Research Article

Compositional Data Analysis for Modelling and Forecasting Mortality with the α-Transformation

Han Ying Lim[1], Dharini Pathmanathan[1,2,3], Sophie Dabo-Niang[4,5]

1. Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, Malaysia; 2. Universiti Malaya Centre for Data Analytics, Universiti Malaya, Malaysia; 3. Center of Research for Statistical Modelling and Methodology, Universiti Malaya, Malaysia; 4. UMR8524-Laboratoire Paul Painlevé, Inria-MODAL, University of Lille, CNRS, France; 5. CRM-CNRS, Université de Montréal, Canada

Mortality forecasting is crucial for demographic planning and actuarial studies, particularly for predicting population ageing rates and future longevity risks. Traditional approaches largely rely on extrapolative methods, such as the Lee-Carter model and its variants which use mortality rates as inputs. In recent years, compositional data analysis (CoDA), which adheres to summability and non-negativity constraints, has gained increasing attention from researchers for its application in mortality forecasting. This study explores the use of the α-transformation as an alternative to the commonly applied centered log-ratio (CLR) transformation for converting compositional data from the Aitchison simplex to unconstrained real space. The α-transformation offers greater flexibility through the inclusion of the α parameter, enabling better adaptation to the underlying data structure and handling of zero values, which are the limitations inherent to the CLR transformation. Using age-specific life table death counts for males and females in 31 selected European countries/regions from 1983 to 2018, the proposed method demonstrates comparable performance to the CLR transformation in most countries with improved forecast accuracy in some cases. These findings highlight the potential of the α-transformation as a competitive alternative transformation technique for real-world mortality data within a non-functional CoDA framework.

Corresponding author: Dharini Pathmanathan, dharini@um.edu.my

1. Introduction

Mortality forecasting plays a pivotal role in demographic analysis, informing strategic planning across critical sectors, including healthcare, insurance, workforce management and social welfare. Over the years, numerous studies have been conducted to enhance forecast accuracy and minimize the risks associated with inaccurate mortality predictions. These approaches are generally classified into expert judgment, extrapolative models and epidemiological models[1]. Among these, the Lee-Carter (LC) model has been widely regarded as a benchmark in mortality forecasting since its establishment. Following that, numerous variants and applications have been developed over the decades[2].

The success of LC model is attributed to its powerful mechanism that employs a single time-varying mortality index along with statistical time series models to project age-specific mortality trends. This is a purely extrapolative process with minimal subjective judgment[2]. The LC framework extracts the time index through singular vector decomposition

 doi.org/10.32388/Z69TVG

method of the logarithm of age-specific mortality rates, in order to forecast the density of deaths in the life table[3][4][5]. It leverages the approximately log-linear decline in mortality rates by age over time and allows the application of multivariate statistical techniques for unbounded variables[6].

In addition to age-specific mortality rates, alternative demographic indicators have been used as inputs of the model, such as death probabilities[7][8][9], age-at-death distributions[10][11][6][12], survival probabilities[13] and life expectancy at birth[14]. Among these indicators, Bergeron-Boucher et al.[15] observed that while death rates and death probabilities generally yield similar trends, they often lead to more pessimistic forecasts compared to other indicators. As different indicators may influence the outcomes, their selection should align with the research interests[15]. Since the age-at-death distribution provides information on the mortality conditions[12], central measures of longevity[16] and lifespan variability[17], it has gained increasing attention.

Compositional data are positive vectors carrying relative information, representing part of a whole with a fixed sum[18]. Some commonly found compositional data include geochemical elements, atmospheric components and food compositions. Pioneered by Oeppen[6], the compositional data analysis (CoDA) has been adopted in mortality forecasting by treating age-specific life table death counts $d_{t,x}$ as compositional data. This approach models and forecasts the redistribution of the density of $d_{t,x}$, in which deaths are shifted gradually from younger to older ages. Such technique is deemed appropriate because $d_{t,x}$ values are bounded between 0 and the life table radix, summing to the radix for each year. The bounded nature of compositional data leads to the need of deploying transformations to map them from constrained simplex to unconstrained real space, where the transformed values can vary freely between $(-\infty, \infty)$. In this context, Aitchison[18] has introduced log-ratio transformations, with the centered log-ratio transformation (CLR) being widely adopted for its interpretability and ability to preserve distances.

