x,t}^j$ is the error term.

- The idea of gradient boosting is to fit a **weak learner**, compute residuals, refit the weak learner, recompute residuals, etc.
- Weak learners are typically decision trees. But can also be:
 - Generalised linear models (Bühlmann and Hothorn, 2007),
 - Generalised additive models (Tutz and Binder, 2006),
 - Copulas (Hans et al., 2023).
- **Novel idea 1: Do gradient boosting with the Li and Lee model:**

$$\sum_{g=1}^I \gamma_g \hat{y}_{x,g}^j,$$

- $\hat{y}_{x,g}^j$ is the g -th prediction from the Li and Lee model for country j ,
- γ_g is the associated coefficient,
- I is the number of iterations.

- **Stopping criteria:**

- (a) all 120 residuals (30 countries $\times$ 4 age groups) have p-values ≥ 0.05 under the Ljung–Box test (Ljung and Box, 1978);
- (b) maximum in-sample MAPE satisfies

$$\max_{x,j,t}(\text{MAPE}_{x,j,t}) = \max_{x,j,t} \left(\frac{|\hat{m}_{x,t}^j - \tilde{m}_{x,t}^j|}{\tilde{m}_{x,t}^j} \right) < 0.001.$$

- We set the **maximum number of boosting iterations** to $G = 50$.
- **Novel Idea 2: Harmonising seasonality across hemispheres:**

Apply a reciprocal transformation to observed mortality rates in Southern Hemisphere countries.

Methodology: GBLLE estimation

Algorithm 1: Estimation of the GBLLE Model

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

Output: $\hat{\mathbf{r}}_{t,l}^p, \hat{\mathbf{a}}_{x,l}^p, \hat{\mathbf{b}}_{x,l}^p$ and γ_l for $x \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$ and $l \in \{1, \dots, G\}$

$\hat{\mathbf{r}}_{t,l}^j, \hat{\mathbf{a}}_{x,l}^j, \hat{\mathbf{b}}_{x,l}^j$ for $x \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$, $l \in \{1, \dots, G\}$ and $j \in \{1, \dots, J\}$

Procedure: GBLLE($\mathbf{Y}^j$)

1: $\hat{\mathbf{y}}_{x,1}^j, \hat{\mathbf{r}}_{t,1}^p, \hat{\mathbf{a}}_{x,1}^p, \hat{\mathbf{b}}_{x,1}^p, \hat{\mathbf{r}}_{t,1}^j, \hat{\mathbf{a}}_{x,1}^j, \hat{\mathbf{b}}_{x,1}^j \leftarrow \text{LL}(\mathbf{Y}^j)$, $\forall x, t$ and j

2: $L_1 \leftarrow \sum_{j=1}^J \sum_{x=1}^N (\mathbf{y}_x^j - \gamma \hat{\mathbf{y}}_{x,1}^j)' (\mathbf{y}_x^j - \gamma \hat{\mathbf{y}}_{x,1}^j)$

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

4: $\mathbf{e}_{x,1}^j \leftarrow \mathbf{y}_x^j - \gamma_1 \hat{\mathbf{y}}_{x,1}^j$, $\forall x, j$

5: $\mathbf{E}_1^j \leftarrow (\mathbf{e}_{1,1}^j, \dots, \mathbf{e}_{N,1}^j)$, $\forall j$

6: $g \leftarrow 1$

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

$p_{x,g}^j \leftarrow 1 - F_{\chi_{52}^2} \left(T(T+2) \sum_{k=1}^{52} (\rho_{k,x,g}^{(j)})^2 / (T-k) \right)$;

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

$l \leftarrow g$;

break;

else

$\hat{\mathbf{y}}_{x,g+1}^j, \hat{\mathbf{r}}_{t,g+1}^p, \hat{\mathbf{a}}_{x,g+1}^p, \hat{\mathbf{b}}_{x,g+1}^p, \hat{\mathbf{r}}_{t,g+1}^j, \hat{\mathbf{a}}_{x,g+1}^j, \hat{\mathbf{b}}_{x,g+1}^j \leftarrow \text{LL}(\mathbf{E}_g^j)$, $\forall x, t$ and j ;

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

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

$\mathbf{e}_{x,g+1}^j \leftarrow \mathbf{e}_{x,g}^j - \gamma_{g+1} \hat{\mathbf{y}}_{x,g+1}^j$, $\forall x, j$;

