Astin Bulletin (2026), **56**, pp. 367–388
doi:10.1017/asb.2026.10092

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Mortality Forecasting in Geographical Space and Time

Gábor Szentkereszti[1] and Péter Vékás[2]

[1]Allianz Hungária Insurance Corporation, Hungary, Corvinus University of Budapest, Hungary
[2]Department of Operations Research and Actuarial Science, Corvinus University of Budapest, Hungary
Corresponding author: Péter Vékás; Email: peter.vekas@uni-corvinus.hu

Received: 22 September 2025; **Revised:** 6 February 2026; **Accepted:** 21 February 2026;
First published online: 25 March 2026

Keywords: Mortality forecasting; actuarial science; spatial statistics; machine learning; artificial intelligence

JEL codes: C51; C52; C53; G22; C45

Abstract

The vast majority of researchers, actuaries, and demographers use standard time series analysis techniques to project time-varying parameters of popular mortality forecasting methods such as the Lee–Carter and Li–Lee models. However, spatial dependence can be as significant as temporal autocorrelation in these time series, and the underlying panel structure of the data is often neglected. We draw on techniques from panel and spatial econometrics, including ordinary and spatial dynamic panel linear models, spatiotemporal autoregressive integrated moving average processes, and spatial eigenvector filters, to capture such dependence and improve projections. We present a methodology to estimate the parameters of these techniques from spatial multipopulation mortality series, select their optimal hyperparameters, and use them for forecasting. We propose a tailor-made robust selection framework to identify the best model–technique combinations for each country, as well as a bootstrap-based procedure to quantify projection uncertainty with accurate nominal coverage on a separate validation period and a strategy for assessing the quality of the resulting prediction intervals. We test these methods on mortality data from 22 European countries. The results show that the proposed techniques yield a clear advantage in both point and interval forecasts for several populations, and these findings are corroborated by a robust selection design and additional robustness checks. These improvements have the potential to deliver meaningful gains for life insurance, pensions, and other contexts involving longevity risk.

1. Introduction

Mortality forecasting is crucial for demographers and actuaries, enabling them to predict future population trends and financial liabilities with greater accuracy. For demographers, understanding mortality patterns helps in projecting population growth, age distribution, and life expectancy, which are essential for planning public health initiatives, social services, and economic policies. Actuaries rely on mortality forecasts to assess the long-term viability of pension plans, life insurance policies, and annuity products. Accurate mortality predictions ensure that these financial instruments are adequately funded, mitigating the risk of insolvency.

The Lee–Carter (LC) model of Lee and Carter (1992) is a seminal approach in mortality forecasting, capable of decomposing mortality rates into a time-varying mortality index and age-specific components. By employing standard time series techniques to forecast the mortality index, it enables the modeler to effectively capture temporal dependencies in mortality trends. However, when used for a set of countries, this framework is limiting as it overlooks the cross-sectional and spatial dependencies evident in mortality data across different populations.

Mortality rates are influenced by a myriad of factors, some of which vary geographically and are interconnected across regions, such as those related to climate and culture. The fields of spatial and panel

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Actuarial Association.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

econometrics offer a rich literature and a wide range of methodologies that can address these complex dependencies. Incorporating these into mortality forecasting models has the potential to enhance their accuracy and predictive power, accounting for nuanced temporal, spatial, and spatiotemporal interactions.[1] Despite its potential, this integration remains poorly explored, presenting an opportunity to significantly advance the precision and reliability of mortality forecasts.

We present, adopt, and test four advanced techniques that account for cross-sectional and spatiotemporal dependencies to forecast the mortality index. Beyond the LC model, we also apply the proposed methodology to country-specific mortality indices from the Li–Lee (LL) model for coherent multipopulation mortality forecasting (Li and Lee, 2005). Besides refining predictions, we aim to highlight the importance of accounting for spatial and spatiotemporal correlations and cross-sectional interactions in demographic and actuarial models. We emphasize robustness in performance evaluation and the accurate assessment of prediction uncertainty. Finally, we present a numerical analysis involving a comprehensive performance comparison of various forecasting techniques across 22 countries.

The remainder of our paper is organized as follows: Section 2 presents a literature review, Section 3 provides a summary of the applied methodology, Section 4 details our results, and Section 5 discusses the conclusions and limitations of our study.

2. Literature review

We present our literature review in three parts: a concise and nonexhaustive overview of mortality forecasting models, a survey of dynamic panel and spatiotemporal forecasting methodologies relevant to our analysis, and a brief literature review on Bayesian spatiotemporal mortality models not considered in this paper.

2.1. Mortality forecasting

Mortality forecasting has seen significant advancements over the decades. The groundbreaking paper of Lee and Carter (1992) decomposed age-specific logarithmic mortality rates into components representing age patterns, a time trend modeled as a discrete-time stochastic process, and random noise, providing a flexible yet parsimonious framework for forecasting. Building on this foundation, Lee and Miller (2001) proposed the Lee–Miller variant to better fit historical data and improve forecast accuracy. Booth *et al*. (2002) further extended the LC framework with their Booth–Maindonald–Smith variant, incorporating age-specific improvements and adjustments for variable mortality decline. Meanwhile, Brouhns *et al*. (2002) introduced a Poisson log-bilinear approach, integrating the LC methodology within a statistical framework that models death counts as Poisson-distributed variables, enabling rigorous probabilistic inference.

The Cairns–Blake–Dowd model (Cairns *et al*., 2006) represented another significant contribution to the field, followed by several extensions in Cairns *et al*. (2009). This method focuses on higher ages and models the logit transform of death probabilities using stochastic processes for the intercept and slope parameters, making it particularly useful for pension and insurance applications.

Cohort effects are important extensions of these models, and the Age–Period–Cohort model (Clayton and Schifflers, 1987) is one of the foundational approaches in this field. It accounts separately for the effects of age, calendar time (period), and birth year (cohort) on mortality rates. Building on this, Renshaw and Haberman (2006) proposed the Renshaw–Haberman model, which incorporates a cohort effect into the LC framework. The generalized age-period-cohort framework (Hunt and Blake, 2014), implemented by Villegas *et al*., (2015), provides a unifying approach to modeling mortality and generalizes several models by incorporating a set of probability distributions and link functions, additional age–period interactions, and cohort effects.

[1] We include the word "geographical" in the title to emphasize that we are using the word "space" in its traditional—and not in an abstract mathematical—sense.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

P-spline smoothing, introduced by Currie *et al*. (2004), is a flexible method for modeling and forecasting mortality rates while effectively balancing fit and smoothness, thereby capturing underlying trends while avoiding overfitting. Hyndman and Ullah (2007) proposed a functional data approach to mortality forecasting, emphasizing robust handling of mortality and fertility rates by leveraging smooth functional representations of demographic data.

Multipopulation mortality forecasting has also been a topic of great interest. Li and Lee (2005) extended the LC framework to handle groups of populations, ensuring coherent projections for related demographic units, such as neighboring countries or different sexes. Enchev *et al*. (2017) examined several multipopulation mortality models on data from England and Wales. Antonio *et al*. (2017) developed a stochastic multipopulation standard for Dutch and Belgian mortality projections. Lam and Wang (2023) introduced the weighted multivariate functional principal component approach for coherent multipopulation forecasting.

The LC model assumes constant rates of mortality improvement across ages, contrary to observed trends where mortality decline decelerates at younger ages and accelerates at older ages—a phenomenon known as the "rotation" of mortality decline. Li *et al*. (2013) incorporated this into the LC model for better long-term projections. Vékás (2019) conducted an empirical analysis of this rotation in countries of the European Union using a new data-driven methodology, revealing the presence of this phenomenon in several member states, with notable differences between sexes.

Guibert *et al*. (2019) proposed penalized vector autoregression (VAR) to capture the dependence structure of mortality dynamics. Guibert *et al*. (2020) presented locally coherent forecasts, which enforce coherence only in subgroups of countries and bridge the gap between the independence assumption of the LC model and the strict coherence enforced by the LL model. Li and Shi (2021) introduced a global VAR approach that decomposes multipopulation mortality dynamics into local systems parsimoniously. Njenga and Sherris (2020) proposed Bayesian VAR as a related alternative. Jarner and Jallbjørn (2020) discussed the application of cointegration for forecasting mortality factors. Li *et al*. (2021) presented a multifactor mortality model based on the Gompertz law. He *et al*. (2021) introduced time-varying factor loadings into mortality forecasting models.

Recently, machine learning approaches, including neural network-based models, have significantly advanced mortality forecasting by accommodating the automatic detection of complex nonlinear dependencies (Hainaut, 2018; Richman and Wüthrich, 2021; Perla *et al*., 2024, 2021; De Mori *et al*., 2025).

Comprehensive reviews of mortality forecasting methods and recent extensions are provided by Basellini *et al*. (2023) and Pascariu *et al*. (2018), while Yao and Wang (2023) offer an overview of machine learning frameworks. A full survey of the mortality forecasting literature is beyond the scope of this paper.

2.2. Dynamic panel and spatiotemporal forecasting techniques

Pfeifer and Deutsch (1980) introduced the spatiotemporal autoregressive moving average (STARMA) process as an extension of the well-known ARMA framework that is also capable of capturing spatiotemporal dependencies.

Arellano and Bond (1991) popularized the generalized method of moments (GMM) estimator for dynamic panel linear models, and Blundell and Bond (1998) introduced the improved system-GMM estimator. Yu *et al*. (2008) presented the spatial dynamic panel model with fixed effects and investigated its asymptotic properties. Yu and Lee (2010) proposed a spatiotemporal dynamic panel model that works for both stationary and non-stationary data. Elhorst (2014) and Salima (2018) provided broad overviews of spatial panel models, and Astuti *et al*. (2020) presented theoretical developments and potential research directions in this field.

Griffith (2010, 2012) proposed eigenvector spatial filters and eigenvector spatiotemporal filters (ESTF) for modeling spatial and spatiotemporal relationships, and presented the contemporaneous and lagged spatiotemporal Moran's I as two adequate measures of spatiotemporal autocorrelation in spatial panel data. Islam *et al*. (2022) embedded this eigenvector approach into machine learning models.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

2.3. Bayesian spatiotemporal mortality models

While Bayesian methods are not considered in our analysis, we include selected examples for context. Bayesian spatiotemporal mortality models (Ocaña-Riola and Mayoral-Cortés, 2010; Álvaro-Meca *et al*., 2011; Greco and Scalone, 2013; Bennett *et al*., 2015; Gibbs *et al*., 2020; Liu, 2021) explicitly model structured spatiotemporal dependence through prior distributions. Hassan (2021) extended the LC model using functional PCA, Ugarte *et al*. (2010) introduced Bayesian P-spline smoothing, Pickle (2000) embedded spatiotemporal mortality forecasting in a linear mixed effects framework, and Carracedo *et al*. (2018) identified spatiotemporal mortality clusters in European data.

