# Modeling mortality with Kernel Principal Component Analysis (KPCA) method

Yuanqi Wu[1], Andrew Chen[2], Yanbin Xu[3], Guangming Pan[1] and Wenjun Zhu[3] (iD)

[1]School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore; [2]Department of Electrical and Electronic Engineering, Imperial College London, London, UK; and [3]Division of Banking & Finance, Nanyang Business School, Nanyang Technological University, Singapore
**Corresponding author:** Wenjun Zhu; Email: wjzhu@ntu.edu.sg

**Abstract**

As the global population continues to age, effective management of longevity risk becomes increasingly critical for various stakeholders. Accurate mortality forecasting serves as a cornerstone for addressing this challenge. This study proposes to leverage Kernel Principal Component Analysis (KPCA) to enhance mortality rate predictions. By extending the traditional Lee-Carter model with KPCA, we capture nonlinear patterns and complex relationships in mortality data. The newly proposed KPCA Lee-Carter algorithm is empirically tested and demonstrates superior forecasting performance. Furthermore, the model's robustness was tested during the COVID-19 pandemic, showing that the KPCA Lee-Carter algorithm effectively captures increased uncertainty during extreme events while maintaining narrower prediction intervals. This makes it a valuable tool for mortality forecasting and risk management. Our findings contribute to the growing body of literature where actuarial science intersects with statistical learning, offering practical solutions to the challenges posed by an aging world population.

## 1. Introduction

The aging of populations around the globe has elevated the importance of longevity risk management (World Bank, 2024). This challenge underscores the need for accurate and reliable mortality forecasting, a fundamental prerequisite for various stakeholders such as governments, pension funds, insurance and reinsurance companies, and individuals (Cairns *et al.*, 2009; Gaille & Sherris, 2011; Blake *et al.*, 2013). As depicted in Figure 1, which illustrates global and regional life expectancy at birth, it is evident that life expectancy has been steadily rising across all regions. Notably, developed regions such as North America, the European Union, and Australia exhibit the highest life expectancy levels. This remarkable global phenomenon has triggered a surge of interest in machine learning techniques within the field of actuarial science, particularly concerning the intricate domain of mortality prediction (LeCun *et al.*, 2015; Richman, 2021a, 2021b).

The extended life expectancy observed in more developed regions and income classes can be attributed to a confluence of environmental, healthcare, and socioeconomic factors, most notably characterized by economic progress and improved healthcare access, resulting in a significant decline in mortality rates among older age groups (Rau *et al.*, 2008; Maier *et al.*, 2010; Desjardins & Bourbeau, 2010; Woolf & Schoomaker, 2019). High-income regions have experienced a substantial increase in the old-age dependency ratio, defined as the proportion of
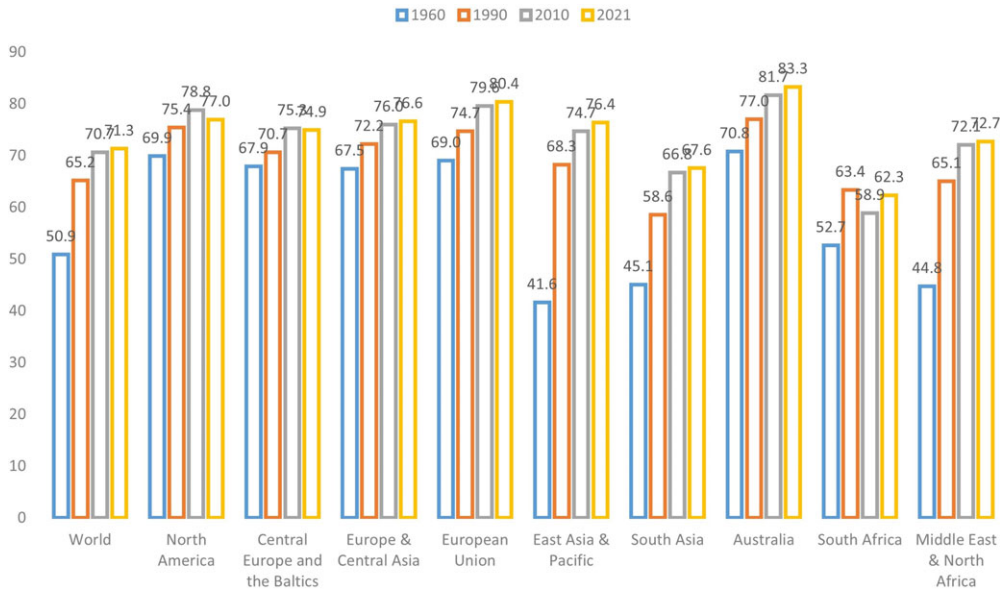
**Figure 1.** Life expectancy at birth (total) of the world and different regions (Source: World Bank).

the population aged above 65 in comparison to the working-age population, surpassing the global average. Take Hong Kong as an example: approximately 19.1% of the population in Hong Kong is aged over 65, with nearly 45% aged over 50. As a result, forecasting mortality has become a pivotal task in demographic analysis, exemplifying the challenges and financial obligations faced by pension plans and related institutions along with the ongoing global increase in life expectancy.

Nevertheless, accurately predicting mortality rates presents a formidable challenge. The complexity arises from intricate nonlinear effects inherent in mortality rates, exacerbated by external factors like pandemics (e.g., COVID-19) and climate change. For instance, for the period of January 2020 to December 2021, the overall excess deaths are estimated to be 14.83 million, 2.74 times higher than the COVID-19 deaths reported to the WHO (Msemburi *et al.*, 2023), underscoring the complexity of mortality dynamics.

However, traditional mortality models, including the seminal work by Lee and Carter (1992), primarily rely on linear extrapolation methods, often overlooking these nonlinear effects. In recent years, there is an evolving landscape of machine learning applications in different fields of actuarial science (e.g., (Wüthrich, 2017; Gabrielli & V. Wüthrich, 2018; Lally & Hartman, 2018; Lee & Lin, 2018; Gao *et al.*, 2019; Ghahari *et al.*, 2019; Brock Porth *et al.*, 2020; Noll *et al.*, 2020; Devriendt *et al.*, 2021; Henckaerts *et al.*, 2021; Lee, 2021; Gomes *et al.*, 2021; Wuthrich & Buser, 2021; Gao *et al.*, 2022; Hu *et al.*, 2022; Meng *et al.*, 2022; Xin & Huang 2024; Fissler *et al.*, 2023; Chen *et al.*, 2023; Debener *et al.*, 2023)), particularly in the context of mortality modeling and forecasting. Pioneering research by Deprez *et al.* (2017) demonstrated the application of machine learning techniques for the analysis and evaluation of stochastic mortality models, shedding light on the potential of machine learning to enhance our understanding of mortality models. Subsequent studies expanded on this foundation by incorporating tree-based models to improve prediction accuracy (Levantesi and Nigri, 2020; Bjerre, 2022). Hainaut (2018) introduced an innovative framework that used auto-encoders to capture nonlinearities in mortality data, extending the traditional Lee–Carter model. Richman and Wüthrich (2019) proposed a neural network extension of the Lee–Carter model for multiple populations, allowing neural networks to automatically select optimal structures for improved mortality forecasting. Since then, numerous studies have delved into improving mortality prediction performance through various

neural network architectures (Nigri *et al.*, 2019; Petneházi & Gáll, 2019; Richman & Wuthrich, 2019; Bravo, 2021a, 2021b; Nigri *et al.*, 2021; Perla *et al.*, 2021; Wang *et al.*, 2021; Chen & Khaliq, 2022; Lindholm & Palmborg, 2022; Zhang *et al.*, 2022; Marino *et al.*, 2023). Some studies employed tensor-based approaches for mortality modeling (Dong *et al.*, 2020; Cardillo *et al.*, 2022). Most recently, transformer models have emerged as a powerful tool for mortality prediction in major countries (Wang *et al.*, 2024). Furthermore, researchers have explored the uncertainty associated with machine learning models for mortality forecasting (Schnürch & Korn, 2022; Marino *et al.*, 2023).

This paper contributes to the dynamic field of mortality modeling and forecasting by introducing a novel framework that leverages machine learning techniques, specifically Kernel Principal Component Analysis (KPCA), to enhance the accuracy of mortality predictions. This innovation holds the potential to address the growing challenges posed by increasing life expectancy, volatile mortality risks, and population dependencies. By surpassing the limitations of traditional linear extrapolation models, our research empowers stakeholders with a more accurate tool for decision-making in the face of evolving mortality dynamics.