$\mathbf{E}_{g+1}^j \leftarrow (\mathbf{e}_{1,g+1}^j, \dots, \mathbf{e}_{N,g+1}^j)$, $\forall j$;

$g \leftarrow g + 1$

Methodology: Forecasting the time trends

- The common and country-specific time trends, κ_t^P and κ_t^j , are forecasted using a **Fourier–ARIMA approach**.
- While SARIMA can capture annual seasonality, seasonal differencing at lag 52 is undesirable for our short weekly sample.
- We therefore model seasonality using **Fourier regressors** (Serfling, 1963) and the remaining stochastic dynamics using ARIMA(p, d, q):

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

- Significant Fourier terms (5% level) are retained; otherwise, forecasting relies solely on ARIMA.
- The final h -step-ahead forecast of log mortality is

$$\log(\hat{m}_{x,T+h}^j) = \sum_{g=1}^I \gamma_g \left(\hat{a}_{x,g}^p + \hat{b}_{x,g}^p \hat{\kappa}_{T+h|T,g}^p + \hat{a}_{x,g}^j + \hat{b}_{x,g}^j \hat{\kappa}_{T+h|T,g}^j \right)$$

Empirical results: GBLL components breakdown

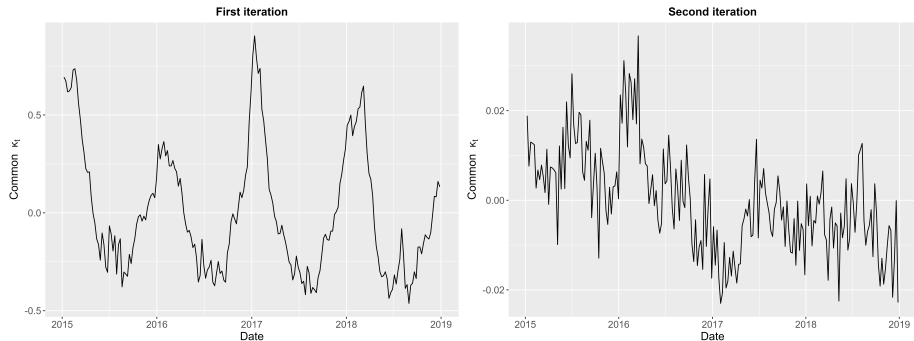


Figure 2: The common time trend under the GBLL model for the first iteration (left) and second iteration (right)

- **First iteration:** dominant seasonal cycles with W-shaped patterns.
- **Second iteration:** Smaller cyclical patterns with fluctuations and a mild downward trajectory.

Empirical results: Out-of-sample forecast accuracy

- **Error measure:** mean absolute percentage error (MAPE).
- **Expanding window** (expanding by one month each step):
 - Starting period: fixed at the second week of 2015,
 - End period: final week of March 2018 to December 2018.
- **Forecast horizon:** 52 weeks.

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

h	LL	HBV	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

- **Marginal calibration** (Czado et al., 2009): proportion of observed mortality rates falling within the 95% prediction intervals:

$$f_{[\hat{\rho}_{2.5}, \hat{\rho}_{97.5}]} = \frac{1}{52 \times 10 \times 4 \times 30} \sum_{h_w, r, x, j} \mathbb{1} \left(\hat{\rho}_{2.5, x, t}^j \leq \tilde{m}_{x, t}^j \leq \hat{\rho}_{97.5, x, t}^j \right).$$

LL, HBV, and GBLL achieve empirical coverage of **90%, 94%, and 96%**, respectively.

Empirical results: Alternative parameter choices of the GBLL model

Table 2: Mean MAPE of out-of-sample forecasts across 30 countries ($\times 100$ scale) for the GBLL model on different choices of the components

h	Original	Absolute loss	Stopping criteria	SARIMA	Seasonal adjustment
1	5.278	5.279	5.278	5.621	5.262
2	5.620	5.621	5.620	6.085	5.640
3	5.582	5.582	5.582	6.243	5.597
4	5.602	5.602	5.602	6.277	5.610
5	5.671	5.672	5.671	6.247	5.685
6	5.718	5.719	5.718	6.214	5.720
7	5.753	5.754	5.753	6.212	5.745
8	5.829	5.829	5.829	6.297	5.811
9	5.878	5.879	5.878	6.366	5.852
10	5.914	5.914	5.914	6.409	5.875
11	5.922	5.922	5.922	6.419	5.877
12	5.911	5.911	5.911	6.420	5.866