3. Methods

3.1. Spatial and spatiotemporal autocorrelation

Spatial autocorrelation refers to the correlation of a variable with itself across space—a manifestation of Tobler's First Law of Geography (Tobler 1970). Moran's I, introduced by Moran (1950), is a foundational statistic used to quantify global spatial autocorrelation, measuring the extent to which geographically proximate observations tend to exhibit similar (positive autocorrelation), dissimilar (negative autocorrelation) or spatially random (zero autocorrelation) values. The methodological foundations and theoretical properties of Moran's I were significantly advanced by Cliff and Ord (1973), establishing it as a core tool for exploratory spatial data analysis.

Given a set of N spatial observations $\{u_i\}_{i=1}^{N}$ and an $N \times N$ spatial weights matrix $\mathbf{W}_{\text{space}} \triangleq [w_{ij}]$, Moran's I is defined as:

$$I \triangleq \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(u_i - \bar{u})(u_j - \bar{u})}{\sum_i (u_i - \bar{u})^2},$$

where $\bar{u}$ is the mean of the observations, and w_{ij} encodes spatial relationships (e.g., *contiguity weights*, where $w_{ij} = 1$ if regions i and j are neighbors and 0 otherwise, when values at neighboring locations are assumed to be autocorrelated; or *inverse-distance weights*, where $w_{ij} = 1/\delta_{ij}$, when autocorrelation is assumed to decay with geographical distance δ_{ij} measured between pairs of capital cities or centroids).[2]

Local indicators of spatial association (LISA, Anselin, 1995) provide a standard framework for detecting spatial clustering and spatial outliers that may be obscured in global analyses. By decomposing the global Moran's I statistic into contributions for each geographic unit, LISA identifies areas where neighboring values are significantly similar (high–high or low–low clusters) or dissimilar (high–low or low–high outliers). Statistical significance is typically assessed through permutation testing.

Spatiotemporal autocorrelation refers to the correlation of a variable with itself across space and time; in other words, it describes the similarity or dissimilarity of values measured at proximate locations and time points. In extending Moran's I to spatiotemporal settings, the spatial weights matrix $\mathbf{W}_{\text{space}}$ is expanded using Kronecker products to form a spatiotemporal weights matrix $\mathbf{W}_{\text{space-time}}$.

A common contemporaneous specification is given by:

$$\mathbf{W}^{(\text{contemporaneous})}_{\text{space-time}} \triangleq \mathbf{I}_T \otimes \mathbf{W}_{\text{space}} + \mathbf{W}_{\text{time}} \otimes \mathbf{I}_N,$$

where $\mathbf{I}_T$ and $\mathbf{I}_N$ are identity matrices of size T and N, respectively, and $\mathbf{W}_{\text{time}}$ is a temporal adjacency matrix of size T connecting each time period with its predecessor and successor, with ones on the first lower and upper off-diagonals and zeros elsewhere (Griffith, 2010, 2012; Griffith and Paelinck, 2018).

Alternatively, the lagged specification is given by:

$$\mathbf{W}^{(\text{lagged})}_{\text{space-time}} \triangleq \mathbf{W}_{\text{time}} \otimes \mathbf{W}_{\text{space}} + \mathbf{W}_{\text{time}} \otimes \mathbf{I}_N,$$

capturing the relationship between observations at each location and their neighbors' past values.

[2]Moran's I values are slopes and not correlation coefficients, and are thus not constrained to the interval between –1 and 1.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Given these weights matrices, spatiotemporal Moran's I is defined as:

$$I^{(\text{spec})}_{\text{space-time}} \triangleq \frac{NT}{\sum_{(i,t)}\sum_{(j,s)} w^{(\text{spec})}_{(i,t),(j,s)}} \frac{\sum_{(i,t)}\sum_{(j,s)} w^{(\text{spec})}_{(i,t),(j,s)}(u_{(i,t)}-\bar{u})(u_{(j,s)}-\bar{u})}{\sum_{(i,t)}(u_{(i,t)}-\bar{u})^2},$$

where spec $\in$ {contemporaneous,lagged} is the chosen specification, $u_{(i,t)}$ denotes observations at location i and time t, and $\bar{u}$ is the overall mean across all spatial and temporal observations (Griffith 2010).

It is both theoretically plausible and empirically supported to assume that mortality improvements exhibit spatiotemporal autocorrelation, thereby motivating the integration of spatiotemporal econometric methods into mortality forecasting.

3.2. Stochastic mortality models

Let D_{xt} and E_{xt} represent the number of deaths and the central exposure at age $x \in \mathcal{X}$ and year $t \in \{1, 2, \ldots, T\}$, where $\mathcal{X}$ denotes the age range included in the analysis, and let $m_{xt} \triangleq D_{xt}/E_{xt}$ be the corresponding central mortality rate. The LC model assumes

$$\ln m_{xt} = a_x + b_x k_t + \varepsilon_{xt},$$

where a_x and b_x are age-specific parameters, k_t is a time series, and ε_{xt} are error terms with mean 0 and variance $\sigma^2_\varepsilon > 0$. For parameter estimation, we adopt the approach of maximizing the Poisson log-likelihood following Brouhns *et al*. (2002), which provides a statistically rigorous alternative to the singular value decomposition (SVD) used by Lee and Carter (1992).

The parameters a_x denote the average log-mortality rates by age in the sample, while the mortality index series k_t represents mortality dynamics in the entire population, and b_x determines sensitivities to changes in the mortality index by age group. To address the issue of identification, the following constraints are imposed:

$$\sum_{t=1}^{T} k_t = 0 \quad \text{and} \quad \sum_{x\in\mathcal{X}} b_x = 1.$$

In the remainder of this paper, we will refer to country $i \in \{1, 2, \ldots, N\}$ using an additional first subscript i in the notation of all parameters.

The LL model (Li and Lee, 2005) closely resembles the LC model in structure:

$$\ln m_{ixt} = a_{ix} + B_x K_t + b_{ix} k_{it} + \varepsilon_{ixt},$$

where B_x and K_t are common parameters, k_{it}, b_{ix}, and a_{ix} are population-specific parameters, and ε_{ixt} are error terms, and it is assumed that

$$\sum_{t=1}^{T} K_t = \sum_{t=1}^{T} k_{it} = 0 \quad \text{and} \quad \sum_{x\in\mathcal{X}} B_x = \sum_{x\in\mathcal{X}} b_{ix} = 1 \quad (i = 1, 2, \ldots, N).$$

For the sake of coherent forecasts with

$$\lim_{t\to\infty} \frac{m_{ixt}}{m_{jxt}} = \text{constant} \quad (\forall x \in \mathcal{X},\ i \neq j),$$

all k_{it} should converge to constants as $t \to \infty$. Consequently, the authors advocate for employing AR(1) models with autoregressive coefficients less than 1 in absolute value or driftless random walk (RW) processes for forecasting the mortality indices k_{it}.

3.3. Forecasting techniques

We consider the following forecasting techniques for the time series of the mortality indices k_{it} of the LC or LL models, with N and T denoting the number of countries and calendar years, respectively, and the

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

notations $\boldsymbol{\xi}_t \triangleq (\xi_{1t}, \xi_{2t}, \ldots, \xi_{Nt})^\intercal$ $(t = 1, 2, \ldots, T)$ for the time-dependent and $\boldsymbol{\xi} \triangleq (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_T)^\intercal$ for the overall vector of innovations of the time series models with mean $\boldsymbol{\mu} \triangleq \mathbb{E}(\boldsymbol{\xi})$ and covariance matrix $\boldsymbol{\Sigma} \triangleq \mathbb{E}\left[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})^\intercal\right]$:

1. Random walk with drift (RWD):

$$\begin{aligned} &k_{it} = \alpha_i + k_{i,t-1} + \xi_{it}, \\ &\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\ &\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2 \mathbf{I}_{T-1}, \sigma_2^2 \mathbf{I}_{T-1}, \ldots, \sigma_N^2 \mathbf{I}_{T-1}), \end{aligned}$$

   so that innovations are multivariate normal with mean zero, constant variance within countries, and no correlations across both countries and time periods. Assuming that the first-order differenced series is stationary, parameters can be estimated by maximum likelihood estimation (MLE). Lee and Carter (1992) used this specification for US data, and it has been shown to perform well across a large number of populations (Tuljapurkar *et al*., 2000).
2. Autoregressive integrated moving average (ARIMA):

$$\begin{aligned} &y_{it} \triangleq \Delta^{(d_i)} k_{it} = \alpha_i + \sum_{\ell=1}^{p_i} \phi_{i\ell} y_{i,t-\ell} + \sum_{\ell=1}^{q_i} \theta_{i\ell} \xi_{i,t-\ell} + \xi_{it}, \\ &\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\ &\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2 \mathbf{I}_{T-d_1}, \sigma_2^2 \mathbf{I}_{T-d_2}, \ldots, \sigma_N^2 \mathbf{I}_{T-d_N}). \end{aligned}$$

   Again, it is assumed that innovations are multivariate normal with mean zero, homoskedastic within countries and uncorrelated across countries and time periods. Differencing is applied to ensure the removal of unit roots for stationarity. As with RWD, the standard estimation method is MLE. The numbers of lags p_i and q_i by country can be optimized by minimizing the Bayesian information criterion (BIC) or related alternatives. The RWD is a special case with $p_i = 0$, $d_i = 1$, and $q_i = 0$ $(i = 1, 2, \ldots, N)$. There are no interactions among cross-sectional units.
3. Dynamic panel linear model (DPLM):

$$\begin{aligned} &\mathbf{y}_t \triangleq \Delta^{(d)} \mathbf{k}_t = \boldsymbol{\alpha} + \sum_{\ell=1}^{p} \phi_\ell \mathbf{y}_{t-\ell} + \boldsymbol{\xi}_t, \\ &\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \operatorname{diag}\left(\sigma_1^2 \mathbf{I}_{T-d}, \sigma_2^2 \mathbf{I}_{T-d}, \ldots, \sigma_N^2 \mathbf{I}_{T-d}\right). \end{aligned}$$

   It is assumed that innovations have mean zero and are homoskedastic within countries and uncorrelated across countries and time periods. This in turn amounts to uniform ARIMA specifications with $d_i = d$, $p_i = p$, $\phi_{i\ell} = \phi_\ell$, and $q_i = 0$ $(i = 1, 2, \ldots, N)$, but without the assumption of multivariate normality. Popular estimators for this model are the generalized method of moments (GMM, Arellano and Bond, 1991) and system-GMM (Blundell and Bond, 1998). The lag order p can be determined based on the statistical significance of the autoregressive terms. There are no interactions among cross-sectional units.
4. Vector autoregression (VAR):

$$\begin{aligned} &\mathbf{y}_t \triangleq \Delta^{(d)} \mathbf{k}_t = \boldsymbol{\alpha} + \sum_{\ell=1}^{p} \mathbf{A}_\ell \mathbf{y}_{t-\ell} + \boldsymbol{\xi}_t, \\ &\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \operatorname{diag}\left(\sigma_1^2 \mathbf{I}_{T-d}, \sigma_2^2 \mathbf{I}_{T-d}, \ldots, \sigma_N^2 \mathbf{I}_{T-d}\right). \end{aligned}$$

   Innovations are assumed to have mean zero and equal variance within countries, with no correlation across countries and time periods. Differencing ensures the removal of unit roots, as stationarity is required to fit the model. Estimation is typically performed by ordinary least squares (OLS), which does not require normality of the innovations. However, in our application, the ratio of observations to parameters is well below the commonly applied rule of thumb of at least five (Van Belle, 2008), so OLS is not recommended. Instead, coefficients can be

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

estimated using elastic net—or, as in our special case, LASSO or L_1—regularization (Guibert *et al*., 2019) in order to achieve a parsimonious description of the dependence structure. The lag order p can be determined by minimizing the cross-validated mean squared error (CV-MSE). The VAR framework allows for cross-sectional interactions, which are not necessarily spatial in nature.