The rest of the paper proceeds as follows. Section 2 introduces the KPCA modeling methodology. Section 3 presents the empirical test results. Section 4 tests the robustness of the proposed method. Section 5 concludes the paper.

## 2. Modeling methodology

### 2.1 The Lee–Carter model

In their seminal paper, Lee and Carter (1992) proposed the following model that has become the standard model for the mortality forecast literature and the preferred methodology for the U.S. Census Bureau. Let $m_{xt}$ be the central death rate at age $x$ and year $t$, $x_0 \leq x \leq X$, $0 \leq t \leq T$, then

$$\log(m_{xt}) = \alpha_x + \beta_x \kappa_t + \epsilon_{xt}, \tag{1}$$

where $\alpha_x$ is a static age function specifying the general shape of mortality by age; $\beta_x \kappa_t$ captures the age-period effect, with $\kappa_t$ reflecting overall mortality trend (period-related effect) and $\beta_x$ modulating its effect across ages (age-related effect). In particular, $\kappa_t$ is commonly known as the mortality index, contributing to capturing the overall level of mortality improvement.

The Lee–Carter model is only identifiable up to a transformation. As a result, in the literature, it is conventional to impose the following parameter constraints to circumvent the identification problem:

$$\begin{cases} \sum_t \kappa_t = 0, \\ \sum_x \beta_x = 1. \end{cases} \tag{2}$$

### 2.2 Estimation of mortality trend

In this section, we introduce two well-established methods to estimate the Lee–Carter coefficients to build the context: the singular value decomposition method (SVD) and the maximum-likelihood estimation method (MLE). They are closely related to the KPCA Lee–Carter model proposed in this paper.

#### 2.2.1 Singular value decomposition

The SVD method is a conventional approach to estimate the parameters in Model (1), as proposed by Lee and Carter (1992). This method first forms the centered logarithms of the mortality rates matrix, denoted as $\mathbf{A}_{X \times T}$, where each element of the matrix is the age-specific intercept $\hat{\alpha_x}$

adjusted logarithm of the mortality rates. $\hat{\alpha}_x$ is estimated as the long-term mean of the logged mortality for each age. Specifically:

$$\hat{\alpha}_x = \frac{1}{T} \sum_{t=1}^{T} \ln(m_{xt}). \tag{3}$$

The logic behind this is that

$$\sum_{t=1}^{T} \ln(m_{xt}) = \sum_{t=1}^{T} (\alpha_x + \beta_x \kappa_t + \epsilon_{xt})$$

$$\approx T\alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t = T\alpha_x, \tag{4}$$

where the identification constraint $\sum_{t=1}^{T} \kappa_t = 0$ is used (see Constrants (2))

The centered logarithms of the mortality rates matrix $\mathbf{A}_{X \times T}$ are formed as follows:

$$\mathbf{A}_{X \times T} = \begin{pmatrix} \ln(m_{11}) - \hat{\alpha}_1 & \ln(m_{12}) - \hat{\alpha}_1 & \cdots & \ln(m_{1T}) - \hat{\alpha}_1 \\ \ln(m_{21}) - \hat{\alpha}_2 & \ln(m_{22}) - \hat{\alpha}_2 & \cdots & \ln(m_{2T}) - \hat{\alpha}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \ln(m_{X1}) - \hat{\alpha}_X & \ln(m_{X2}) - \hat{\alpha}_X & \cdots & \ln(m_{XT}) - \hat{\alpha}_X \end{pmatrix} \tag{5}$$

Applying SVD method to matrix $\mathbf{A}_{X \times T}$, the factorization form is $\mathbf{A}_{X \times T} = U\Sigma V^T$, where the columns of $U$ and $V$ are orthonormal and the matrix $\Sigma$ is diagonal with positive real entries. Thus, $\mathbf{A}_{X \times T}$ is decomposed with right singular vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_r$, left singular vectors $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_r$, and corresponding singular values $\lambda_1, \lambda_2, \cdots, \lambda_r$. Then $\mathbf{A}_{X \times T}$ can be written as

$$\mathbf{A}_{X \times T} = \lambda_1 \vec{u}_1 \vec{v}_1^T + \lambda_2 \vec{u}_2 \vec{v}_2^T + \cdots + \lambda_r \vec{u}_r \vec{v}_r^T \tag{6}$$

The first left and right singular vectors, along with leading value of the SVD, provide the approximation of $\beta_x$ and $\kappa_t$, respectively. Specifically,

$$\hat{\beta}_x = \frac{u_{1x}}{\sum_{i=1}^{X} u_{1i}}, \tag{7}$$

$$\hat{\kappa}_t = \lambda_1 \left( v_{1t} - \frac{1}{T} \sum_{t=1}^{T} v_{1t} \right). \tag{8}$$

Notably, weighting and adjustments are applied to $\hat{\beta}_x$ and $\hat{\kappa}_t$ to satisfy the conventional Constraints (2).

### 2.2.2 Maximum-likelihood estimation

Besides the SVD method, using MLE to estimate the Lee–Carter coefficients is another classic approach (Wilmoth, 1993). The MLE method requires specifying a probabilistic model. Following Wilmoth, we assume the number of deaths, denoted as $D_{xt}$, at age $x$ and year $t$ follows a Poisson distribution, with the mean parameter $\lambda_{xt}$ set to the product of the number of lives at risk ($E_{xt}$) and the mortality rate ($m_{xt}$) for the corresponding age and year. That is:

$$D_{xt} \sim Poisson(\lambda_{xt})$$
$$\lambda_{xt} = E_{xt} m_{xt} \tag{9}$$

Assuming $D_{xt}$ is independent across time-age, we sum over all time-age to obtain the full log-likelihood:

$$
\begin{aligned}
l(\theta) &= \sum_{xt} D_{xt} log(\lambda_{xt})) - \lambda_{xt} - log(D_{xt}) \\
&= \sum_{xt} D_{xt} log(E_{xt}) + D_{xt} log(m_{xt}) - E_{xt} m_{xt} - log(D_{xt}) \\
&= \sum_{xt} \left\{ D_{xt} log(m_{xt}) - E_{xt} m_{xt} \right\} + C \\
&= \sum_{xt} \left\{ D_{xt}(\alpha_x + \beta_x \kappa_t) - E_{xt} e^{\alpha_x + \beta_x \kappa_t} \right\} + C.
\end{aligned} \tag{10}
$$

Here, we substitute the Lee–Carter model (Equation 1) into the log-likelihood function to obtain the optimization maximizer. The term $C$ is a constant that aggregates the non-variable terms. By finding the parameter set $\theta = (\alpha_x, \beta_x, \kappa_t)$ that maximizes the log-likelihood function $l(\theta)$, we estimate the Lee–Carter coefficients.

A notable advantage of MLE method is that it is flexible in including additional covariates or explanatory variables to account for other factors affecting mortality rates, making it a more adaptable method for complex models, while SVD method is limited to the structure of the decomposition and does not easily accommodate additional covariates.

## 2.3 KPCA Lee–Carter model

KPCA extends traditional Principal Component Analysis (PCA) into nonlinear feature spaces, offering a means to better capture the intricate spatial structures and nonlinear patterns present in high-dimensional mortality data. While traditional PCA is effective for capturing linear relationships and patterns in data, it may underperform when data exhibits nonlinear structures. KPCA addresses this limitation by using a kernel trick to project the data into a higher-dimensional space where the linear relationships are more easily captured. By applying PCA in this transformed space, KPCA effectively captures nonlinear patterns present in the original data.

As discussed previously, nonlinearity may exist in the mortality data. KPCA excels at capturing these nonlinear structures and enables the estimation of the time-varying mortality index $\kappa_t$ through the use of kernel function. The estimation stage of the Lee–Carter model using SVD can be viewed as a specific instance of PCA. It involves summarizing the log-mortality data by only considering the first principal component (PC), which is $\kappa_t$ in Equation (1), while the variance of other PCs is encapsulated within the error term $\epsilon_{xt}$.

In this section, we illustrate how the conventional SVD based solution of Lee–Carter models can be extended to the KPCA context. Section 2.3.1 briefly introduces the kernel transformation and the kernel functions used in this paper. Section 2.3.2 illustrates the estimation of Lee–Carter coefficients with KPCA method.

### 2.3.1 Kernel functions

KPCA is a statistical learning algorithm that allow us to study the data in the feature space $\mathcal{F}$ using kernel function $k$.