Oeppen's[6] CoDA framework has inspired subsequent extensions with the aim to improve forecast accuracy. Bergeron-Boucher et al.[10] adapted it for coherent mortality forecasting akin to the Li-Lee model[5]. Their findings indicate that both coherent and non-coherent CoDA models produce less biased forecasts with increased accuracy for many of the selected Western European countries, as compared to their LC model counterparts. This is partly attributed to the changing rate of mortality improvement over time as a consequence of the use of $d_{t,x}$ as input along with the CLR transformation. The summability constraint in compositional data also preserves coherence across populations, addressing the limitations of the LC-based models.

Building on this, Kjærgaard et al.[11] applied CoDA to forecast causes of cancer death, so that the dependency between causes is accounted for at the aggregate level. Shang and Haberman[12], on the other hand, introduced weights within a functional CoDA framework. The findings from these studies have proven the potential of CoDA method in improving the mortality forecasting accuracy. Nevertheless, most of them employ the log-ratio approaches to transform $d_{t,x}$. Such methods have a major shortcoming where they fail to handle zero values due to the nature of the logarithm.

To address this limitation, Tsagris et al.[19] proposed a Box-Cox type transformation that is capable of handling zeros[20][21], namely the α-transformation. This method offers greater flexibility in terms of the degree of transformation via the parameter α, ranging from an isometric log-ratio (ILR) transformation to a linear transformation[20][19]. This parameter can be fine-tuned to fit different datasets, which in turns help in improving the accuracy. While the α-transformation violates the properties such as perturbation and permutation invariance and subcompositional dominance and coherence suggested by Aitchison[18], this does not affect its applicability since these properties are meant to support the log-ratio transformations[19][22]. The practicality of α-transformation has been demonstrated in various contexts, including regression[21], classification[23] and forecasting[24][25] settings.

This paper extends prior research by evaluating the α-transformation within a non-functional CoDA framework. Using CLR transformation as a benchmark, it is aim to compare forecast accuracy between the two transformations. The analysis focuses on male and female populations across 31 European countries/regions from year 1983 to 2018 using the data retrieved from the Human Mortality Database (HMD)[26]. These countries/regions are selected to obtain the longest possible common timeframe.

Section 2 provides a detailed description of methodology, covering the key steps and analytical framework employed in this study. Subsequently, Section 3 presents the results, along with a comprehensive discussion revolving around the forecast accuracy of each transformation. Lastly, Section 4 concludes the paper by summarizing key findings and suggesting possible extensions for future research.

2. Methodology

2.1. Data preprocessing

Similar to the pipeline of Bergeron-Boucher et al.[10], age-specific mortality rates over the years M_x for each country are first calculated based on the observed death counts D_x and exposure-to-risk estimates E_x provided by the HMD. However, the observed M_x at older ages have considerable random variation as a consequence of E_x being smaller than D_x[27].

To address these anomalies, the Kannisto model of old-age mortality is fitted to smooth the mortality rates of ages 80 and above, up to the open age interval of 110+, separately for male and female[27]. This smoothing model that is originally proposed by Thatcher et al.[28] utilizes a Poisson log-likelihood procedure where the logistic curve was found to better fit the old-age mortality patterns compared to other mortality models. Such approach also helps avoid the presence of zeros and missing values at older ages.

For ages below 80, zeros are still present in the observed M_x for some countries, resulting from zeros in D_x, can lead to undefined results in CLR transformation due to the nature of logarithm. In this context, a multiplicative replacement strategy is applied on the observed death counts to remove zeros before constructing life tables, as implemented in[10]. Even though the α-transformation can theoretically accommodate zeros, zero imputation is performed uniformly across transformations to compare forecast accuracy consistently.