5. Spatiotemporal ARIMA (STARIMA, Pfeifer and Deutsch, 1980):

$$z_{it} \triangleq \frac{\Delta^{(d)}k_{it} - \underset{i,t}{\text{mean}}\left(\Delta^{(d)}k_{it}\right)}{\underset{i,t}{\text{sd}}\left(\Delta^{(d)}k_{it}\right)} =$$

$$= \alpha + \sum_{\ell=1}^{p}\sum_{r=0}^{\lambda_\ell}\phi_{\ell r}W^{(r)}z_{i,t-\ell} + \sum_{\ell=1}^{q}\sum_{r=0}^{\mu_\ell}\theta_{\ell r}W^{(r)}\xi_{i,t-\ell} + \xi_{it},$$

$$\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \sigma^2\mathbf{I}_{N(T-d)},$$

where $W^{(r)}$ is the r-th order spatial lag operator representing adjacency of r-th order.[3] Innovations are assumed to have mean 0 and equal variance *across all countries*, with no correlation across countries or time periods. Differencing ensures the removal of unit roots. Estimation is performed using the Kalman filter (KF), and the numbers of time lags p and q and space-time lags λ_ℓ ($\ell = 1, 2, \ldots, p$) and μ_ℓ ($\ell = 1, 2, \ldots, q$) can be determined by minimizing the BIC. STARIMA is capable of capturing complex spatiotemporal interactions.

6. Spatial dynamic panel linear model (SDPLM):

$$\mathbf{y}_t \triangleq \Delta^{(d)}\mathbf{k}_t = \boldsymbol{\alpha} + \phi_{\text{space}}\,\mathbf{W}_{\text{space}}\mathbf{y}_t + \phi_{\text{time}}\,\mathbf{y}_{t-1} + \phi_{\text{space-time}}\,\mathbf{W}_{\text{space}}\mathbf{y}_{t-1} + \boldsymbol{\xi}_t,$$

$$\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \text{diag}\left(\sigma_1^2\mathbf{I}_{T-d}, \sigma_2^2\mathbf{I}_{T-d}, \ldots, \sigma_N^2\mathbf{I}_{T-d}\right).$$

It is assumed that innovations have mean zero and are homoskedastic within countries and uncorrelated across countries and time periods. There is no normality assumption, and parameters are estimated by quasi maximum likelihood (QML), with the Lee–Yu transformation removing finite-sample bias (Yu and Lee, 2010). The presence or absence of a time lag and a space-time lag, the choice between adjacency and inverse-distance spatial weights, and the necessity of the Lee–Yu transformation can be determined by minimizing the quasi-BIC. SDPLM allows for structured and easily interpretable spatiotemporal interactions.

7. Eigenvector spatiotemporal filter (ESTF):

$$\Delta^{(d)}\mathbf{k}_t = \boldsymbol{\alpha} + \mathbf{S}_{\text{space-time}}\,\boldsymbol{\beta} + \mathbf{S}_{\text{space}}\,\boldsymbol{\gamma} + \mathbf{P}\boldsymbol{\phi} + \boldsymbol{\xi}_t,$$

$$\boldsymbol{\mu} = \mathbf{0}, \quad \boldsymbol{\Sigma} = \text{diag}(\sigma_1^2\mathbf{I}_{T-d}, \sigma_2^2\mathbf{I}_{T-d}, \ldots, \sigma_N^2\mathbf{I}_{T-d}),$$

where $\mathbf{S}_{\text{space}}$ and $\mathbf{S}_{\text{space-time}}$ are matrices of eigenvectors of the spatial and spatiotemporal weights matrices $\mathbf{W}_{\text{space}}$ and $\mathbf{W}_{\text{space-time}}$, and $\mathbf{P}$ is a matrix of country dummy variables. Griffith (2010) proposes keeping all country dummy variables in the equation and using stepwise OLS (SW-OLS) estimation for the sake of parsimony, or alternatively, shrinkage by LASSO can be applied.[4] Choices between adjacency and inverse-distance spatial weights and between contemporaneous and lagged specifications can be made by minimizing the BIC in the case of SW-OLS and the deviance ratio (DR) in the case of LASSO. ESTF allows for complex spatiotemporal interactions, but is typically less interpretable than an SDPLM.

Chief properties of the proposed techniques are displayed in Table 1. Additionally, we consider two ensemble techniques. The first one, denoted by ENS1, computes the mean predicted log-mortality rate across the seven forecasting techniques for every country, age, and calendar year in both the LC and LL

[3]For example, $W^{(2)}$ takes the average value of neighbors of neighbors, excluding the unit itself.

[4]As warranted by widespread criticism of the SW-OLS procedure (Derksen and Keselman, 1992).

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Table 1. *Properties of the various techniques (*: not used previously in this context).*

Technique	Dependence	Estimation	Hyperparameters	Criterion
RWD	temporal	MLE	–	–
ARIMA	temporal	MLE	p_i, q_i $(i = 1, 2, \ldots, N)$	BIC
DPLM*	temporal	GMM	p	Significance
VAR	temporal and cross-sectional	LASSO	p	CV-MSE
STARIMA*	spatiotemporal	KF	p, λ_ℓ $(\ell = 1, 2, \ldots, p)$, q, μ_ℓ $(\ell = 1, 2, \ldots, q)$	BIC
SDPLM*	spatiotemporal	QML	weights ∈ {contiguity, inverse-distance}, time lag ∈ {0, 1}, space-time lag ∈ {0, 1}, Lee–Yu transf. ∈ {0, 1}	Quasi-BIC
ESTF*	spatiotemporal	SW-OLS or LASSO	weights ∈ {contiguity, inverse-distance}, spec ∈ {contemporaneous, lagged}, selection ∈ {SW-OLS, LASSO}	BIC or DR

models, whereas the second one, denoted by ENS2, computes a trimmed mean excluding the lowest and highest predictions.

Given the time series k_{it} for each country from the LC model, univariate unit root and second-generation panel unit root tests (Hurlin and Mignon, 2007) may help determine their orders of integration. The latter account for cross-sectional dependence, which is central to our approach. Specifically, for ARIMA, we use augmented Dickey–Fuller (ADF) tests to determine country-specific orders d_i such that the differenced series have no unit roots. For all forecasting techniques other than RWD and ARIMA, we use the covariate-augmented Dickey–Fuller (CADF) test (Costantini and Lupi, 2013), a second-generation panel unit root test, to determine a uniform order of integration $d_i = d$ $(i = 1, 2, \ldots, N)$ across countries.

For the LL model, following Li and Lee (2005), we recommend forecasting the common trend K_t using an RWD process estimated by MLE. For the country-specific k_{it}, we set $d_i = 0$ and $d = 0$ across all countries and forecasting techniques to ensure coherence.

In the remainder of our paper, we will use the word "model" for the LC and LL models, and "technique" for the forecasting techniques presented in this subsection.[5]

3.4. Robust selection[6]

We propose dividing the dataset into a training set, consisting of the initial T_{train} years of available data, and a test set, comprising the remaining T_{test} years, with $T = T_{\text{train}} + T_{\text{test}}$, thereby spanning the entire observation period. Additionally, because model performance may be sensitive to the location of the train–test split, we propose a robust selection procedure involving repeated estimation across multiple training horizons $T^{(h)}_{\text{train}}$, with $T^{(h)}_{\text{test}} = T - T^{(h)}_{\text{train}}$ $(h = 1, 2, \ldots, H)$.

[5]Under this terminology, DPLM and SDPLM are techniques, despite what their names suggest.

[6]We avoid the term "model selection," as "selection of model–technique combinations" would be more accurate but unnecessarily cumbersome.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Forecast accuracy of forecasting technique j under model ℓ (LC or LL in our case) in country i given train–test split h is then evaluated using the mean squared error (MSE) of the forecasted logarithmic mortality rates:

$$\mathrm{MSE}_{i,j,\ell}^{(h)} \triangleq \frac{1}{XT_{\mathrm{test}}^{(h)}} \sum_{x\in\mathcal{X}} \sum_{t=T_{\mathrm{train}}^{(h)}+1}^{T} \left(\ln m_{ixt} - \ln \hat{m}_{ixt}^{(j,\ell)}\right)^2,$$

where $\hat{m}_{(\cdot)}^{(j,\ell)}$ denotes mortality rates forecasted using technique j with its optimal hyperparameters under model ℓ.

As it is preferable to assess model performance relative to the best-performing procedure rather than against a perfect fit or another hypothetical benchmark, for each split, we define the underperformance scores

$$\mathrm{U}_{i,j,\ell}^{(h)} \triangleq \mathrm{MSE}_{i,j,\ell}^{(h)} - \min_{j,\ell} \mathrm{MSE}_{i,j,\ell}^{(h)},$$

which measure how much worse forecasting technique j under model ℓ performs for country i and training horizon h, relative to the best-performing combination.

Model underperformance can then be assessed independently from the location of the train–test split by measuring average or maximum underperformance (AU or MU) scores across the splits, with the maximum criterion being more conservative and reflecting the minimax principle from decision theory:

$$\mathrm{AU}_{i,j,\ell} \triangleq \frac{1}{H} \sum_{h=1}^{H} \mathrm{U}_{i,j,\ell}^{(h)},$$

$$\mathrm{MU}_{i,j,\ell} \triangleq \max_{h=1,2,\ldots,H} \mathrm{U}_{i,j,\ell}^{(h)}.$$

These measures can then be summarized by their medians across all countries, enabling the robust comparison of model–technique combinations:

$$\mathrm{MAU}_{j,\ell} \triangleq \operatorname*{median}_{i=1,2,\ldots,N}(\mathrm{AU}_{i,j,\ell}),$$

$$\mathrm{MMU}_{j,\ell} \triangleq \operatorname*{median}_{i=1,2,\ldots,N}(\mathrm{MU}_{i,j,\ell})$$

MAU captures median average underperformance relative to the best-performing combination, while MMU reflects median worst-case underperformance across train–test splits.