**Theorem 1.** (Aronszajn, 1950) *Let $k$ be a kernel function in some space $\mathcal{X}$, where $\vec{x}_i, \vec{x}_j \in \mathcal{X}$. Then, there exists a Hilbert space of functions $\mathcal{H}$ and a mapping $\phi: \mathcal{X} \to \mathcal{H}$ such that $k(\vec{x}_i, \vec{x}_j) = \langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle_{\mathcal{H}}$.*

Here, the Hilbert space $\mathcal{H}$ is what we referred as feature space $\mathcal{F}$. Theorem 1 says that $k(\vec{x}_i, \vec{x}_j)$ can be obtained without explicitly calculating $\phi(\vec{x})$. This implicit mapping helps us construct kernel matrix without knowing what $\phi$ is. The kernel functions that we tested in this paper include:

- Gaussian Radial Basis Function (RBF) Kernel:

$$k\left(\vec{x}_i, \vec{x}_j\right) = \exp\left(-\sigma \left\|\vec{x}_i - \vec{x}_j\right\|^2\right), \sigma > 0, \sigma \in \mathbb{R} \tag{11}$$

- Laplace Kernel:

$$k\left(\vec{x}_i, \vec{x}_j\right) = \exp\left(-\gamma \left\|\vec{x}_i - \vec{x}_j\right\|\right), \gamma > 0, \gamma \in \mathbb{R} \tag{12}$$

- Polynomial Kernel:

$$k\left(\vec{x}_i, \vec{x}_j\right) = \left(\vec{x}_i^T \vec{x}_j + c\right)^q, c \geq 0, q \in \mathbb{N}^+ \tag{13}$$

All three kernels are effective for handling nonlinear data, each excelling in different scenarios. The Gaussian RBF kernel is particularly well-suited for capturing smooth, continuous variations, whereas the Laplace kernel demonstrates greater robustness to noise and outliers, making it ideal for data characterized by sharp transitions (Wang *et al.*, 2015). The polynomial kernel, on the other hand, is strong at capturing feature interactions and modeling global structures, making it effective for data that can be made linearly separable through polynomial (Weiße *et al.*, 2006).

In untabulated analyses, we found that the Gaussian RBF kernel performed best, as measured by the minimum error in predicting mortality data in validation set, which will be illustrated in detail in Section 3. This superior performance is likely due to the kernel's ability to handle the nonlinear relationships and local variations present in the mortality data, aligning with the theoretical strengths of the Gaussian RBF kernel. Therefore, in our empirical analyses, we use the Gaussian RBF kernel function as the default setting to compare the performance of KPCA relative to other well-established mortality models.

### 2.3.2 Kernel PCA

The procedure of using KPCA to estimate the Lee–Carter coefficients from log-mortality rate is as below:

**Step 1: Estimate $\hat{\alpha}_x$ and form the centered mortality vectors.**

Given that KPCA will transform the log-mortality data to a higher dimension, we first need to estimate the age-specific intercept term $\hat{\alpha}_x$ for each age $x$, following the method described in Equation (3). After estimating $\hat{\alpha}_x$, we subtract it from the log-mortality rate to obtain the centered log-mortality rate.

For the sake of discussion, we denote the centered log-mortality vector for time $t$ as $\vec{A}_t = (A_{1t}, \ldots, A_{Xt})^T$, where the element

$$A_{xt} = \log\left(m_{xt}\right) - \hat{\alpha}_x. \tag{14}$$

The vector $\vec{A}_t$ corresponds to the $t$-th column of the matrix $\mathbf{A}_{X \times T}$ in Equation (5). We will then transform the set $\{\vec{A}_t\}_{t=1}^T$ from its original space to the extended space. We note that this approach is transforming the log-mortality data over time for each age group rather than the other way around since we intend to capture the time-varying patterns of the mortality nonlinearity.

**Step 2: Construct centered kernel matrix in featured space.**

The data points should be centered in the feature space, that is, $\sum_{t=1}^T \phi(\vec{A}_t) = 0$. We denote the centered mapping of $x_i$ in the feature space as $\phi^c(\vec{A}_i)$ and calculated as:

$$\phi^c(\vec{A}_i) = \phi(\vec{A}_i) - \frac{1}{T}\sum_{j=1}^T \phi(\vec{A}_j). \tag{15}$$

The expression of the $i, j$-th element of centered kernel matrix $\mathbf{K}^c$ is calculated as:

$$
\begin{aligned}
K_{ij}^c &= \langle \phi^c(\vec{A}_i), \phi^c(\vec{A}_j) \rangle_{\mathcal{H}} \\
&= \left\langle \phi(\vec{A}_i) - \frac{1}{T} \sum_{l=1}^{T} \phi(\vec{A}_l), \phi(\vec{A}_j) - \frac{1}{T} \sum_{l=1}^{T} \phi(\vec{A}_l) \right\rangle_{\mathcal{H}} \\
&= \langle \phi(\vec{A}_i), \phi(\vec{A}_j) \rangle_{\mathcal{H}} - \frac{1}{T} \sum_{l=1}^{T} \langle \phi(\vec{A}_i), \phi(\vec{A}_l) \rangle_{\mathcal{H}} \\
&\quad - \frac{1}{T} \sum_{l=1}^{T} \langle \phi(\vec{A}_j), \phi(\vec{A}_l) \rangle_{\mathcal{H}} + \frac{1}{T^2} \sum_{l=1}^{T} \sum_{l'=1}^{T} \langle \phi(\vec{A}_l), \phi(\vec{A}_{l'}) \rangle_{\mathcal{H}} \\
&= K_{ij} - \frac{1}{T} \sum_{l=1}^{T} K_{il} - \frac{1}{T} \sum_{l=1}^{T} K_{jl} + \frac{1}{T^2} \sum_{l=1}^{T} \sum_{l'=1}^{T} K_{ll'}.
\end{aligned}
\tag{16}
$$

Here, $K_{ij}$ represents the $i, j$-th element of the uncentered kernel matrix $\mathbf{K}$. Let $\mathbf{1}_T$ denote a $T \times T$ matrix with all elements equal to $\frac{1}{T}$, the relationship between $\mathbf{K}^c$ and $\mathbf{K}$ is:

$$
\mathbf{K}^c = \mathbf{K} - \mathbf{1}_T \mathbf{K} - \mathbf{K} \mathbf{1}_T + \mathbf{1}_T \mathbf{K} \mathbf{1}_T.
\tag{17}
$$

### Step 3: Perform PCA to the centered kernel matrix.

In this step, we perform PCA on $\mathbf{K}^c$. Since PCA is a well-established method, we will not delve into the details of the procedure (see comprehensive procedures in (Jolliffe, 2002) and (Shlens, 2014)). Given that $\mathbf{K}^c$ is real and symmetric, it can be eigendecomposed as follows:

$$
\mathbf{K}^c = \mathbf{Q} \Lambda \mathbf{Q}^T,
\tag{18}
$$

where $\mathbf{Q}$ is the matrix of $\mathbf{K}^c$'s eigenvectors (or PCs), and $\Lambda$ is a diagonal matrix of the corresponding eigenvalues. In the eigendecomposition, the eigenvalues are ranked from the largest to the smallest, with the leading PCs capturing more variance of $\mathbf{K}^c$.

### Step 4: Estimate $\hat{\beta}_x$ and $\hat{\kappa}_t$.

The KPCA method of obtaining the Lee–Carter coefficients is similar in form to the SVD method (see Section 2.2.1). However, due to the properties of the matrix $\mathbf{K}^c$, the left and right singular matrices are identical. Denote the $t$-th column of $\mathbf{Q}$ as $\vec{v}_t$, where $\vec{v}_t = (v_{1t}, \ldots, v_{Tt})^T$

To estimate the age-specific sensitivity $\hat{\beta}_x$ for age $x$, we project the centered log-mortality for age $x$ onto the first PC (i.e., $\vec{v}_1$):

$$
\hat{\beta}_x^{(1)} = \frac{\sum_{t=1}^{T} A_{xt} v_{t1}}{\sum_{t=1}^{T} v_{t1}},
\tag{19}
$$

where $A_{xt}$ is calculated as described in Equation (14), and the weighting $\sum_{t=1}^{T} v_{t1}$ is applied to meet the conventional Constraints (2). The estimated $\hat{\beta}_x^{(1)}$ is labeled with a superscript to indicate it is derived from the first PC. The Lee–Carter model can incorporate more than one pair of $\beta_x$ and $\kappa_t$ estimated from different principal components (Booth *et al.*, 2002).

The calculation of $\hat{\beta}_x$ by projecting the centered log-mortality data onto the eigenvectors allows us to quantify how each age group's mortality rates respond to the primary time-varying factors captured by KPCA. This makes $\hat{\beta}_x$ a valid representation of age-specific sensitivity, aligning with the objectives of the Lee–Carter model in capturing the dynamics of mortality rates over time.