Basically, a composition $\mathbf{x} = [x_1, x_2, \ldots, x_D]$ of D_x containing zeros is replaced by a composition $\mathbf{r} = [r_1, r_2, \ldots, r_D]$ without zeros as follows[10]:

$$r_j = \begin{cases} \delta, & \text{if } x_j = 0, \\ \left(1 - \frac{z\delta}{\sum x_j}\right) x_j, & \text{if } x_j > 0, \end{cases}$$

 doi.org/10.32388/Z69TVG

where z is the number of zeros counted in $\mathbf{x}$ and δ is the imputed value for part x_j computed as follows:

$$\delta = \frac{\left(\min_{t,x} D_{t,x}\right)/2}{\sum_{x=0}^{110} D_{t,x}}, \quad \forall D_{t,x} > 0.$$

Subsequently, $\mathbf{r}$ is multiplied by $\sum_{x=0}^{110} D_{t,x}$ to obtain a new set of death counts without zeros which are then used to calculate mortality rates for ages below 80[10]. Combining these with the mortality rates for ages 80 and above, a smoothed and imputed set of observed mortality rates is ready to be used for constructing life tables.

To meet the constant-sum constraint in CoDA, the life table radix is assumed to be unity, ensuring that $d_{t,x}$ values fall within a standard simplex[29][6][10]. For visualization purposes, $d_{t,x}$ values are multiplied by 100,000, which is a commonly used radix in demographic research[27][12][25].

The average number of years lived within the age interval $[x, x+1)$ for people dying at that age, denoted as a_x, is assumed to be 0.5 for all single-year ages except age 0[10][27]. Unlike in[10] where the infant a_0 is fixed across all countries, the value of a_0 in this study is computed in a similar manner to that employed by the HMD[27]. The gender-specific a_0 for each country is computed as the average of a_0 values, which are calculated based on the range of infant life table mortality rates m_0 for each year in the training set, as outlined in Table 1. This method that is inspired by the core idea suggested by Andreev and Kingkade[30], incorporates averaging to allow for country-specific adjustments which reflects the infant mortality more accurately than a fixed a_0 value across all countries and genders.

Subsequently, country-specific life tables are constructed from the preprocessed mortality rates separately for males and females, using a readily available LifeTable() function in the R package called `MortalityLaws`[31].

Gender	m_0 range	Formula
Male	$[0, 0.02300)$	$\frac{1}{N}\sum_{t=1}^{N}(0.14929 - 1.99545 m_{t,0})$
	$[0.02300, 0.08307)$	$\frac{1}{N}\sum_{t=1}^{N}(0.02832 + 3.26021 m_{t,0})$
	$[0.08307, \infty)$	0.29915
Female	$[0, 0.01724)$	$\frac{1}{N}\sum_{t=1}^{N}(0.14903 - 2.05527 m_{t,0})$
	$[0.01724, 0.06891)$	$\frac{1}{N}\sum_{t=1}^{N}(0.04667 + 3.88089 m_{t,0})$
	$[0.06891, \infty)$	0.31411

Table 1. Formulas for computing a_0 based on m_0.

Note: N here refers to the length of the training set, which is equivalent to 28 years (1983-2010) in this context.

2.2. *CLR transformation*

A positive compositional data vector typically satisfies a unit sum constraint and lies within a sample space called the standard simplex[20][19], defined by

$$\mathbb{S}^D = \left\{ \mathbf{x} = [x_1, \ldots, x_D] \in \mathbb{R}^D \mid x_i > 0, \sum_{i=1}^{D} x_i = 1 \right\}.$$

Positive points can be mapped to the simplex $\mathbb{S}^D$ using the closure operator[20][24] defined as:

$$C : (0, \infty)^D \to \mathbb{S}^D,$$

$$C[\mathbf{x}] = \left[\frac{x_1}{\sum_{i=1}^{D} x_i}, \ldots, \frac{x_D}{\sum_{i=1}^{D} x_i} \right].$$