3.5. Stochastic forecasts

Stochastic forecasts may be generated by incorporating some or all of the following four sources of uncertainty (Cairns 2000):

1. *Measurement uncertainty* arises from the random error terms ε_{ixt} in the LC and LL log-mortality equations. It captures fluctuations in observed log-mortality rates that are not explained by the systematic components of the models.
2. *Process uncertainty* arises from the inherent randomness in the future evolution of the stochastic time series components k_{it} of the models.
3. *Parameter uncertainty* results from potential errors in the estimation of the model parameters a_{ix}, b_{ix}, k_{it}, B_x, and K_t based on finite historical data.
4. *Model uncertainty* reflects the possibility that the chosen model structure (e.g., LC–RWD) may be misspecified, or that alternative models could better represent the underlying mortality dynamics.

Following the approach recommended by Brouhns *et al*. (2005), we address parameter uncertainty by generating synthetic training datasets of bootstrapped death counts $\{D_{ixt}^{(r)}\}_{r=1}^{R}$ drawn independently from

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Poisson distributions with means $\lambda_{ixt} \triangleq D_{ixt}$, where R is the number of bootstrap samples. This yields bootstrapped estimates $a_{ix}^{(r)}$, $b_{ix}^{(r)}$, and $k_{it}^{(r)}$ separately for the LC and LL models and $B_x^{(r)}$ and $K_t^{(r)}$ for the LL model.

We incorporate process uncertainty via resampling with replacement from the past residuals of the fitted time series model for the given country:[7]

$$\Delta^{(d_i)} k_{i,T_{\text{train}}+t}^{(r)} = \Psi_{it}^{(r)} + \xi_{it}^{(r)} \quad (t \geq 1,\ r = 1, 2, \ldots, R),$$

and additionally, for the LL model, we apply

$$\Delta K_{T_{\text{train}}+t}^{(r)} = \Theta^{(r)} + \eta_t^{(r)} \quad (t \geq 1,\ r = 1, 2, \ldots, R),$$

where $\Psi_{it}^{(r)} \triangleq \mathbb{E}_{T_{\text{train}}}\left(\Delta^{(d_i)} k_{i,T_{\text{train}}+t}^{(r)}\right)$ with $\mathbb{E}_{T_{\text{train}}}(\cdot)$ denoting a conditional expectation[8] based on bootstrapped samples up to T_{train}, $\Theta^{(r)}$ is the drift term of the RWD process of the common mortality index K_t, and $\xi_{it}^{(r)}$ and $\eta_t^{(r)}$ are randomly drawn with replacement from training-set residuals of $\Delta^{(d_i)} k_{it}$ and ΔK_t ($t = 1, 2, \ldots, T_{\text{train}}$). We draw from the residuals rather than from normal distributions, as the normality assumption may not hold and methods outside the ARIMA family do not rely on it.

We account for measurement uncertainty by adding bootstrapped residuals $\varepsilon_{i,x,T_{\text{train}}+t}^{(r)}$ drawn from past values $\{\hat{\varepsilon}_{ixt}\}_{t=1}^{T_{\text{train}}}$ to the systematic components of the log-mortality equations (with $B_x^{(r)} \triangleq K_t^{(r)} \triangleq 0$ in the case of the LC model):

$$\ln m_{i,x,T_{\text{train}}+t}^{(r)} \triangleq \ln m_{i,x,T_{\text{train}}}^{(r)} + b_{ix}^{(r)}\left(k_{i,T_{\text{train}}+t}^{(r)} - k_{i,T_{\text{train}}}^{(r)}\right) + B_x^{(r)}\left(K_{T_{\text{train}}+t}^{(r)} - K_{T_{\text{train}}}^{(r)}\right) + \varepsilon_{i,x,T_{\text{train}}+t}^{(r)}.$$

Model uncertainty could be accounted for within a fully Bayesian framework; however, we leave this aspect outside the scope of the present study, as subjective priors would introduce a new layer of uncertainty, and model averaging with BIC-derived weights is not feasible as some forecasting techniques we have proposed are not estimated by maximum likelihood. Instead, we apply post hoc adjustment and ensembling to mitigate this limitation.

To reduce sampling bias, we apply mean calibration, that is, a pointwise shift to the bootstrapped trajectories so that their sample mean equals the expected trajectory $\hat{m}_{i,x,T_{\text{train}}+t}$ computed under the given model–technique combination at the end of the training period:

$$\ln \tilde{m}_{i,x,T_{\text{train}}+t}^{(r)} \triangleq \ln m_{i,x,T_{\text{train}}+t}^{(r)} + \ln \hat{m}_{i,x,T_{\text{train}}+t} - \frac{1}{R}\sum_{r=1}^{R} \ln m_{i,x,T_{\text{train}}+t}^{(r)}.$$

Then, we construct bootstrapped prediction intervals with confidence level $1 - \tau$ for the log-mortality rates $\ln m_{i,x,T_{\text{train}}+t}$ as

$$\left(\pi_{i,x,T_{\text{train}}+t}^{(L,\tau)}, \pi_{i,x,T_{\text{train}}+t}^{(U,\tau)}\right) \triangleq \left(Q_{\frac{\tau}{2}}\left(\{\ln \tilde{m}_{i,x,T_{\text{train}}+t}^{(r)}\}_{r=1}^{R}\right),\quad Q_{1-\frac{\tau}{2}}\left(\{\ln \tilde{m}_{i,x,T_{\text{train}}+t}^{(r)}\}_{r=1}^{R}\right)\right),$$

where $Q_\tau(\cdot)$ denotes the empirical τ-quantile of the resampled estimates.

To correct potentially inaccurate interval coverage (e.g., due to model uncertainty or structural breaks), we apply a post hoc scaling adjustment by solving for scaling factors ν_t that achieve the empirical nominal coverage probability $1 - \tau$ on a validation set of length T_{val}, for each model–technique combination and forecast horizon $t = 1, 2, \ldots, T_{\text{val}}$:

$$\frac{1}{N\,|\mathcal{X}|}\sum_{i=1}^{N}\sum_{x\in\mathcal{X}} \mathbb{1}_{\left(\pi_{i,x,T_{\text{train}}+t}^{*(L,\tau)} \leq \ln m_{i,x,T_{\text{train}}+t} \leq \pi_{i,x,T_{\text{train}}+t}^{*(U,\tau)}\right)} = 1 - \tau,$$

[7] We suppress the dependence on the model–technique combination to simplify the notation. Past residuals are computed from the original parameter estimates and not the bootstrapped ones.

[8] We are avoiding the sigma-algebra notation.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

where $\mathbb{1}_{(\cdot)}$ are indicator functions, and

$$\left(\pi^{*(L,\tau)}_{i,x,T_{\text{train}}+t}, \pi^{*(U,\tau)}_{i,x,T_{\text{train}}+t}\right) \triangleq \left(\pi^{(L,\tau)}_{i,x,T_{\text{train}}+t} - \frac{\nu_t - 1}{2}\left(\pi^{(U,\tau)}_{i,x,T_{\text{train}}+t} - \pi^{(L,\tau)}_{i,x,T_{\text{train}}+t}\right),\right.$$
$$\left.\pi^{(U,\tau)}_{i,x,T_{\text{train}}+t} + \frac{\nu_t - 1}{2}\left(\pi^{(U,\tau)}_{i,x,T_{\text{train}}+t} - \pi^{(L,\tau)}_{i,x,T_{\text{train}}+t}\right)\right).$$

We assess the quality of these prediction intervals on a separate evaluation period of length T_{eval}, with the original test period partitioned as $T_{\text{val}} + T_{\text{eval}} = T_{\text{test}}$, using average interval scores (Gneiting and Raftery, 2007) by country, computed as

$$\begin{aligned}\text{AIS}_{i,1-\tau} \triangleq \frac{1}{|\mathcal{X}|\ T_{\text{eval}}} \sum_{x\in\mathcal{X}} \sum_{t=1}^{T_{\text{eval}}} \Big[& \pi^{*(U,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t} - \pi^{*(L,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t} \\ & + \frac{2}{\tau} \max\left(\pi^{*(L,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t} - \ln m_{i,x,T_{\text{train}}+T_{\text{val}}+t}, 0\right) \\ & + \frac{2}{\tau} \max\left(\ln m_{i,x,T_{\text{train}}+T_{\text{val}}+t} - \pi^{*(U,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t}, 0\right)\Big],\end{aligned}$$

where we sample from the first $T_{\text{train}} + T_{\text{val}}$ years, simulate for the remaining T_{eval} years, use mean calibration to expected trajectories estimated at the end of the validation period, and apply the scaling factors ν_t previously estimated on the validation period.

Finally, we compute median average interval scores across countries as a robust summary of prediction interval quality across model–technique combinations:

$$\text{MAIS}_{1-\tau} \triangleq \underset{i=1,2,\ldots,N}{\text{median}}\ \text{AIS}_{i,1-\tau},$$

and assess out-of-sample coverage using empirical coverage probabilities:

$$\text{ECP}_{1-\tau} \triangleq \frac{1}{N\,|\mathcal{X}|\,T_{\text{eval}}} \sum_{i=1}^{N} \sum_{x\in\mathcal{X}} \sum_{t=1}^{T_{\text{eval}}} \mathbb{1}_{\left(\pi^{*(L,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t} \le \ln m_{i,x,T_{\text{train}}+T_{\text{val}}+t} \le \pi^{*(U,\tau)}_{i,x,T_{\text{train}}+T_{\text{val}}+t}\right)}.$$

This procedure could be made even more robust by using multiple train–test splits, following the analogy of the robust selection pipeline described in Subsection 3.4.

4. Empirical analysis

We used the R programming language (R Core Team, 2025) for our computations.[9] Figure 1 illustrates the steps of the analysis, from data collection to robust selection and the evaluation of the prediction intervals.

4.1. Data collection and preparation

We obtained our dataset from the Human Mortality Database (HMD, 2025). We collected unisex data on the number of deaths and exposures in all European countries for which data were available for the $T = 60$ years between 1960 and 2019, categorized by country, age, and calendar year, and excluded the remaining countries. We did not use data for ages above 99 years due to missing data and other potential data quality issues. Consequently, our dataset comprised the $N = 22$ countries depicted in Figure 2.

We used data only up to 2019 to avoid testing on the years of the COVID-19 pandemic, which would have led to a bias toward models that tend to overestimate mortality. Assuming that COVID-19 was a one-off event, the pandemic years could be included after subtracting COVID-19 deaths from total counts

[9] And in particular, the packages `StMoMo` (Villegas *et al*., 2015), `MultiMoMo` (Robben, 2021), `punitroots` (Kleiber and Lupi, 2011), `forecast` (Hyndman and Khandakar, 2008), `sparsevar` (Vazzoler, 2021), `plm` (Croissant and Millo, 2008), `starma` (Cheysson, 2016), and `SDPDmod` (Simonovska, 2025).