The estimated time-varying mortality trend $\hat{\kappa}_t$ under the KPCA framework is the first principal component adjusted by the long-term mean and then weighted by the corresponding eigenvalue.

---

**Algorithm 1:** KPCA Lee–Carter Algorithm for Mortality Forecasting

---

**Input** Mortality Data $m_{xt}$, Death distribution $D_{xt}$, $t = 1, \ldots, T$; $x = 1, \ldots, X$
**Output** $h$-steps ahead forecasts of $\hat{m}_{x(T+h)}$, $h = 1, 2, \ldots$
    **Dimension Reduction Step:**
1:  Estimate the age-specific intercept term $\hat{\alpha}_x$ by Equation (3).
2:  Select kernel and corresponding parameters.
3:  Compute kernel matrix $\mathbf{K} = kernelMatrix(kernel, parameter, Y)$.
4:  Compute the centered log-mortality $A_{xt}$ for each $x$ and $t$ by Equation (14) and construct the centered log-mortality matrix as depicted in (5).
5:  Compute the mortality kernel matrix in the feature space $\mathbf{K}^c$ following Equation (17).
6:  Conduct eigendecomposition on $\mathbf{K}^c$, and get the first PC $v_{t1}$ and the corresponding eigenvalue $\lambda_1$.
7:  Compute the first set of age-specific sensitivity $\hat{\beta}_x^{(1)}$ for $x = 1, \ldots, X$ following Equation (19) and time-varying mortality trend $\hat{\kappa}_t^{(1)}$ for $t = 1, \ldots, T$ following Equation (20).
8:  **if** include two PCs to the model
9:     Get the second PC $v_{t2}$ and the corresponding eigenvalue $\hat{\lambda}_2$.
10:    Compute the second set of age-specific sensitivity $\hat{\beta}_x^{(2)}$ for $x = 1, \ldots, X$ and time-varying mortality trend $\hat{\kappa}_t^{(2)}$ for $t = 1, \ldots, T$.
11:  **end if**
    **Forecasting Step:**
12:  Fit $\hat{\kappa}_t^{(1)}$ with ARIMA(0,1,0) time series model.
13:  Compute $\hat{\kappa}_{t+h}^{(1)}$, for $h = 1, 2, \ldots$, the h-step ahead forecasts.
14:  **if** include one PC to the model
15:    Compute h-step ahead forecasts $\hat{m}_{x(T+h)} = \hat{\alpha}_x + \hat{\beta}_x^{(1)} \hat{\kappa}_{t+h}^{(1)}$
16:  **else if** include two PCs to the model
17:    Fit $\hat{\kappa}_t^{(2)}$ with ARIMA(0,1,0) time series model.
18:    Compute $\hat{\kappa}_{t+h}^{(2)}$, $h = 1, 2, \ldots$, the h-step ahead forecasts.
19:    Compute h-step ahead forecasts $\hat{m}_{x(T+h)} = \hat{\alpha}_x + \hat{\beta}_x^{(1)} \hat{\kappa}_{t+h}^{(1)} + \hat{\beta}_x^{(2)} \hat{\kappa}_{t+h}^{(2)}$
20:  **end if**

---

Since the transformation was done over time, the time variation is preserved, and no additional projection is required:

$$\hat{\kappa}_t^{(1)} = \lambda_1 \left( v_{t1} - \frac{1}{T} \sum_{t=1}^{T} v_{t1} \right). \tag{20}$$

For the same reason as $\hat{\beta}_x^{(1)}$, we label the estimated $\hat{\kappa}_t^{(1)}$ with a superscript to indicate it is derived from the first principal component.

    Thus, the KPCA is constructed using the obtained eigenvectors and eigenvalues. We can then select the kernel PCs that explain the most variance in the data. For future mortality projections, we fit $\hat{\kappa}_t^{(1)}$ with an ARIMA(0,1,0) model to project future trends and calculate the future mortality using the Lee–Carter Model (1). The procedure of the KPCA Lee–Carter model is summarized in Algorithm 1.
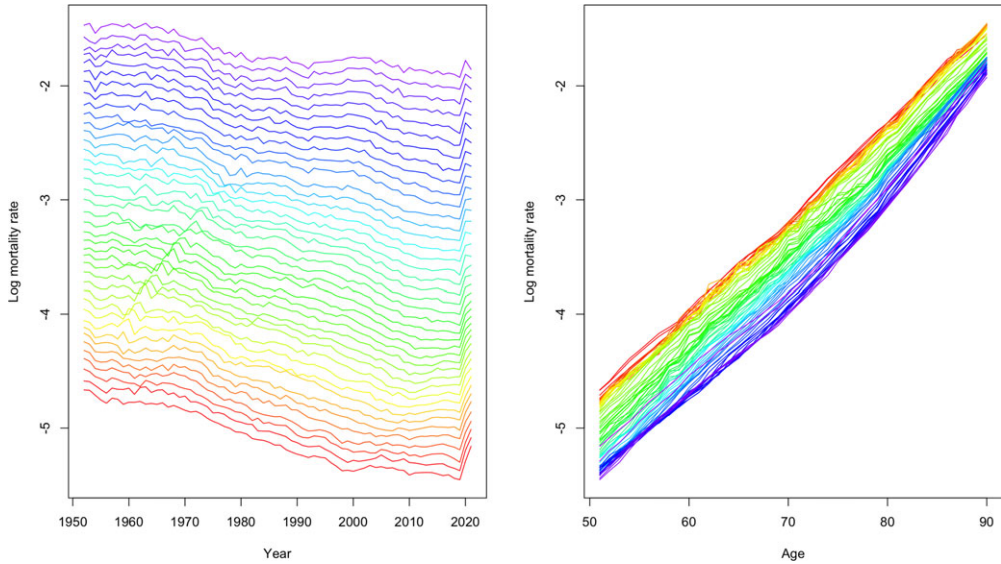
**Figure 2.** Log-mortality rate by year and age, with different colors representing age groups. The oldest age group is shown in violet, while the youngest is highlighted in red.

## 3. Empirical analysis

### 3.1 Data description

In this section, the proposed KPCA Lee–Carter method is applied to the U.S. mortality data from the Human Mortality Database (HMD),[1] a database that provides detailed mortality and population data. All the death numbers ($D_{xt}$), exposure-to-risk ($E_{xt}$) and central mortality rates ($m_{xt}$) are derived from the dataset to facilitate the analysis of mortality patterns across age ($x$) and time ($t$). The sample period is from 1952 to 2021 and we use total population mortality data between age 51 and 90.

The logarithmic patterns of mortality rates over time and across age groups are represented in Figure 2. In the year-specific plot on the left, distinct colors are employed to distinguish between various age groups, with the oldest age group depicted in violet and the youngest in red. Conversely, in the age-specific plot on the right, the color scheme denotes different years, with the most recent years represented in violet and the earliest in red.

The year-specific plot illustrates a consistent decline in mortality rates across all ages spanning from 1959 to 2019, then increased due to COVID-19 from 2020 to 2021. Notably, the decline exhibits varying trajectories for different age groups, highlighting the importance of our task to forecast mortality rates as a means of managing longevity risk, while the COVID-19 period poses new challenges for researchers to deal with excess life expectancy calculation. Conversely, the age-specific plot reveals a predictable trend of increasing mortality rates with advancing age, aligning with conventional wisdom regarding age-related mortality patterns.

### 3.2 Prediction results

We split the U.S. mortality data with time period from 1952 to 2021 into three subsets, training-validation-test, preserving the temporary order. The training set contains observations of a total of 50 years from 1952 to 2001, the validation set contains observations of a total of 10 years from 2002 to 2011, and the test set contains observations of a total of 10 years from 2012 to 2021 for model forecast.