When the simplex is equipped with Aitchison geometry and its associated operations, it is referred to as the Aitchison simplex which forms a vector space[18]. Some key features of Aitchison geometry[18][10][20] are:

$$\begin{aligned}
\text{Perturbation: } & \mathbf{x} \oplus \mathbf{y} = C[x_1 y_1, \ldots, x_D y_D] \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \\
\text{Power transformation: } & \alpha \odot \mathbf{x} = C[x_1^\alpha, \ldots, x_D^\alpha] \quad \forall \mathbf{x} \in \mathbb{S}^D \text{ and } \alpha \in \mathbb{R}, \\
\text{Negative perturbation: } & \mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus (-1 \odot \mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \\
\text{Inner product: } & \langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(\ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} \right) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \\
\text{Norm: } & \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a} \quad \forall \mathbf{x} \in \mathbb{S}^D, \\
\text{Distance: } & d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D.
\end{aligned}$$

Since compositional data carry only relative proportions, Aitchison[18] introduced log-ratio based transformations, including the widely used centered log-ratio (CLR) transformation. The CLR transformation[18] that maps the simplex onto a hyperplane passing through the origin of $\mathbb{R}^D$ is defined as:

$$\begin{aligned}
& \text{clr: } \mathbb{S}^D \to \mathbb{H} \subset \mathbb{R}^D, \\
& \mathbf{w} = \text{clr}(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ldots, \ln \frac{x_1}{g(\mathbf{x})} \right], \\
& \text{clr}^{-1}(\mathbf{w}) = C\left[e^{\mathbf{w}} \right],
\end{aligned}$$

where $g(\mathbf{x}) = \left(\prod_{i=1}^{D} x_i \right)^{1/D}$ is the geometric mean of the composition.

The CLR transformation is a one-to-one mapping between $\mathbb{S}^D$ and $\mathbb{R}^D$ under a zero-sum constraint. It also preserves distances where $d_a \langle \mathbf{x}, \mathbf{y} \rangle = d_e(\text{clr}(\mathbf{x}), \text{clr}(\mathbf{y}))$ with d_e representing the Euclidean distance[20].

2.3. α *transformation*

Proposed by Tsagris et al.[19], the α-transformation is a one-parameter Box-Cox type transformation that maps compositional data $\mathbf{x}$ from D-dimensional Aitchison simplex $\mathbb{S}^D$ to $(D-1)$-dimensional unconstrained real space $\mathbb{R}^{D-1}$ using the Helmert sub-matrix $\mathbf{H}_D$, as defined below:

$$A_\alpha : \mathbb{S}^D \to \mathbb{R}^{D-1}.$$

Given a compositional vector $\mathbf{x} \in \mathbb{S}^D$, the transformation and its inverse are expressed as:

$$\begin{aligned}
& \mathbf{z}_\alpha = A_\alpha(\mathbf{x}) = \frac{1}{\alpha} \mathbf{H}_D \left(D \mathbf{u}_\alpha(\mathbf{x}) - \mathbf{1}_D \right), \\
& A^{-1}(\mathbf{z}_\alpha) = C\left[\left(\alpha \mathbf{H}'_D \mathbf{z}_\alpha + \mathbf{1}_D \right)^{1/\alpha} \right],
\end{aligned}$$

where $\mathbf{z}_\alpha \in \mathbb{R}^{D-1}$, $\mathbf{u}_\alpha(\mathbf{x}) = \left[\frac{x_1^\alpha}{\sum_{j=1}^{D} x_j^\alpha}, \ldots, \frac{x_D^\alpha}{\sum_{j=1}^{D} x_j^\alpha} \right]$ is the power transformed vector in $\mathbb{S}^D$, $\mathbf{H}_D$ is the $(D-1) \times D$ Helmert sub-matrix and $\mathbf{1}_D$ is the D-dimensional vector of ones. Note that $\mathbf{H}_D$ satisfies

$\mathbf{H}_D\mathbf{H}'_D = \mathbf{I}_{D-1}$ and $\mathbf{H}'_D\mathbf{H}_D = \mathbf{G}_D$ where $\mathbf{G}_D = \mathbf{I}_D - \frac{1}{D}\mathbf{1}_D$ is the D-dimensional centering matrix[20].