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

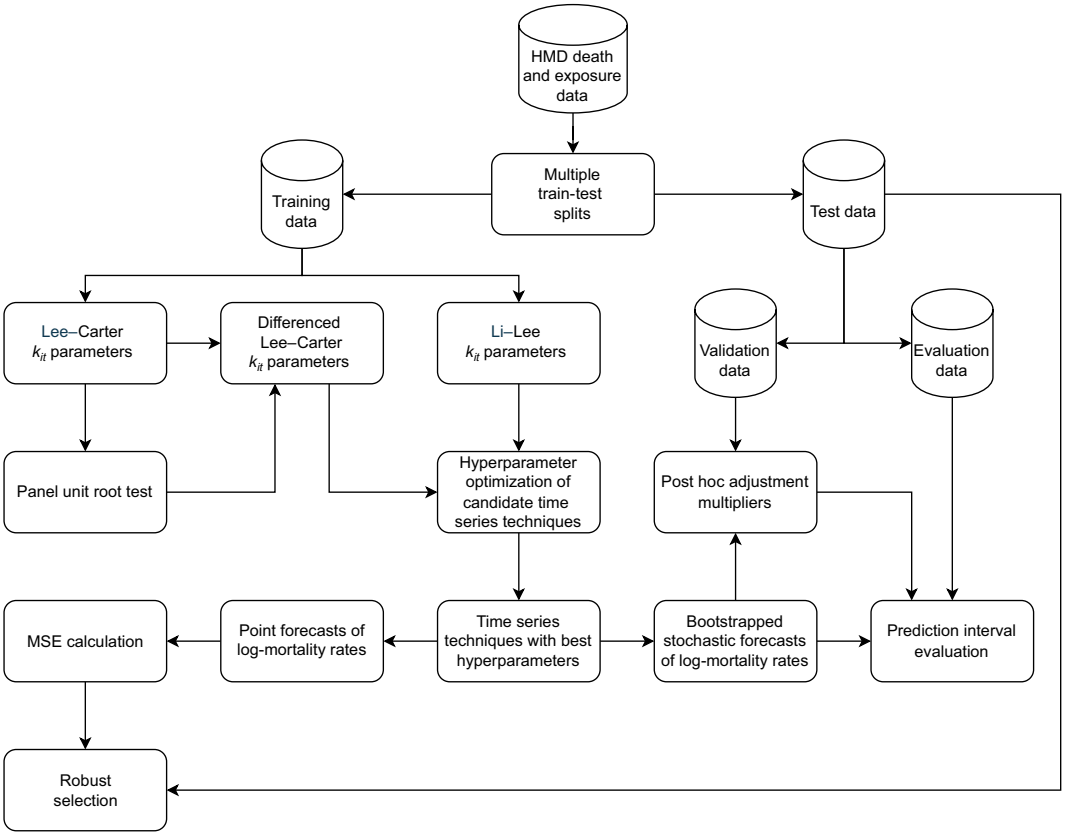


Figure 1. *Analysis workflow.*

or by interpolating data for the years of the pandemic using rates from neighboring years. However, a significant problem with the subtraction approach is that criteria for recording the cause of death and the extent of testing varied widely by country, while the second approach is not yet viable as data for years following the end of the pandemic are not yet available for some countries in the Human Mortality Database. Due to these constraints, we opted to perform our analysis on this truncated period.

By default, we used the first $T_{\text{train}} = 45$ years, up to 2004, as training data, the remaining $T_{\text{test}} = 15$ years as test data, and split the test period into a validation period of $T_{\text{val}} = 8$ years and an evaluation period of $T_{\text{eval}} = 7$ years.

4.2. Model estimation and interpretation

We estimated the parameters of the LC and LL models on data from all $N = 22$ countries. Table 2 displays the spatial and spatiotemporal Moran's I statistics of the differenced mortality indices Δk_{it} of the LC and LL models using contiguity and inverse-distance spatial weights[10]. These results indicate that there is significant spatial and spatiotemporal autocorrelation in mortality improvements in both the LC and LL models, which motivates the application of spatiotemporal forecasting techniques, and contiguity weighting yields stronger spatiotemporal autocorrelation than inverse-distance weighting.

[10]This descriptive step is not shown in Figure 1. Averages $\frac{1}{T-1}\sum_{t=2}^{T}\Delta k_{it}$ across time periods are used in the computation of spatial Moran's I. Contemporaneous spatiotemporal autocorrelation is not significant at the 5% level for the LL model, but it is significant at the 10% level using contiguity weighting.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press



Figure 2. *Map of the countries included in the analysis with their ISO-2 codes.*

We performed LISA clustering at the 5% significance level on the means of Δk_{it} from the LC and LL models by country. We used contiguity weighting in this step, based on the findings reported in Table 2. LISA identified two spatial clusters for the LC model: France and Switzerland and the Baltic states of Estonia, Latvia, and Lithuania, both showing high mean differences in the mortality index. For the LL model, the first significant cluster consisted of France and Germany, and the second of the three Baltic states, again both with high mean differences.

In the ESTFs, the spatiotemporal eigenvectors with the largest absolute coefficients revealed clear spatial patterns. In the LC model, they separated the United Kingdom and Ireland, as well as the Scandinavian countries of Finland, Norway, and Sweden, from the rest of Europe under both Stepwise and LASSO selections. In the LL model, they again isolated these same clusters under LASSO, while under stepwise the three Baltic states were separated from the rest of Europe. This has a natural geographic interpretation, as the separated groups of countries are divided by bodies of water from the rest of the countries in the analysis.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Table 2. *Spatial and spatiotemporal Moran's I statistics of the differenced mortality indices Δk_{it} from the LC and LL models under alternative spatial weighting schemes (p-values in parentheses).*

	Contiguity weights			Inverse-distance weights		
Model	I	$I^{(C)}_{\text{space-time}}$	$I^{(L)}_{\text{space-time}}$	I	$I^{(C)}_{\text{space-time}}$	$I^{(L)}_{\text{space-time}}$
LC	0.709	0.097	−0.150	0.273	0.062	−0.147
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)
LL	0.774	0.039	−0.057	0.326	0.023	−0.090
	(0.000)	(0.093)	(0.005)	(0.000)	(0.273)	(0.000)

4.3. Forecasting techniques and hyperparameter optimization

We assessed the stationarity of the LC mortality indices using the CADF test, which yielded the p-values 0.847 on the original panel data and 0.000 after differencing, indicating the successful removal of unit roots, and $d = 1$ as the uniform order of integration.

For both the LC and LL models, we estimated the parameters and optimized the hyperparameters of the time series forecasting techniques as we outlined in Table 1. In all cases, the contiguity spatial weights matrix outperformed the version constructed from inverse distances between capital cities, in line with the findings from Table 2.

4.4. Point forecasts

We forecasted the country-specific mortality indices of the LC and LL models and the log-mortality rates using all forecasting techniques presented in Subsection 3.3, each with its optimal set of hyperparameters, using observed jump-off rates to enhance forecast accuracy, as recommended by Lee and Miller (2001), and applied an RWD for the common mortality index of the LL model[11] The first row of Figure 3 illustrates the LC model predictions for Norway across selected ages, and the second row shows the corresponding coherent LL forecasts.[12]

4.5. Robust selection

Following forecasting, we applied the robust selection procedure described in Subsection 3.4, using the years 2001 through 2007 as the endpoints of the training period (i.e., $T_{\text{train}} \in \{42, 43, 44, 45, 46, 47, 48\}$) to evaluate the predictive performance of the combinations of the LC and LL models and the seven forecasting techniques and the two ensembles. The numbers and ISO-2 codes of the countries in which each technique performs best under each model, based on AU and MU, are summarized in Table 3.[13]

Within the LC framework, LC–ESTF emerges as the strongest overall performer, with 6 and 7 wins under AU and MU, respectively. LC–SDPLM also performs well, achieving 4 wins under both criteria. Overall, the spatiotemporal techniques account for 17 and 14 wins under AU and MU, respectively, out of the 22 populations.

Within the LL model, LL–DPLM stands out with 6 and 5 wins under AU and MU, respectively, making it the best performer across criteria. LL–STARIMA follows with 3 and 5 wins. Spatiotemporal techniques are best in 9 and 11 countries under AU and MU, respectively—less frequently than in the LC framework. This is expected, as the common trend in the LL model already captures much of the cross-country dependence, and the spatiotemporal techniques can only leverage residual dependence beyond this.

[11] Since the SW-OLS and LASSO methods rely on different selection criteria and are not directly comparable, we applied an unweighted average ensemble of the best SW-OLS and LASSO specifications for forecasting in both the LC and LL cases.

[12] Ensemble forecasts are not displayed to avoid visual clutter, but can be interpreted as (trimmed) average paths.

[13] In the case of the LL model, RWD, ARIMA, and STARIMA reduce to RW, ARMA, and STARMA, respectively, since no differencing is applied to ensure coherence.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Table 3. *Number of wins by technique within each model, and ISO-2 codes of countries where they won, based on AU and MU. Asterisks denote newly proposed techniques in this context, and highest win counts by column are typeset in bold underlined font.*

	LC		LL	
Technique	AU	MU	AU	MU
RWD	1 (FR)	5 (AT,CH,FR,IE,PL)	3 (GB,IT,SE)	3 (CH,GB,IT)
ARIMA	1 (EE)	1 (EE)	1 (ES)	0
DPLM*	1 (HU)	1 (HU)	**6** (DK,EE, HU,IE,NO,PT)	**5** (DK,EE,IE,NO,PT)
VAR	2 (BE,IT)	1 (BE)	3 (CZ,LV,PL)	3 (CZ,HU,PL)
STARIMA*	1 (PL)	0	3 (DE,FR,NL)	5 (AT,BE,DE,FR,NL)
SDPLM*	4 (AT,DE,GB,PT)	4 (DE,GB,IT,PT)	0	0
ESTF*	**6** (CH,DK, ES,FI,NL,NO)	**7** (CZ,DK, ES,FI,LV,NL,NO)	2 (AT,FI)	2 (FI,SE)
ENS1*	4 (CZ,IE,LV,SK)	1 (SK)	1 (BE)	3 (ES,LT,LV)
ENS2*	2 (LT,SE)	2 (LT,SE)	3 (CH,LT,SK)	1 (SK)

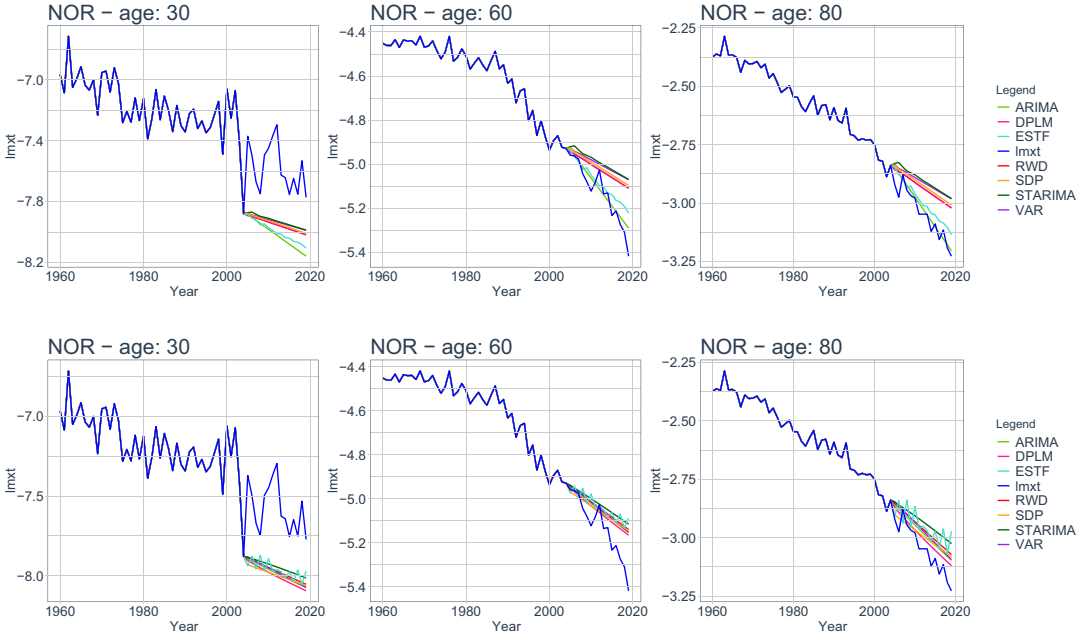


Figure 3. *Point projections of log-mortality for selected ages in Norway from the LC (first row) and LL (second row) models.*

Spatiotemporal techniques often outperform generic VAR as they rely on more parsimonious dependence structures.