[1]HMD can be accessed at http://www.mortality.org

**Table 1.** Validation result for Kernel Principal Component Analysis 1 principal component model

| $\sigma$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| MAPE | 0.05422343 | 0.05410058 | 0.0540364 | 0.05399309 |
| $\sigma$ | 10 | 15 | 20 | 25 |
| MAPE | 0.05385088 | 0.05379326 | 0.05375357 | 0.05370333 |
| $\sigma$ | 30 | 35 | 40 | 45 |
| MAPE | 0.05360421 | 0.05341044 | 0.05325787 | 0.05326042 |
| $\sigma$ | 0.1 | 0.01 | 0.001 | 0.0001 |
| MAPE | 0.05448813 | 0.05452792 | 0.05453205 | 0.05453254 |

**Table 2.** Validation result for Kernel Principal Component Analysis 2 principal component model

| $\sigma$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| MAPE | 0.05422343 | 0.05410059 | 0.0540364 | 0.05399309 |
| $\sigma$ | 5 | 6 | 7 | 8 |
| MAPE | 0.0539594 | 0.0539313 | 0.0539071 | 0.05388595 |
| $\sigma$ | 0.01 | 0.05 | 0.1 | 0.5 |
| MAPE | 0.05755043 | 0.05106988 | 0.05448818 | 0.05434232 |
| $\sigma$ | 0.0001 | 0.0005 | 0.001 | 0.005 |
| MAPE | 0.05453239 | 0.1062232 | 0.1058115 | 0.0595571 |

The proposed KPCA Lee–Carter method is applied to the mortality data. Standard SVD and MLE methods are natural benchmarks to assess the prediction performance of KPCA Lee–Carter method. In addition, we also include auto-encoder method proposed by Hainaut (2018) as benchmarks for performance assessment comparison, as auto-encoder model is able to incorporate nonlinearity as well. To evaluate the forecast performance, we consider the mean absolute percentage error (MAPE) between the forecast mortality rates and actual mortality rates. More specifically,

$$MAPE = \frac{1}{XT_f} \sum_{t=1}^{T_f} \sum_{x=1}^{X} \left| \frac{m_{xt} - \hat{m}_{xt}}{m_{xt}} \right|, \qquad (21)$$

where $T_f$ is the number of forecast years.

The validation set result shown in Tables 1 and 2 is used for choosing the hyper-parameters of KPCA 1 PC and 2 PC model. We choose the parameters that give the lowest validation MAPE. More specifically, for 1 PC Gaussian RBF kernel, we choose $\sigma = 40$, and for 2 PC Gaussian RBF kernel, we choose $\sigma = 0.05$.

The MAPE for the KPCA method with RBF kernel and chosen hyper-parameter, in comparison with MAPE of benchmark models, are shown in Table 3. For the auto-encoder model, we use three neurons in both the input and output layers and two neurons in the intermediate layer. The model is trained with a learning rate of 0.001, over 800 epochs, and the MSE loss function. The auto-encoder 1 PC model extracts a single-dimensional factor, while the auto-encoder 2 PC model extracts two-dimensional factors from the constructed model.

When compared to all other models, it is evident that the 1 PC and 2PCs KPCA approach yields the lowest MAPE, indicating the efficacy of our proposed KPCA Lee–Carter model. Notably, the root mean squared error (RMSE) and MAPE values obtained from the 1 PC and 2PCs KPCA

**Table 3.**　Mean absolute percentage error (MAPE) for out-sample test set on all models

| Model | SVD | MLE | KPCA 1 PC |
|---|---|---|---|
| MAPE | 0.1019752 | 0.09558255 | 0.09079403 |
| Model | KPCA 2PCs | Auto-encoder 1PC | Auto-encoder 2PCs |
| MAPE | 0.0866092 | 0.1159562 | 0.1125127 |

**Table 4.**　Proportion of variance for Kernel Principal Component Analysis 2 principal component model

| PC1 | PC2 |
|---|---|
| 0.9324705 | 0.05252356 |



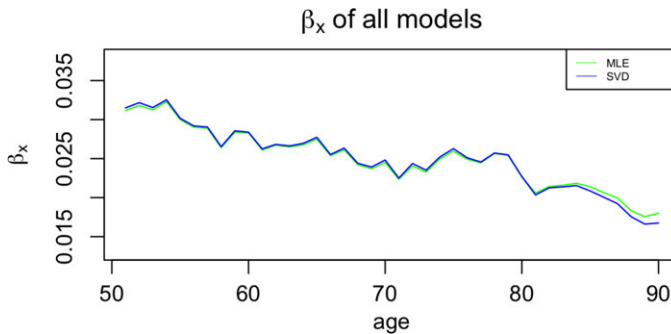**Figure 3.** $\beta_x$ estimated across models.

methods are quite similar. Based on the calculation of proportion of variance for KPCA 2 PC model shown in Table 4, this similarity is attributed to the fact that the first PC can effectively capture a substantial portion of the variance in the data, rendering the inclusion of the second PC less impactful in enhancing our KPCA model's performance.

The fitted value of $\alpha_x$ is the same across all models, as it represents the long-term mean of logged mortality for each age and follows the same formula across different models. The fitted $\beta_x$ based on different models are plotted in Figure 3. Estimation results of $\kappa_t$ from different models are displayed in Figure 4. Figure 5 shows the corresponding fan charts.

Analyzing Figure 3 and Figure 4, we can discern consistent trends in all three parameters when employing the SVD and MLE methods. These methods primarily capture the linear trend present in the mortality data. However, the KPCA method stands out by estimating a distinct mortality index $\kappa_t$ and different age interactions with mortality index $\beta_x$, in contrast to the SVD and MLE methods. This distinction arises from the KPCA's ability to uncover underlying nonlinear patterns within the mortality data through kernel methods in the feature space.

Furthermore, Figure 5 reveals that the width of the fan chart corresponds to the level of uncertainty around the baseline forecast, and Table 5 shows the confidence intervals of out-of-sample predictions for the year 2021. The SVD method produces the narrowest confidence interval for $\kappa_t$, due to its overly optimistic estimation of the upper bound, which is driven by the excessive information reduction inherent in the SVD approach. In contrast, the KPCA-estimated $\kappa_t$ has a similar range to the MLE and auto-encoder methods. Notably, the KPCA method exhibits the lowest level of uncertainty when compared to the other two methods.

**Table 5.**  Projected $\kappa_t$ at 2021

|              | Lower   | Mean    | Higher  | Width of CI |
|--------------|---------|---------|---------|-------------|
| SVD          | −23.75  | −19.89  | −16.04  | 7.71        |
| MLE          | −24.28  | −19.52  | −14.77  | 9.51        |
| Auto-encoder | −25.52  | −20.64  | −15.76  | 9.76        |
| KPCA         | −23.24  | −18.58  | −13.92  | 9.32        |
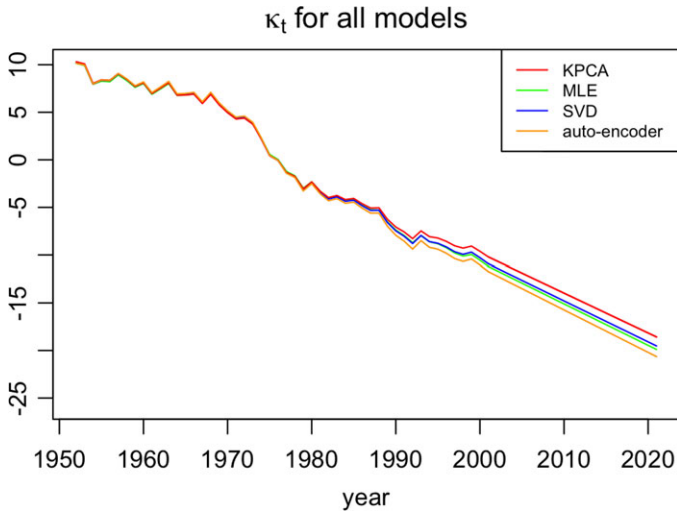