The parameter $\alpha \in [0,1]$ can be tuned using criteria tailored to the type of analysis, such as the pseudo-R^2, profile log-likelihood (excluding zeros) or Kullback-Leibler divergence[19][21]. As noted by Tsagris et al.[19], when α approaches 1, it simplifies to a scaled Euclidean transformation. When $\alpha = 0$, it is equivalent to the ILR transformation, which requires the data to be free of zeros for it to be well-defined.

$$\text{ilr: } \mathbb{S}^D \to \mathbb{R}^{D-1}, \quad \text{ilr}(\mathbf{x}) = \mathbf{H}_D \text{clr}(\mathbf{x}).$$

2.4. Modeling

Following the approach of Bergeron-Boucher et al.[10], after constructing the life tables, $d_{t,x}$ are extracted to form a matrix $\mathbf{D}_{T\times(X+1)}$ for each country and gender, where T rows representing the number of years and $X+1$ columns representing the ages $x = 0, 1, \ldots, 110+$. Obeying the summability constraint of compositional data, each row sums up to the life table radix.

The matrix $\mathbf{D}$ is then centered by subtracting the column-specific geometric means α_x for each age using the negative perturbation operator, resulting in matrix $\mathbf{F}$. Since compositional data are constrained between 0 and the radix, a transformation is applied to project them into unconstrained real space. For this purpose, the α-transformation with a fine-tuned α and the CLR transformation are applied separately to $\mathbf{F}$, forming matrix $\mathbf{H}$.

Subsequently, singular vector decomposition (SVD) is applied to matrix $\mathbf{H}$. A rank-1 approximation is used to extract the time index κ_t and those corresponds to the training period are extrapolated using a forecasting model. Note that the data are split into a training set and test set by a typically chosen ratio of 80:20[32], covering the years 1983 to 2010 (28 years) for training and the remaining 8 years for testing. The rank-1 approximation is considered adequate as supported by the prior studies[10][6].

The approximately linear characteristic of κ_t allows it to be forecast using an ARIMA model, particularly the random walk with drift which gives a good fit[4]. In addition to that, previous research[10] has shown that ARIMA (0,1,1) with drift performs well for most Western European countries. Thus, two ARIMA models are fitted to the data in this study, namely ARIMA (0,1,1) with drift and the automatic ARIMA model proposed by Hyndman and Khandakar[33] that uses step-wise algorithmic selection of the best order.

After forecasting, the inverse of the respective transformations is applied to the matrix to convert it back to compositional data, forming matrix $\mathbf{F}^*$. Finally, the perturbation operator is used to add back the column-specific geometric means to matrix $\mathbf{F}^*$, obtaining the fitted $\hat{d}_{t,x}$ values in matrix $\mathbf{D}^*$. A fitted life table can then be constructed based on $\hat{d}_{t,x}$.

2.5. Accuracy measures

The predictive performance of models fitted to data under both transformations is compared between the actual and forecast life table death counts using root mean squared error (RMSE) and mean absolute error (MAE). The model that yields lower out-of-sample forecast errors is considered as the best model for each country and is chosen for further comparative analysis.

3. Application to Real Data

3.1. α-parameter tuning

As this study focuses on predictive analysis of mortality, the value of α should be chosen via cross-validation on a data-driven basis[21]. In order to determine the optimal α value for transforming $d_{t,x}$ for each country, an expanding window approach[32] is adopted. Starting with a subset of 15 years of data, the training window size is increased by one year each time, with a fixed size of 4 years of validation data. This results in a total of ten iterations.

Making good use of the optim() function in R[34], optimal α values within the range of $[0, 1]$ are selected to minimize the average RMSE in the validation data. It is important to note that penalization is applied to α values resulting in negative detection limits by excluding them from consideration in order to avoid returning negative $d_{t,x}$.