- **Alternative stopping criterion:** stop boosting when $|L_g - L_{g-1}| < 10^{-6}$.
- **SARIMA:** specified by orders $(p, d, q) \times (P, D, Q)_s$.
- **Seasonal adjustment:** flip the observed mortality peaks and troughs around the historical mean for Southern Hemisphere countries: $\log(\tilde{m}_{x,t}^{j*}) = 2\tilde{b}_x^j - \log(\tilde{m}_{x,t}^j)$.

- **Context:** Clustering defines sub-populations based on characteristics such as sex and states (Hyndman et al., 2013), or ages (see, e.g., Léger and Mazzuco, 2021; Tsai and Cheng, 2021; Debon et al., 2024).
- **Novel idea 3: Construction of clustering data inputs**
Exploiting short-term dynamics and seasonality unique to high-frequency data.
- **Four main components:**
 - **Data input:** raw single-population time trends; improvement rates; improvement rates and seasonal strength,
 - **Clustering:** k -means,
 - **Number of clusters:** “elbow” method,
 - **Distance:** Euclidean distance.

Method 1: Clustering based on raw time trend series

- We first extract the time trend component κ_t for each country using the Lee–Carter model.
- We analyse the raw time series of the country-specific κ_t .

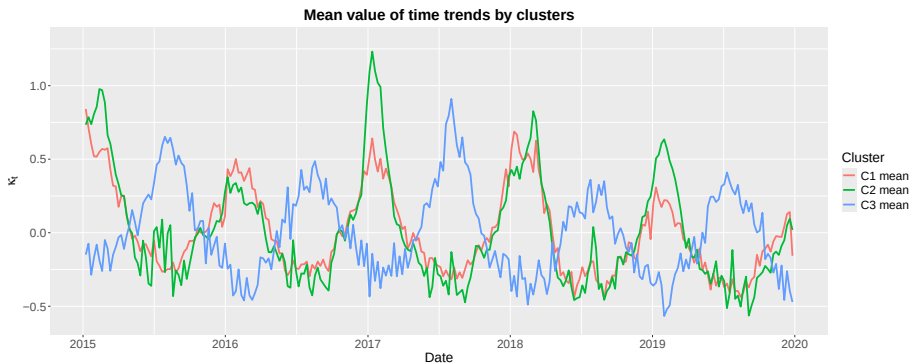


Figure 3: Mean value of the time trends in each cluster

Method 2: Clustering based on estimated trend coefficients

- We apply the Seasonal and Trend decomposition using LOESS (STL) to the country-specific κ_t series (Cleveland et al., 1990).
- We compute and analyse the slope of each trend component.

Clustering based on slopes of the time trend

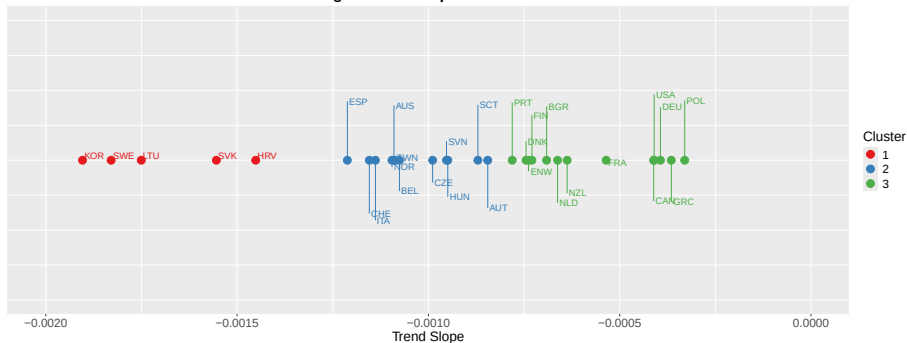


Figure 4: Trend slopes by clusters

Method 3: Clustering based on estimated trend coefficients and seasonal strength

- We incorporate both the estimated trend coefficients and the seasonal strength derived from the STL decomposition.
- We apply the min-max scaling on both features.