**Figure 4.** $\kappa_t$ estimated across models.
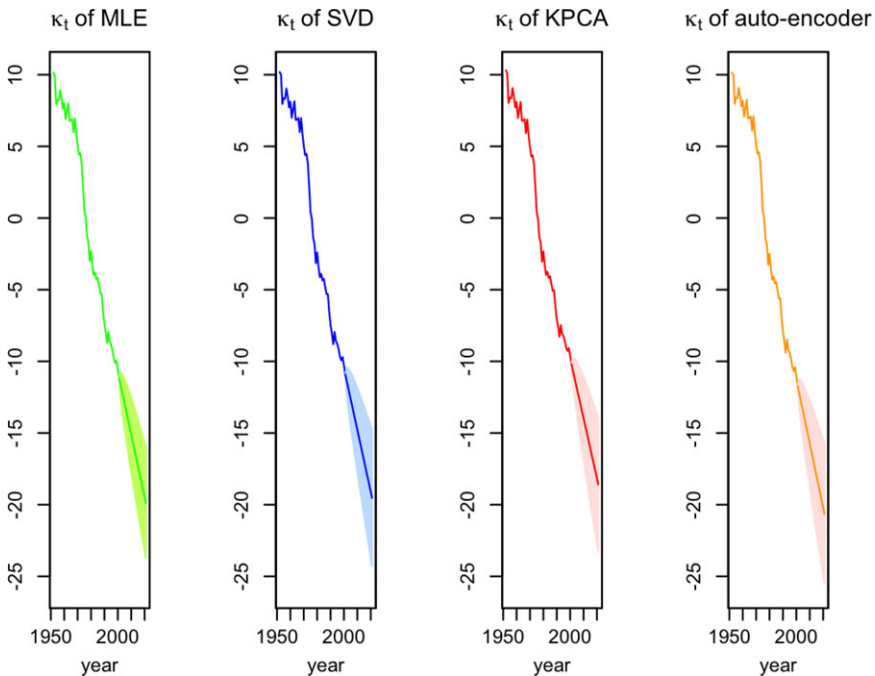


**Figure 5.** Fan chart of $\kappa_t$ estimated across models, 95% CI.

## $\beta_x^{(1)}$ for 2PCs KPCA model    $\beta_x^{(2)}$ for 2PCs KPCA model
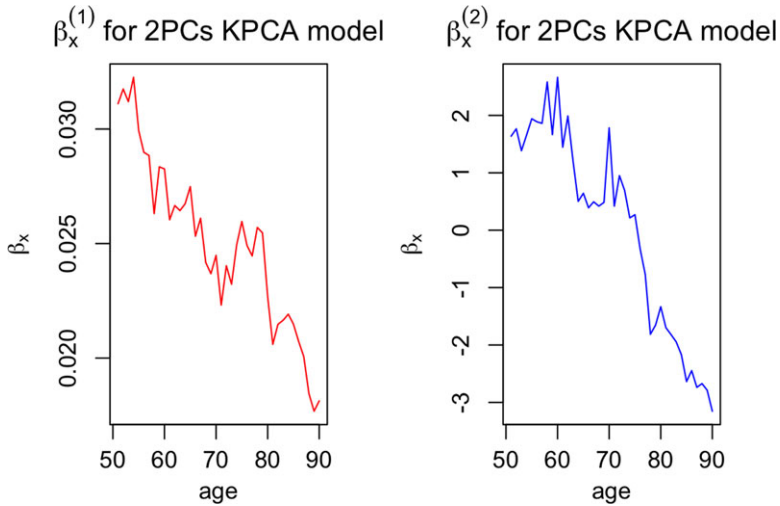


**Figure 6.** $\beta^{(1)}$ and $\beta^{(2)}$ estimated by 2 principal components Kernel Principal Component Analysis model.
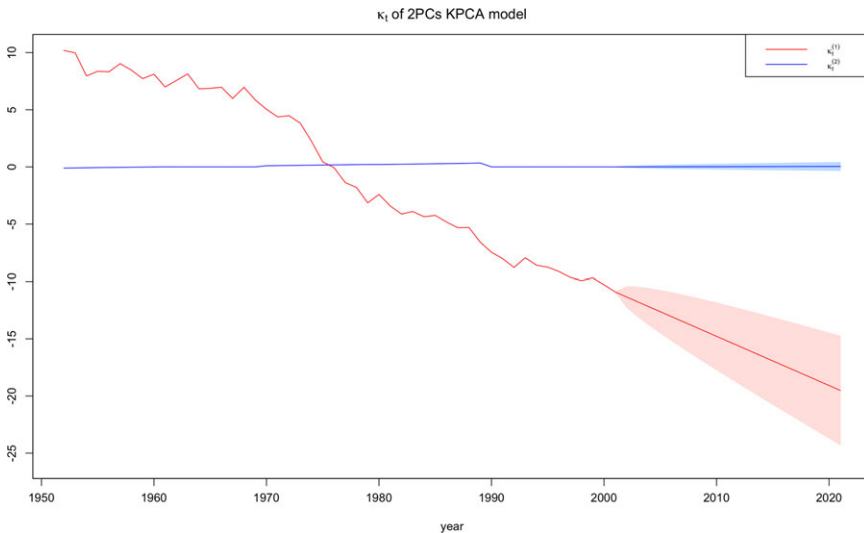


**Figure 7.** Fan plots of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ estimated by 2 principal components Kernel Principal Component Analysis model, 95% CI.

For 2PCs KPCA model, the fitted results of $\beta_x^{(1)}$ and $\beta_x^{(2)}$ are displayed in Figure 6, and prediction results of $\kappa_x^{(1)}$ and $\kappa_x^{(2)}$ are shown in Figure 7. From Figure 6, we can see that the estimated $\beta_x^{(1)}$ and $\beta_x^{(2)}$ follow similar decreasing trend as the original Lee-Carter model. From Figure 7, we can identify that the estimated $\kappa_t^{(2)}$ hovers near zero with a small variance, indicating its small impact on the overall model. This observation aligns with the findings in Table 3 and Table 4, where the forecasting error of the 1 PC and 2 PCs KPCA Lee–Carter model reveals that including the second PC doesn't yield a substantial improvement (only 0.4% improvement).

Our empirical results lead us to the conclusion that the KPCA method effectively enhances the forecast performance of the Lee–Carter model.

**Table 6.**    Model comparison for other countries

|       | SVD      | MLE      | KPCA 1 PC |
|-------|----------|----------|-----------|
| CAN   | 0.08792  | 0.08425  | 0.08372   |
| AUS   | 0.09524  | 0.10337  | 0.08373   |

## 4.  Robustness analysis

In this section, we test the robustness of KPCA in mortality prediction. Section 4.1 extends our analyses from the U.S. to Canada and Australia to ensure the geographical robustness of our empirical results. Section 4.2 compares the models analyzed in the empirical analyses using COVID-19 as a case study.

### 4.1  Geographical robustness

We apply the 1 PC KPCA model to mortality data from Canada and Australia using the same prediction procedure and age groups as the U.S. data and compare it with the SVD and MLE methods. The MAPE results in Table 6 demonstrate that KPCA is applicable to populations across different geographical locations.

### 4.2  Impact of COVID-19 mortality disruption on life expectancy projections

To evaluate the robustness of the KPCA model in extreme cases, in this subsection, we conduct an analysis of the model's performance during the COVID-19 pandemic, which presented an unprecedented disruption in mortality rates. The objective of this analysis is to assess how a disruption in mortality would affect the model's prediction performance. We created two sets of training datasets: (1) COVID-exclusive, using pre-COVID mortality data from 1952 to 2019; (2) COVID-inclusive, which includes mortality data from the COVID years, spanning from 1952 to 2021. To make the results of different models more tangible, we transformed the mortality projections into remaining life expectancy projections using the life table method (Dickson *et al.*, 2019).

Figure 8 depicts the projection of life expectancy under the two training sets. Table 7 presents summary statistics of these projections, including the mean, and the 20th and 80th quantiles of the life expectancy projections at age 60, with the two training sets across different models. We have the following interesting observations. First, the differences between predictions of average remaining life expectancy from samples including and excluding the COVID period are large and economically significant. In particular, the inclusion of the COVID period in the training sample significantly reduces the projected remaining life expectancy by about 2 years on average, ranging from 2.11 years (auto-encoder model) to 2.19 years (MLE model). The projections from KPCA and SVD are similar. However, the average reduction from the KPCA model is 2.13 years, slightly higher than auto-encoder model while lower than SVD model (2.15 years). The confidence interval width of the auto-encoder model remains the smallest both before and after the inclusion of the COVID period. Second, after including the COVID period, confidence intervals from all models are widened, but the width of the 20th-80th confidence interval from the KPCA model increases by 1.31 years, compared to 1.29 years (SVD) and 1.27 years (MLE & auto-encoder) from benchmarks. Although the KPCA model's confidence interval widens the most, we also notice that it remains narrower than those of traditional models such as SVD and MLE. This indicates that while the KPCA model maintains narrower prediction intervals, it effectively captures the increased uncertainty during the pandemic.