Taking female data as an example, Table 2 tabulates the α values chosen for the best model that yields higher forecast accuracy. Interestingly, only five countries have $\alpha = 0$ as the optimal value, indicating that the α-transformation outperforms the log-ratio transformations in these cases[24]. This is because ILR, the special case when α approaches 0, simply involves multiplying CLR by the Helmert sub-matrix[25].

doi.org/10.32388/Z69TVG

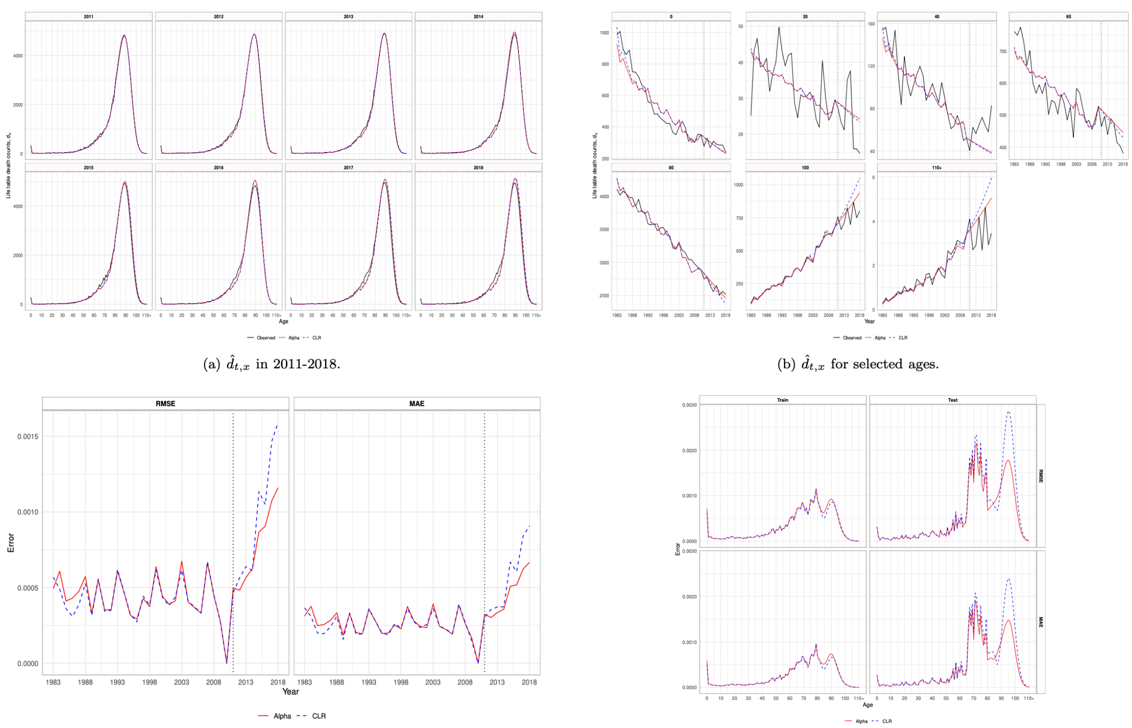
Country	Optimal α	Average validation RMSE
Austria	0.257989	0.000485
Belgium	0.592753	0.000492
Bulgaria	0.099998	0.000997
Belarus	0.100013	0.000866
Switzerland	0.099967	0.000504
Czechia	0.099954	0.000565
East Germany	0.683139	0.000532
West Germany	0.524746	0.000378
Denmark	0.099999	0.000691
Spain	0.099964	0.000482
Estonia	0.099989	0.001215
Finland	0.000000	0.000838
France	0.547718	0.000483
England & Wales	0.000000	0.000478
Northern Ireland	0.100039	0.001273
Scotland	0.099990	0.000610
Greece	0.099967	0.000567
Hungary	0.099980	0.000575
Ireland	0.000000	0.000989
Iceland	1.000000	0.002776
Italy	0.514915	0.000481
Lithuania	0.099994	0.000772
Luxembourg	0.099984	0.001799
Latvia	0.099989	0.000911
Netherlands	0.099992	0.000667
Norway	0.100001	0.000677
Poland	0.000000	0.000558
Portugal	0.000000	0.000659
Slovakia	0.099995	0.000710

Country	Optimal α	Average validation RMSE
Slovenia	0.100024	0.000883
Sweden	0.100003	0.000470

Table 2. Optimal α values for female mortality data.