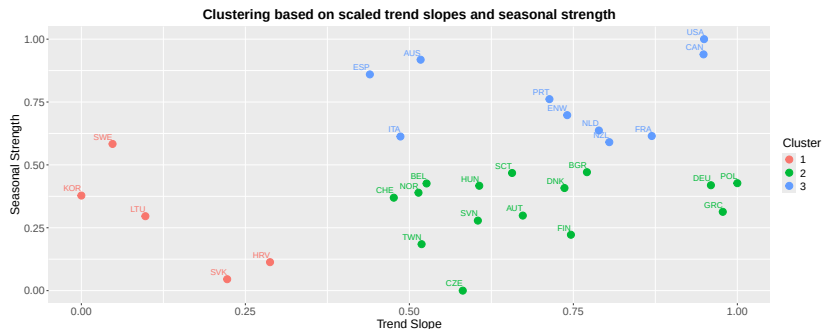


Figure 5: Clusters based on min-max scaled trend slopes and seasonal strength

Table 3: Mean MAPE of out-of-sample forecasts across 30 countries with clustering ($\times 100$ scale)

	Method 1			Method 2			Method 3			Whole
h	LL	HBV	GBLL	LL	HBV	GBLL	LL	HBV	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

- The GBLL model produces the **most accurate forecasts** under each clustering method.
- Method 2 shows an improvement from the case without clustering.
- The GBLL model is **insensitive** to cross-country heterogeneity, with superior forecast performance.

- We propose a flexible GBLM framework to model and forecast high-frequency mortality rates within a multi-population context.
- The proposed model is insensitive to the choice of modeling components and country groupings, maintaining stable and superior forecast accuracy.
- Potential future research:
 - **Risk hedging:** exploit opposite seasonal mortality patterns across the Northern and Southern Hemispheres.
 - **Model extensions:** apply the boosting framework to other mortality models and incorporate finer age groups or additional dimensions.
 - **Within-year forecasts:** enable early detection of elevated mortality associated with seasonal or extreme events, such as influenza and pandemics.
- Full Paper: Miao, Z., Li, H. and Chen, Y. (2026). Gradient boosted multi-population mortality modeling with high-frequency data. *ASTIN Bulletin: The Journal of the IAA*, 1–26. [[doi:10.1017/asb.2026.10104](https://doi.org/10.1017/asb.2026.10104)]

- Bühlmann, P. and T. Hothorn (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* 22(4), 477–505.
- Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics* 6(1), 3–73.
- Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics* 65(4), 1254–1261.
- Debon, A., S. Haberman, and G. Piscopo (2024). Multipopulation mortality analysis: bringing out the unobservable with latent clustering. *Quality & Quantity* 58(6), 5107–5123.
- Hans, N., N. Klein, F. Faschingbauer, M. Schneider, and A. Mayr (2023). Boosting distributional copula regression. *Biometrics* 79(3), 2298–2310.
- Hyndman, R. J., H. Booth, and F. Yasmeeen (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* 50(1), 261–283.
- Islam, N., V. M. Shkolnikov, R. J. Acosta, I. Klimkin, I. Kawachi, R. A. Irizarry, G. Alicandro, K. Khunti, T. Yates, D. A. Jdanov, et al. (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.
- Jdanov, D. A., A. A. Galarza, V. M. Shkolnikov, D. Jasilionis, L. Németh, D. A. Leon, C. Boe, and M. Barbieri (2021). The short-term mortality fluctuation data series, monitoring mortality shocks across time and space. *Scientific Data* 8(1), 235.
- Kontis, V., J. E. Bennett, T. Rashid, R. M. Parks, J. Pearson-Stuttard, M. Guillot, P. Asaria, B. Zhou, M. Battaglini, G. Corsetti, et al. (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.
- Léger, A.-E. and S. Mazzucco (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.

- Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography* 42, 575–594.
- Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika* 65(2), 297–303.
- Serfling, R. E. (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports* 78(6), 494.
- 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). Data downloaded on 18 October 2024.
- Tsai, C. C.-L. and E. S. Cheng (2021). Incorporating statistical clustering methods into mortality models to improve forecasting performances. *Insurance: Mathematics and Economics* 99, 42–62.
- Tutz, G. and H. Binder (2006). Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics* 62(4), 961–971.