Moreover, the discrepancies between future life expectancy projections from different models are smaller in COVID-exclusive cases. In contrast, after including the COVID sample, the

**Table 7.**  Projected remaining life expectancy at age 60

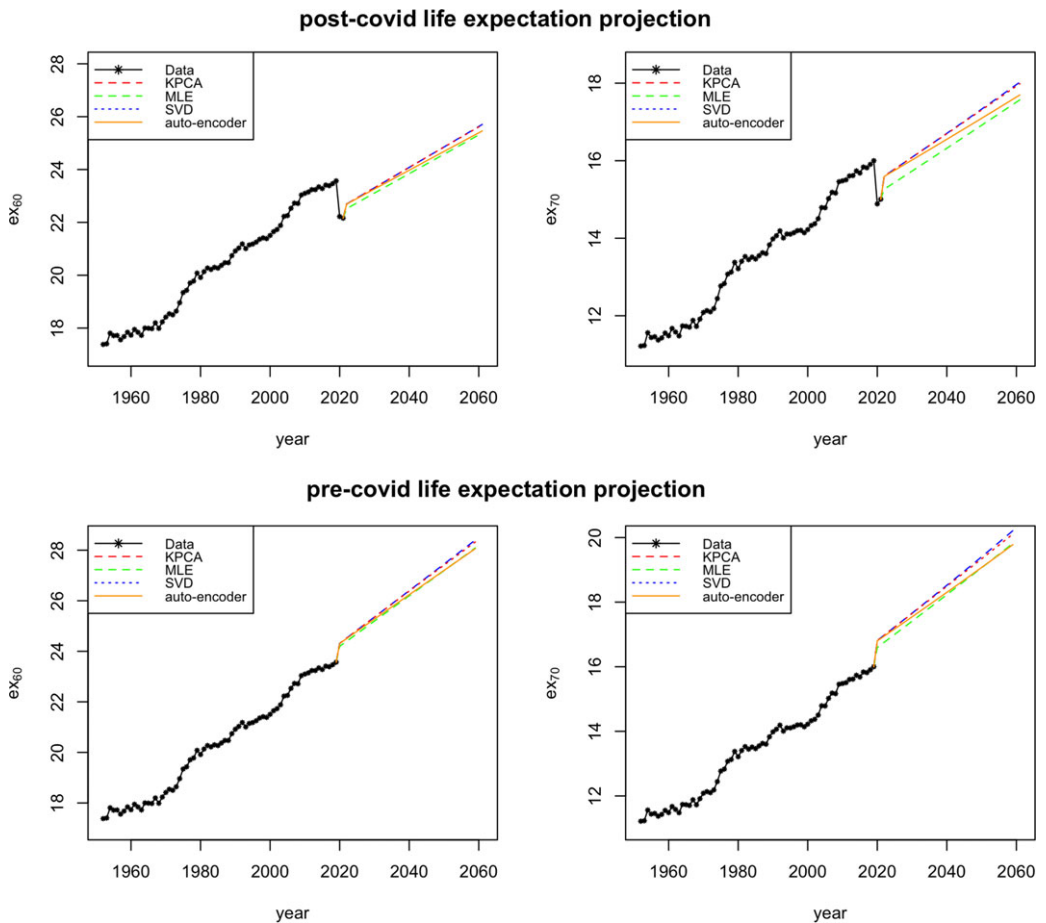| | 20th | Mean | 80th | Width of CI | Mean diff (Pre-post) | CI diff (Pre-post) |
|---|---|---|---|---|---|---|
| *Panel A: COVID inclusion* | | | | | | |
| SVD | 22.59 | 24.19 | 25.81 | 3.22 | 2.15 | 1.29 |
| MLE | 22.62 | 23.95 | 25.82 | 3.20 | 2.19 | 1.27 |
| Auto-encoder | 22.55 | 24.08 | 25.59 | 3.04 | 2.11 | 1.27 |
| KPCA | 22.59 | 24.19 | 25.78 | 3.19 | 2.13 | 1.31 |
| *Panel B: COVID exclusion* | | | | | | |
| | 20th | Mean | 80th | Width of CI | | |
| SVD | 25.39 | 26.35 | 27.32 | 1.93 | | |
| MLE | 25.40 | 26.14 | 27.33 | 1.93 | | |
| Auto-encoder | 25.30 | 26.19 | 27.07 | 1.77 | | |
| KPCA | 25.38 | 26.32 | 27.27 | 1.89 | | |



**Figure 8.** Projected remaining life expectancy based on COVID inclusion and exclusion datasets.

life expectancy predictions deviate more between different models. For example, the standard deviation of mean projection between different models is 0.10 years without the COVID data and increases to 0.12 years after the COVID period is included. This highlights the increased variability and uncertainty introduced by the pandemic. Finally, from our analysis, we can see that while the KPCA model maintains narrower prediction intervals, it effectively captures the increased uncertainty during the pandemic. This indicates that KPCA is more robust in extreme cases. In particular, KPCA model provides conservative and realistic projections even when faced with extreme disruptions. The model's predictions remain robust, with the intervals appropriately widening to reflect the increased variability during the COVID-19 period.

We acknowledge that due to the limited post-COVID sample, there are some limitations to our analysis in this subsection. It is interesting to include more observations after COVID when more data are available, to see the long-term impact of nonlinear disruptive events and the performance of different models. It's also interesting to develop new model that can incorporate the shocks from extreme events, e.g., an extension of the model in Zhou and Li (2022).

## 5. Conclusion

In this study, we introduced an extension of classic Lee–Carter model to enhance mortality forecasting by incorporating KPCA. Our empirical results demonstrate that the KPCA-enhanced Lee–Carter model significantly improves the accuracy of mortality rate predictions. Both RMSE and MAPE are consistently lower for the KPCA model compared to SVD and MLE. This enhancement in forecasting performance is due to the ability of KPCA Lee–Carter model to capture nonlinear patterns and complex relationships in mortality data. In addition, we observed that including more PCs beyond the first one in KPCA does not lead to substantial improvements in forecasting accuracy. The first PC effectively characterizes a significant portion of the variance in mortality data, highlighting the efficiency and practicality of our proposed KPCA Lee–Carter model. Moreover, our analysis indicates that our KPCA model is more robust to nonlinear disruptions in the extreme case and provides more reliable mortality forecasts with lower uncertainty. This robustness makes the KPCA model a reliable tool for stakeholders managing longevity risk, ensuring they can make well-informed decisions even during significant mortality shocks like the COVID-19 pandemic.

Our research highlights the potential of KPCA as a valuable tool for enhancing mortality forecasting. As global populations continue to age, and the challenges of longevity risk intensify, accurate mortality predictions are essential for governments, pension funds, insurance companies, and individuals. Our findings contribute to the growing body of literature where actuarial science intersects with statistical learning, offering practical solutions to the challenges posed by an aging world population. We acknowledge that due to the limited sample of post-COVID, there are some limitations to our analysis. It will be beneficial to include more observations as more data become available to better understand the long-term impact of nonlinear disruptive events and the performance of different models. Additionally, it is worth exploring the development of new models that can incorporate shocks from extreme events, e.g., an extension of the model in Zhou and Li (2022). We leave discussions of these interesting topics for future research.

## References

**Aronszajn**, **N.** (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**(3), 337–404.

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

**Blake**, **D.**, **Cairns**, **A.**, **Coughlan**, **G.**, **Dowd**, **K.** & **MacMinn**, **R.** (2013). The new life market. *Journal of Risk and Insurance*, **80**(3), 501–558.

**Booth**, **H.**, **Maindonald**, **J.** & **Smith**, **L.** (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, **56**(3), 325–336.

**Bravo**, **J. M.** (2021a). Forecasting longevity for financial applications: a first experiment with deep learning methods. In Joint European conference on machine learning and knowledge discovery in databases(pp. 232–249). Springer.

**Bravo**, **J. M.** (2021b). Forecasting mortality rates with recurrent neural networks: A preliminary investigation using Portuguese data. *CAPSI 2021 Proceedings*, **7**. https://aisel.aisnet.org/capsi2021/7

**Brock Porth**, **C.**, **Porth**, **L.**, **Zhu**, **W.**, **Boyd**, **M.**, **Tan**, **K. S.** & **Liu**, **K.** (2020). Remote sensing applications for insurance: a predictive model for pasture yield in the presence of systemic weather. *North American Actuarial Journal*, **24**(2), 333–354.

**Cairns**, **A. J.**, **Blake**, **D.**, **Dowd**, **K.**, **Coughlan**, **G. D.**, **Epstein**, **D.**, **Ong**, **A.** & **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.

**Cardillo**, **G.**, **Giordani**, **P.**, **Levantesi**, **S.** & **Nigri**, **A.** (2022). A tensor-based approach to cause-of-death mortality modeling. *Annals of Operations Research*, 1–20. https://link.springer.com/article/10.1007/s10479-022-05042-2#citeas

**Chen**, **Y.** & **Khaliq**, **A. Q.** (2022). Comparative study of mortality rate prediction using data-driven recurrent neural networks and the Lee–Carter model. *Big Data and Cognitive Computing*, **6**(4), 134.

**Chen**, **Z.**, **Lu**, **Y.**, **Zhang**, **J.** & **Zhu**, **W.** (2023). Managing weather risk with a neural network-based index insurance. *Management Science*, **70**(7), 4306–4327.

**Debener**, **J.**, **Heinke**, **V.** & **Kriebel**, **J.** (2023). Detecting insurance fraud using supervised and unsupervised machine learning. *Journal of Risk and Insurance*, **90**(3), 743–768.