We provide another comprehensive summary of our results in Figure 4, where the red bars display MAU, the red and blue bars combined correspond to values of the more cautious MMU, and the left and right panels refer to the LC and LL models, respectively. This comparison is chiefly relevant if the same model–technique combination is applied in all countries. Otherwise, country-specific combinations are clearly superior.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

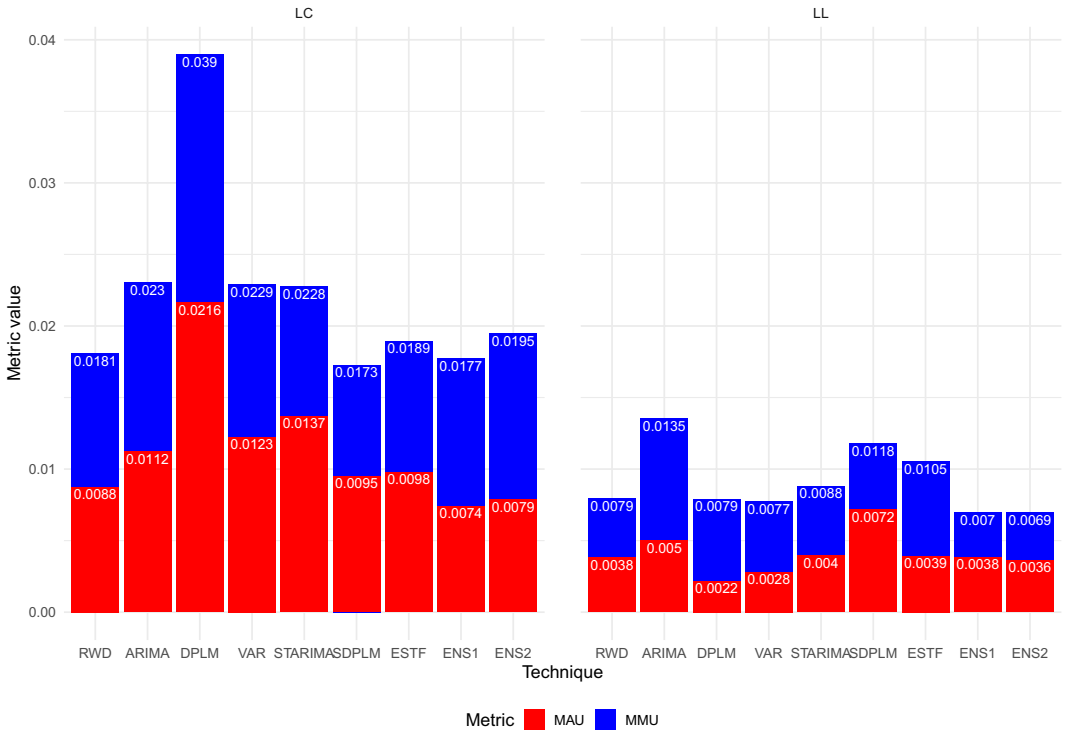


Figure 4. *Comparison of MAU and MMU values across models and techniques (MMU is the sum of the heights of the red and blue columns).*

As shown in Figure 4, the LL–DPLM model achieves the lowest overall MAU, but is slightly outperformed by the ensembles LL–ENS1 and LL–ENS2 in terms of the more conservative MMU. Within the LC model, the ensembles LC–ENS1 and LC–ENS2 achieve the lowest MAU, while LC–SDPLM is the best in terms of MMU.

Notably, ensembles achieve excellent overall MAU and MMU, and LL-based forecasts outperform LC-based ones, as they incorporate information on the common trend distilled from a larger set of data in addition to country-specific patterns, while remaining coherent.

Overall, on the set of countries, ages and calendar years under consideration, LL–DPLM, LC–ESTF, LL–VAR, and LL–STARIMA emerge as strong performers, with notable variation by country. The widely used RWD and ARIMA specifications are clearly outperformed by other procedures in both the LC and LL frameworks.

4.6. Stochastic forecasts

We performed the procedure presented in Subsection 3.5 using $T_{\text{train}} = 45$, $T_{\text{val}} = 8$, and $T_{\text{eval}} = 7$ years, $R = 100$ bootstrap samples, and a confidence level of $1 - \tau = 0.9$. For the sake of illustration, the first and second rows of Figure 5 display mean-calibrated, post hoc adjusted, bootstrapped 90% prediction intervals of log-mortality rates for selected ages in Norway, based on the LC–ESTF and LL–ESTF frameworks.[14]

As ENS1 and ENS2 lack their own mortality index series, we computed their prediction intervals as quantiles of the means and trimmed means, respectively, of the bootstrapped samples of the log-mortality rates from the seven individual techniques and applied all subsequent modeling steps in the same way as to the original techniques.

[14]Prediction uncertainty at age 30 is dominated by measurement uncertainty, which does not accumulate over time.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Table 4. *Number of countries in which each forecasting technique yields the lowest average 90% interval score, and ISO-2 codes of countries where they won. Asterisks denote newly proposed techniques, and highest win counts by model are typeset in bold underlined font.*

Technique	LC	LL
RWD	1 (IE)	0
ARIMA	1 (LT)	2 (ES, IT)
DPLM*	3 (DK, PL, SK)	**7** (BE, CH, CZ, HU, NO, PL, PT)
VAR	1 (DE)	2 (IE, NL)
STARIMA*	4 (EE, FI, NO, PT)	1 (GB)
SDPLM*	**6** (AT, FR, GB, HU, IT, LV)	1 (EE)
ESTF*	2 (BE, CZ)	6 (AT, DK, FI, LT, LV, SK)
ENS1*	3 (CH, ES, SE)	1 (SE)
ENS2*	1 (NL)	2 (DE, FR)

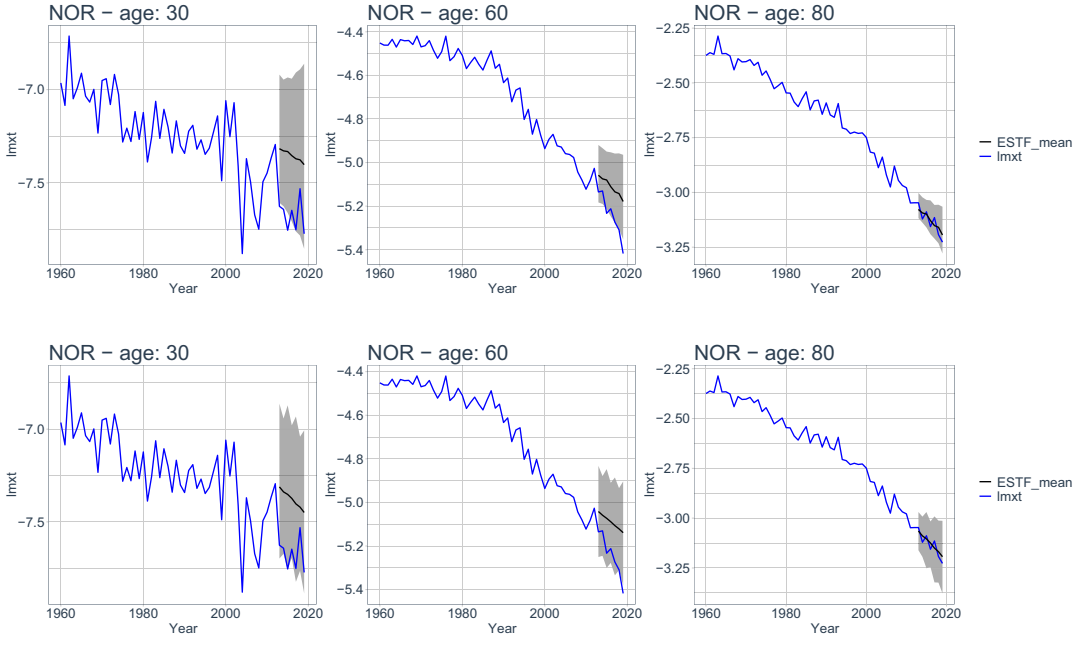


Figure 5. *Bootstrapped 90% prediction intervals for selected ages in Norway from the LC–ESTF (first row) and LL–ESTF (second row) frameworks.*

Table 4 reports the number and list of countries in which each technique achieves the lowest average 90% interval score under the LC and LL models. Under LC, SDPLM performs best, winning in six countries, followed by STARIMA, DPLM and ENS1, while the classical benchmarks RWD and ARIMA win in one country each. Under LL, DPLM attains the highest number of wins (seven), closely followed by ESTF (six), whereas ARIMA wins in only two countries and RWD in none.

Table 5 reports median average 90% interval scores and empirical coverage probabilities across forecasting techniques. Under the LC model, ENS1 achieves the lowest median interval score, closely followed by ENS2 and SDPLM, while classical benchmarks exhibit higher scores. Under the LL model, ESTF attains the lowest MAIS, with DPLM performing similarly well. Empirical coverage probabilities

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

Table 5. *Median average 90% interval scores and empirical coverage probabilities by forecasting technique under the LC and LL models. Asterisks denote newly proposed techniques, and lowest MAIS by model are typeset in bold underlined font.*

	LC		LL	
Technique	$MAIS_{0.90}$	$ECP_{0.90}$	$MAIS_{0.90}$	$ECP_{0.90}$
RWD	0.997	0.864	1.029	0.929
ARIMA	0.946	0.845	0.898	0.888
DPLM*	0.878	0.894	0.827	0.913
VAR	0.865	0.900	0.896	0.909
STARIMA*	0.892	0.913	0.883	0.908
SDPLM*	0.865	0.896	0.931	0.922
ESTF*	0.897	0.878	**0.827**	0.934
ENS1*	**0.845**	0.893	0.885	0.916
ENS2*	0.848	0.897	0.873	0.899

are generally close to the nominal level across methods, while LL-based frameworks tend to slightly overcover.

4.7. Robustness check

To assess the robustness of our forecasts, we compared period life expectancy at birth at the end of the test period in 2019, computed using the various forecasting techniques estimated on the training period 1950–2004, for all $N = 22$ countries and both the LC and LL models. This choice was motivated by four considerations: (i) life expectancy at birth is an easily interpretable measure with an intuitive scale; (ii) it summarizes mortality across all ages into a single indicator, allowing for a parsimonious comparison; (iii) by the end of the test period, forecasts from different techniques are expected to have diverged sufficiently to allow for an informative comparison; and (iv) unlike cohort life expectancies, it does not require re-estimating the models and techniques on the full dataset and forecasting an additional 100 years, which would be more speculative as well as computationally demanding.

By careful individual examination of the projected mortality trajectories by country, model, and technique, we identified a few anomalies. In the Baltic states of Estonia, Latvia, and Lithuania, the LC mortality index series k_{it} increases over time, accompanied by negative values of b_{ix} at childhood and some other ages and positive values at middle ages. This pattern captures the idiosyncratic mortality experience of these countries, where middle-aged mortality stagnated or worsened during the late Soviet period (Krūmiņš and Zvidriņš 1992), while other age groups experienced improving mortality. STARIMA introduces decreases into k_{it} over several periods, which leads to projected increases in child mortality and decreases in middle-aged mortality in the forecasts for all three countries. A similar inversion is observed for ESTF in Estonia and Latvia. As this behavior is clearly a modeling artifact, we exclude LC–STARIMA for all three countries and LC–ESTF for Estonia and Latvia. This inversion does not affect the robust selection or the evaluation of prediction intervals presented in Subsections 4.5 and 4.6, as these techniques were not selected for the affected countries and the reported medians are robust to outliers.

The results are displayed in Figure 6, which shows little variation across techniques. The standard deviations across techniques, computed separately for each country and model, range from 0.01 to 1.06 years, with a median of 0.44 years, corroborating the robustness of our forecasts.

5. Conclusion

Our paper introduces a suite of multipopulation forecasting techniques for the Lee–Carter and Li–Lee country-specific mortality indices. Traditionally, most authors apply RWD or ARIMA processes. By

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

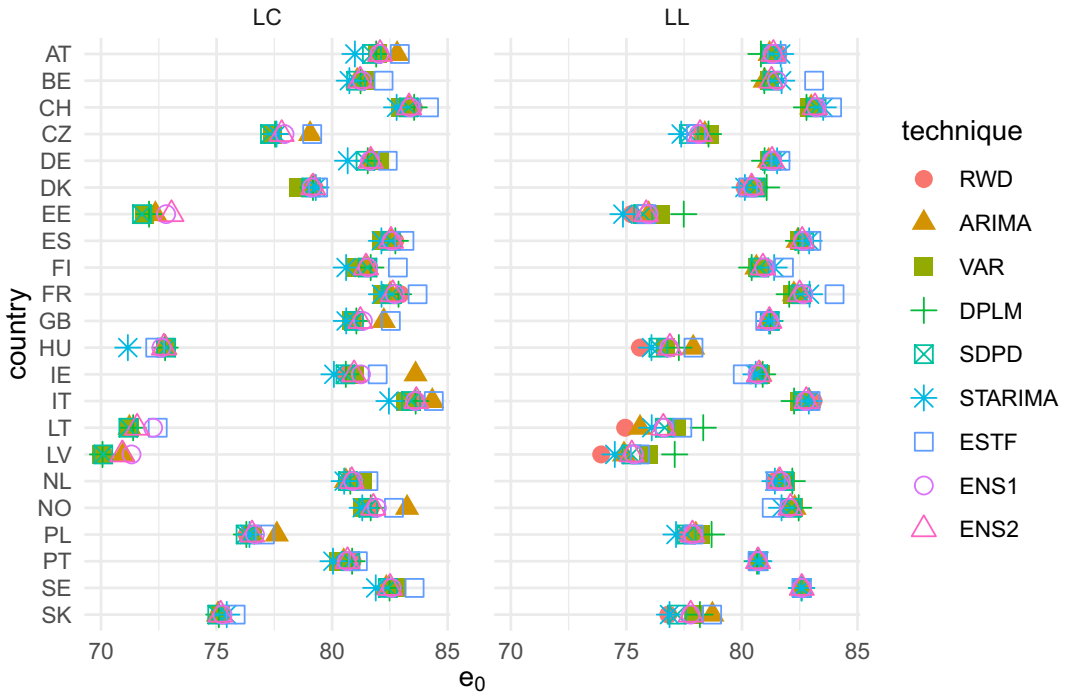


Figure 6. *Comparison of projected period life expectancy at birth in 2019 across countries, models, and techniques.*

contrast, we have also explored various spatial and panel econometric techniques. We applied these to mortality data from 22 countries, spanning the years 1960 to 2019 and covering ages 0–99. We estimated both the Lee–Carter and Li–Lee models on training periods ending in years ranging from 2001 through 2007 and forecasted their mortality indices using various techniques, with coherent forecasts in the Li–Lee frameworks. We then analyzed the forecast accuracy based on our proposed robust selection procedure using test sets spanning the remaining years through 2019.

Based on multiple robust metrics (number of wins by average and maximum underperformance relative to the best model–technique combination, as well as median average underperformance and maximum average underperformance), the combinations of the Li–Lee and the dynamic panel linear models, the Lee–Carter model and the eigenvector spatiotemporal filter, the Li–Lee model and VAR, and the Li–Lee model and spatiotemporal ARIMA emerge as particularly strong performers in point forecasts, with notable variation by country. Out of these frameworks, we are the first to propose all but the third one in the literature to the best of our knowledge.

We propose a mean-calibrated, post hoc adjusted bootstrapping procedure to generate prediction intervals with correct nominal coverage. In interval forecasting, the newly introduced techniques clearly outperform classical benchmarks, with the Li–Lee model combined with the eigenvector spatiotemporal filter or the dynamic panel linear model achieving the lowest median average interval scores.

We examined projected period life expectancies at birth in 2019 across the various model–technique combinations and found that the forecasts were sufficiently robust.

In summary, our results provide evidence that spatial and panel econometric techniques for forecasting the Lee–Carter and Li–Lee mortality indices can be superior to more established approaches and applying them may yield measurable improvements in accuracy in several populations.

Our study is limited in that it only considers the Lee–Carter and Li–Lee models as the two most prominent general and coherent multipopulation stochastic mortality models, and the Random Walk

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

with Drift for the common mortality index in the Li–Lee model, and it does not delve into Bayesian techniques or more sophisticated ensembles for forecasting the mortality indices.

Despite the built-in robustness, the results may change if the set of countries, ages, and calendar years considered in the analysis is modified.

A further limitation is the use of geographical distance, as countries' mortality experiences may be more closely aligned through cultural proximity. This could be accounted for using spatial weights matrices based on Hofstede's cultural dimensions theory (Hofstede, 1984), among other possible spatial weighting schemes.

Additionally, we have omitted the years of the COVID-19 pandemic from the testing period out of necessity, and our analysis is currently limited by the absence of data for the post-pandemic years.

Finally, it would be worthwhile to apply the outlined methodology to more granular data—for example, at the levels of counties, states, or other subnational units—where even more meaningful spatial relationships may emerge.

The proposed techniques lend themselves to practical applications by actuaries and demographers, facilitating more accurate forecasting of mortality patterns and enhancing the design of demographic projections, pension systems, and life insurance products.

Data and code availability. The data and code that support the findings of this study are available from the authors upon request.

Competing interests. The authors declare none.

References

Álvaro-Meca, A., *et al.* (2011) The convergence of European mortality in both sexes in the near future: A spatio-temporal approach. *Population Review*, **50**(2), https://doi.org/10.1353/prv.2011.0018.

Anselin, L. (1995) Local indicators of spatial association–LISA. *Geographical Analysis*, **27**(2), 93–115, https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

Antonio, K., Devriendt, S., de Boer, W., de Vries, R., De Waegenaere, A., Kan, H.-K., Kromme, E., Ouburg, W., Schulteis, T., Slagter, E., van der Winden, M., van Iersel, C. and Vellekoop, M. (2017) Producing the Dutch and Belgian mortality projections: A stochastic multi-population standard. *European Actuarial Journal*, **7**(2), 297–336, https://doi.org/10.1007/s13385-017-0159-x.

Arellano, M. and Bond, S. (1991) Some tests of specification for Panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), 277–297, https://doi.org/10.2307/2297968.

Astuti, A.M., Setiawan, Z.I. and Purnomo, J.D.T. (2020) A review of panel data on spatial econometrics models. *Journal of Physics: Conference Series*, **1490**(1), 012032, https://doi.org/10.1088/1742-6596/1490/1/012032.

Basellini, U., Camarda, C.G. and Booth, H. (2023) Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting*, **39**(3), 1033–1049, https://doi.org/10.1016/j.ijforecast.2022.11.002.

Bennett, J.E., Li, G., Foreman, K., Best, N., Kontis, V., Pearson, C., Hambly, P. and Ezzati, M. (2015) The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. *Lancet*, **386**(9989), 163–170, https://doi.org/10.1016/S0140-6736(15)60296-3.

Blundell, R. and Bond, S. (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**(1), 115–143, https://doi.org/10.1016/S0304-4076(98)00009-8.

Booth, H., Maindonald, J. and Smith, L. (2002) Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, **56**(3), 325–336, https://doi.org/10.1080/00324720215935.

Brouhns, N., Denuit, M. and Keilegom, I.V. (2005) Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, **2005**(3), 212–224, https://doi.org/10.1080/03461230510009754.

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, https://doi.org/10.1016/S0167-6687(02)00185-3.

Cairns, A.J. (2000) A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**(3), 313–330, https://doi.org/10.1016/S0167-6687(00)00055-X.

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, https://doi.org/10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A. and Balevich, I. (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, **13**(1), 1–35, https://doi.org/10.1080/10920277.2009.10597538.

Carracedo, P., Debón, A., Iftimi, A., *et al.* (2018) Detecting spatio-temporal mortality clusters of European countries by sex and age. *International Journal for Equity in Health*, **17**(1), 38, https://doi.org/10.1186/s12939-018-0750-z.