**Deprez**, **P.**, **Shevchenko**, **P. V.** & **Wüthrich**, **M. V.** (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, **7**(2), 337–352.

**Desjardins**, **B.** & **Bourbeau**, **R.** (2010). The emergence of supercentenarians in Canada. In *Supercentenarians* (pp. 59–74). Springer.

**Devriendt**, **S.**, **Antonio**, **K.**, **Reynkens**, **T.** & **Verbelen**, **R.** (2021). Sparse regression with multi-type regularized feature modeling. *Insurance: Mathematics and Economics*, **96**, 248–261.

**Dickson**, **D. C.**, **Hardy**, **M. R.** & **Waters**, **H. R.** (2019). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

**Dong**, **Y.**, **Huang**, **F.**, **Yu**, **H.** & **Haberman**, **S.** (2020). Multi-population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, **2020**(8), 754–775.

**Fissler**, **T.**, **Merz**, **M.** & **Wüthrich**, **M. V.** (2023). Deep quantile and deep composite triplet regression. *Insurance: Mathematics and Economics*, **109**, 94–112.

**Gabrielli**, **A.** & **Wüthrich**, **M. V.** (2018). An individual claims history simulation machine. *Risks*, **6**(2), 29.

**Gaille**, **S.** & **Sherris**, **M.** (2011). Modelling mortality with common stochastic long-run trends. *The Geneva Papers on Risk and Insurance-Issues and Practice*, **36**(4), 595–621.

**Gao**, **G.**, **Meng**, **S.** & **Wüthrich**, **M. V.** (2019). Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal*, **2019**(2), 143–162.

**Gao**, **G.**, **Wang**, **H.** & **Wüthrich**, **M. V.** (2022). Boosting Poisson regression models with telematics car driving data. *Machine Learning*, **111**(1), 243–272.

**Ghahari**, **A.**, **Newlands**, **N. K.**, **Lyubchich**, **V.** & **Gel**, **Y. R.** (2019). Deep learning at the interface of agricultural insurance risk and spatio-temporal uncertainty in weather extremes. *North American Actuarial Journal*, **23**(4), 535–550.

**Gomes**, **C.**, **Jin**, **Z.** & **Yang**, **H.** (2021). Insurance fraud detection with unsupervised deep learning. *Journal of Risk and Insurance*, **88**(3), 591–624.

**Hainaut**, **D.** (2018). A neural-network analyzer for mortality forecast. *ASTIN Bulletin: The Journal of the IAA*, **48**(2), 481–508.

**Henckaerts**, **R.**, **Côté**, **M.-P.**, **Antonio**, **K.** & **Verbelen**, **R.** (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, **25**(2), 255–285.

**Hu**, **C.**, **Quan**, **Z.** & **Chong**, **W. F.** (2022). Imbalanced learning for insurance using modified loss functions in tree-based models. *Insurance: Mathematics and Economics*, **106**, 13–32.

**Jolliffe**, **I. T.** (2002). *Principal component analysis for special types of data*. Springer.

**Lally**, **N.** & **Hartman**, **B.** (2018). Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping. *Insurance: Mathematics and Economics*, **82**, 124–140.

**LeCun**, **Y.**, **Bengio**, **Y.** & **Hinton**, **G.** (2015). Deep learning. *Nature*, **521**(7553), 436–444.

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

**Lee**, **S. C.** (2021). Addressing imbalanced insurance data through zero-inflated Poisson regression with boosting. *ASTIN Bulletin: The Journal of the IAA*, **51**(1), 27–55.

**Lee**, S. C. & **Lin**, S. (2018). Delta boosting machine with application to general insurance. *North American Actuarial Journal*, **22**(3), 405–425.

**Levantesi**, S. & **Nigri**, A. (2020). A random forest algorithm to improve the Lee–Carter mortality forecasting: impact on q-forward. *Soft Computing*, **24**(12), 8553–8567.

**Lindholm**, M. & **Palmborg**, L. (2022). Efficient use of data for LSTM mortality forecasting. *European Actuarial Journal*, **12**(2), 749–778.

**Maier**, H., **Gampe**, J., **Jeune**, B., **Vaupel**, J. W. & **Robine**, J.-M. (2010). *Supercentenarians*. Springer.

**Marino**, M., **Levantesi**, S. & **Nigri**, A. (2023). A neural approach to improve the Lee-Carter mortality density forecasts. *North American Actuarial Journal*, **27**(1), 148–165.

**Meng**, S., **Gao**, Y. & **Huang**, Y. (2022). Actuarial intelligence in auto insurance: claim frequency modeling with driving behavior features and improved boosted trees. *Insurance: Mathematics and Economics*, **106**, 115–127.

**Msemburi**, W., **Karlinsky**, A., **Knutson**, V., **Aleshin-Guendel**, S., **Chatterji**, S. & **Wakefield**, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, **613**(7942), 130–137.

**Nigri**, A., **Levantesi**, S. & **Marino**, M. (2021). Life expectancy and lifespan disparity forecasting: a long short-term memory approach. *Scandinavian Actuarial Journal*, **2021**(2), 110–133.

**Nigri**, A., **Levantesi**, S., **Marino**, M., **Scognamiglio**, S. & **Perla**, F. (2019). A deep learning integrated Lee–Carter model. *Risks*, **7**(1), 33.

**Noll**, A., **Salzmann**, R. & **Wuthrich**, M. V. (2020). Case study: French motor third-party liability claims. Available at SSRN, **3**, 164764.

**Perla**, F., **Richman**, R., **Scognamiglio**, S. & **Wüthrich**, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, **2021**(7), 572–598.

**Petneházi**, G. & **Gáll**, J. (2019). Mortality rate forecasting: can recurrent neural networks beat the Lee-Carter model? arXiv preprint arXiv:1909.05501.

**Rau**, R., **Soroko**, E., **Jasilionis**, D. & **Vaupel**, J. W. (2008). Continued reductions in mortality at advanced ages. *Population and Development Review*, **34**(4), 747–768.

**Richman**, R. (2021a). AI in actuarial science–a review of recent advances–part 1. *Annals of Actuarial Science*, **15**(2), 207–229.

**Richman**, R. (2021b). AI in actuarial science–a review of recent advances–part 2. *Annals of Actuarial Science*, **15**(2), 230–258.

**Richman**, R. & **Wuthrich**, M. V. (2019). Lee and Carter go machine learning: recurrent neural networks. Available at SSRN 3441030.

**Richman**, R. & **Wüthrich**, M. V. (2021). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science* **15**(2), 346–366.

**Schnürch**, S. & **Korn**, R. (2022). Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin: The Journal of the IAA*, **52**(1), 333–360.

**Shlens**, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

**Wang**, C. W., **Zhang**, J. & **Zhu**, W. (2021). Neighbouring prediction for mortality. *ASTIN Bulletin: The Journal of the IAA*, **51**(3), 689–718.

**Wang**, J., **Wen**, L., **Xiao**, L. & **Wang**, C. (2024). Time-series forecasting of mortality rates using transformer. *Scandinavian Actuarial Journal*, **2024**(2), 109–123.

**Wang**, T., **Zhao**, D. & **Tian**, S. (2015). An overview of kernel alignment and its applications. *Artificial Intelligence Review*, **43**, 179–192.

**Weiße**, A., **Wellein**, G., **Alvermann**, A. & **Fehske**, H. (2006). The kernel polynomial method. *Reviews of modern physics*, **78**(1), 275–306.

**Wilmoth**, J. R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report, Department of Demography, University of California, Berkeley.

**Woolf**, S. H. & **Schoomaker**, H. (2019). Life expectancy and mortality rates in the United States, 1959-2017. *JAMA*, **322**(20), 1996–2016.

**World Bank** (2024). Life expectancy at birth, total (years). Retrieved 2024-11-16 from https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=1W

**Wüthrich**, M. V. (2017). Covariate selection from telematics car driving data. *European Actuarial Journal*, **7**(1), 89–108.

**Wuthrich**, M. V. & **Buser**, C. (2021). Data analytics for non-life insurance pricing. *Swiss Finance Institute Research Paper*, 16–68.

**Xin**, X. & **Huang**, F. (2024). Antidiscrimination insurance pricing: regulations, fairness criteria, and models. *North American Actuarial Journal*, 28(2), 285–319.

**Zhang**, N., **Chen**, H. & **Liu**, J. (2022). Mortality forecasting using LSTM-CNN model. Available at SSRN 4261735.

**Zhou**, R. & **Li**, J. S. H. (2022). A multi-parameter-level model for simulating future mortality scenarios with COVID-alike effects. *Annals of Actuarial Science*, **16**(3), 453–477.