Cheysson, F. (2016) *starma: Modelling Space Time AutoRegressive Moving Average (STARMA) Processes*. R package version 1.3, https://CRAN.R-project.org/package=starma.
Clayton, D. and Schifflers, E. (1987) Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine*, **6**(4), 469–481, https://doi.org/10.1002/sim.4780060406.
Cliff, A.D. and Ord, J.K. (1973) *Spatial Autocorrelation*. London: Pion.
Costantini, M. and Lupi, C. (2013) A simple panel-CADF test for unit roots. *Oxford Bulletin of Economics and Statistics*, **75**(2), 276–296, https://doi.org/10.1111/j.1468-0084.2012.00690.x.
Croissant, Y. and Millo, G. (2008) Panel data econometrics in R: The plm package. *Journal of Statistical Software*, **27**(2), 1–43, https://doi.org/10.18637/jss.v027.i02.
Currie, I.D., Durban, M. and Eilers, P.H. (2004) Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**(4), 279–298, https://doi.org/10.1191/1471082X04st080oa.
De Mori, L., Haberman, S., Millossovich, P. and Zhu, R. (2025) Mortality forecasting via multi-task neural networks. *ASTIN Bulletin*, **55**(2), 313–331, https://doi.org/10.1017/asb.2025.10.
Derksen, S. and Keselman, H.J. (1992) Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, **45**(2), 265–282, https://doi.org/10.1111/j.2044-8317.1992.tb00992.x.
Elhorst, J. (2014) *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer. ISBN 9783642403408. https://link.springer.com/book/10.1007/978-3-642-40340-8.
Enchev, V., Kleinow, T. and Cairns, A.J. (2017) Multi-population mortality models: Fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, **2017**(4), 319–342, https://doi.org/10.1080/03461238.2015.1133450.
Gibbs, Z., Groendyke, C., Hartman, B. and Richardson, R. (2020) Modeling county-level spatio-temporal mortality rates using dynamic linear models. *Risks*, **8**(4), https://doi.org/10.3390/risks8040117.
Gneiting, T. and Raftery, A.E. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**(477), 359–378, https://doi.org/10.1198/016214506000001437.
Greco, F. and Scalone, F. (2013) A space-time extension of the Lee–Carter model in a hierarchical Bayesian framework: Modelling and forecasting provincial mortality in Italy, https://api.semanticscholar.org/CorpusID:55707767.
Griffith, D.A. (2010) Modeling spatio-temporal relationships: Retrospect and prospect. *Journal of Geographical Systems*, **12**(2), 111–123, https://doi.org/10.1007/s10109-010-0120-x.
Griffith, D.A. (2012) Space, time, and space-time eigenvector filter specifications that account for autocorrelation. *Estadística Española*, **54**, 4–27 https://www.researchgate.net/publication/288942431_Space_time_and_space-time_eigenvector_filter_specifications_that_account_for_autocorrelation.
Griffith, D.A. and Paelinck, J.H.P. (2018) Space–time autocorrelation. In *Morphisms for Quantitative Spatial Analysis*, pp. 25–34. Cham, Switzerland: Springer.
Guibert, Q., Loisel, S., Lopez, O. and Piette, P. (2020) Bridging the Li–Carter's gap: A locally coherent mortality forecast approach. HAL Working Papers, *hal-02472777*, https://hal.science/hal-02472777/document.
Guibert, Q., Lopez, O. and Piette, P. (2019) Forecasting mortality rate improvements with a high-dimensional VAR. *Insurance: Mathematics and Economics*, **88**, 255–272, https://doi.org/10.1016/j.insmatheco.2019.07.004.
Hainaut, D. (2018) A neural-network analyzer for mortality forecast. *ASTIN Bulletin*, **48**(2), 481–508, https://doi.org/10.1017/asb.2017.33.
Hassan, A. (2021) *Spatial data analysis: Applications to population health*. Ph.D. Thesis, Université de Lille https://www.researchgate.net/publication/361052863_Spatial_data_analysis_applications_to_population_health.
He, L., Huang, F., Shi, J. and Yang, Y. (2021) Mortality forecasting using factor models: Time-varying or time-invariant factor loadings? *Insurance: Mathematics and Economics*, **98**, 14–34, https://doi.org/10.1016/j.insmatheco.2021.01.006.
HMD (2025) Human Mortality *Database*. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). http://www.mortality.org (Accessed 21 June 2025).
Hofstede, G. (1984) Cultural dimensions in management and planning. *Asia Pacific Journal of Management*, **1**(2), 81–99, https://doi.org/10.1007/BF01733682.
Hunt, A. and Blake, D. (2014) A general procedure for constructing mortality models. *North American Actuarial Journal*, **18**(1), 116–138, https://doi.org/10.1080/10920277.2013.852963.
Hurlin, C. and Mignon, V. (2007) *Second generation panel unit root tests*. HAL Working Papers halshs-00159842, https://ideas.repec.org/p/hal/wpaper/halshs-00159842.html.
Hyndman, R.J. and Khandakar, Y. (2008) Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, **27**(3), 1–22, https://doi.org/10.18637/jss.v027.i03.
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, https://doi.org/10.1016/j.csda.2006.07.028.
Islam, M.D., Li, B., Lee, C. and Wang, X. (2022) Incorporating spatial information in machine learning: The Moran eigenvector spatial filter approach. *Transactions in GIS*, **26**, 1–21, https://doi.org/10.1111/tgis.12894.
Jarner, S.F. and Jallbjørn, S. (2020) Pitfalls and merits of cointegration-based mortality models. *Insurance: Mathematics and Economics*, **90**, 80–93, https://doi.org/10.1016/j.insmatheco.2019.10.005.
Kleiber, C. and Lupi, C. (2011) Panel Unit Root Testing with R, https://api.semanticscholar.org/CorpusID:197629737.
Krúmiňš, J. and Zvidriňš, P. (1992) Recent mortality trends in the three baltic republics. *Population Studies*, **46**(2), 259–273, https://doi.org/10.1080/0032472031000146226.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press

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, https://doi.org/10.1080/02664763.2022.2104228.
Lee, R.D. and Carter, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671, https://doi.org/10.2307/2290201.
Lee, R.D. and Miller, T. (2001) Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, **38**(4), 537–549, https://doi.org/10.1353/dem.2001.0036.
Li, H. and Shi, Y. (2021) Forecasting mortality with international linkages: A global vector-autoregression approach. *Insurance: Mathematics and Economics*, **100**, 59–75, https://doi.org/10.1016/j.insmatheco.2021.04.006.
Li, H., Tan, K. S., Tuljapurkar, S. and Zhu, W. (2021) Gompertz law revisited: Forecasting mortality with a multi-factor exponential model. *Insurance: Mathematics and Economics*, **99**, 268–281, https://doi.org/10.1016/j.insmatheco.2021.03.018.
Li, N. and Lee, R.D. (2005) Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, **42**(3), 575–594 http://www.jstor.org/stable/4147363.
Li, N., Lee, R.D. and Gerland, P. (2013) Extending the Lee–Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, **50**(6), 2037–2051, https://doi.org/10.1007/s13524-013-0232-2.
Liu, Z. (2021) *Bayesian Poisson log-normal model with regularized time structure and spatial framework for mortality projection of multi-population*. Ph.D. Dissertation, Clemson University, https://tigerprints.clemson.edu/all_dissertations/2859.
Moran, P.A.P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, **37**(1–2), 17–23, https://doi.org/10.2307/2332142.
Njenga, C.N. and Sherris, M. (2020) Modeling mortality with a Bayesian vector autoregression. *Insurance: Mathematics and Economics*, **94**, 40–57, https://doi.org/10.1016/j.insmatheco.2020.05.011.
Ocaña-Riola, R. and Mayoral-Cortés, J.M. (2010) Spatio-temporal trends of mortality in small areas of Southern Spain. *BMC Public Health*, **10**, 26, https://doi.org/10.1186/1471-2458-10-26.
Pascariu, M.D., Canudas-Romo, V. and Vaupel, J.W. (2018) Advances in mortality forecasting: Introduction. *Genus*, **74**(1), 1–15, https://doi.org/10.1186/s41118-018-0045-7.
Perla, F., Richman, R., Scognamiglio, S. and Wüthrich, M.V. (2021) Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, **2021**(7), 572–598, https://doi.org/10.1080/03461238.2020.1867232.
Perla, F., Richman, R., Scognamiglio, S. and Wüthrich, M.V. (2024) Accurate and explainable mortality forecasting with the LocalGLMnet. *Scandinavian Actuarial Journal*, **2024**(7), 739–761, https://doi.org/10.1080/03461238.2024.2307620.
Pfeifer, P.E. and Deutsch, S.J. (1980) A three-stage iterative procedure for space-time modeling. *Technometrics*, **22**(1), 35–47 http://www.jstor.org/stable/1268381.
Pickle, L.W. (2000) Exploring spatio-temporal patterns of mortality using mixed effects models. *Statistics in Medicine*, **19**(17-18), 2251–2263.
R Core Team (2025) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
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, https://doi.org/10.1016/j.insmatheco.2005.12.001.
Richman, R. and Wüthrich, M.V. (2021) A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, **15**(2), 346–366, https://doi.org/10.1017/S1748499519000071.
Robben, J. (2021) *MultiMoMo: Multi-population mortality models*. R package version 1.0.0 https://github.com/jensrobben/MultiMoMo.
Salima, B.A. (2018) Spatial econometrics on panel data, https://api.semanticscholar.org/CorpusID:214675854.
Simonovska, R. (2025) SDPDmod: Spatial Dynamic Panel Data Modeling, https://CRAN.R-project.org/package=SDPDmod.
Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46**(Supplement), 234–240, https://doi.org/10.2307/143141.
Tuljapurkar, S., Li, N. and Boe, C. (2000) A universal pattern of mortality decline in the G7 countries. *Nature*, **405**(6788), 789–792, https://doi.org/10.1038/35015561.
Ugarte, M.D., Goicoa, T. and Militino, A.F. (2010) Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, **21**(3-4), 270–289, https://doi.org/10.1002/env.1011.
Van Belle, G. (2008) *Statistical Rules of Thumb*, 2nd edn. Hoboken, NJ: Wiley.
Vazzoler, S. (2021) *sparsevar: Sparse VAR/VECM Models Estimation*. R package version 0.1.0, https://CRAN.R-project.org/package=sparsevar.
Villegas, A.M., Kaishev, V.K. and Millossovich, P. (2015) StMoMo: An R package for stochastic mortality modeling. *European Actuarial Journal*, **5**(2), 271–284, https://doi.org/10.18637/jss.v084.i03.
Vékás, P. (2019) Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research*, **28**(3), 1031–1048, https://doi.org/10.1007/s10100-019-00617-0.
Yao, S. and Wang, X. (2023) Machine learning in long-term mortality forecasting. *Annals of Actuarial Science*, **17**(2), 143–168, https://doi.org/10.1057/s41288-024-00320-5.
Yu, J., de Jong, R. and Lee, L. (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, **146**(1), 118–134, https://doi.org/10.1016/j.jeconom.2008.08.002.
Yu, J. and Lee, L. (2010) Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, **26**(5), 1332–1362 http://www.jstor.org/stable/40800885.

https://doi.org/10.1017/asb.2026.10092 Published online by Cambridge University Press