3.2. *Forecasts of mortality*

The country-specific model accuracy shows that the α-transformation performs comparably to the CLR transformation in 25 countries for females and 23 countries for males. Taking Austrian female life table death counts as an example, Figures 1a to 1b illustrate the forecast $\hat{d}_{t,x}$ based on both α and CLR transformations for the holdout period and for selected 20-year-gapped ages over the years, respectively.



(a) $\hat{d}_{t,x}$ in 2011-2018.

(b) $\hat{d}_{t,x}$ for selected ages.

(c) Forecast errors over years.

(d) Forecast errors by age.

Figure 1. Mortality forecasting results for females in Austria.

In general, the life table death counts are negatively skewed and bimodal, with peaks in infancy that later shift towards older ages[12][35]. Infant mortality is primarily driven by genetic errors, infectious diseases or exposure while aging becomes the main cause of mortality at older ages[35]. According to Abouzahr et al.[36], high mortality rates are typically found during infancy and reach their lowest levels between ages 5 and 14, before rising exponentially beyond age 35.

doi.org/10.32388/Z69TVG

There might also have some bumps during female reproductive ages, indicating premature mortality due to maternal deaths.

As the years progress in the test set, forecasts based on α-transformed data retain the age patterns in $d_{t,x}$, similar to those predicted based on CLR-transformed data. Despite yielding similar forecasts, the α-transformation results in lower forecast errors as depicted in Figure 1c. Furthermore, forecasts from the α-transformed data are more accurate, especially beyond age 65, with a slight decrease in accuracy between ages 86 and 90 as shown in Figure 1d. More results for females and males can be found in the supplementary materials due to the limited space.

3.3. Comparison of mean forecast accuracy

3.3.1. Overall mean errors

Table 3 summarizes the overall mean forecast errors for female mortality in both training and test sets. It is observed that although α-transformation has a slightly higher training error with an extremely small difference, it yields more accurate forecasts in the test set. Similar findings are obtained in [25][24], in which α-transformation outperforms conventional log-ratio transformations. This is primarily due to the flexibility introduced by the parameter α in the Box-Cox type power transformation which allows it to better adapt to the underlying data structure [19], particularly the temporal changes in age-specific life table death counts [25].

Phase	**RMSE**		**MAE**	
	α	**CLR**	α	**CLR**
Train	0.000715	0.000700	0.000392	0.000385
Test	0.000812	0.000866	0.000477	0.000502

Table 3. Overall mean forecast errors for female mortality.

3.3.2. Mean errors by country

Breaking down model forecast accuracy of female mortality at the country level, Table 4 presents RMSE and MAE values for both the training and test sets under the α and CLR transformations, along with their best ARIMA models. Notably, ARIMA (0,1,1) with drift demonstrates superior performance across most countries compared to the models selected by the auto ARIMA algorithm. This result further supports that ARIMA (0,1,1) with drift is highly suitable for modeling κ_t for most of the selected European countries, consistent with findings reported in [10].

doi.org/10.32388/Z69TVG

Country	RMSE				MAE				ARIMA Model	
	Train		Test		Train		Test			
	α	CLR	α	CLR	α	CLR	α	CLR	α	CLR
Austria	0.000455	0.000440	0.000811	0.001024	0.000258	0.000249	0.000454	0.000553	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Belgium	0.000520	0.000481	0.000514	0.000615	0.000285	0.000265	0.000319	0.000367	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Bulgaria	0.001000	0.001000	0.001119	0.001124	0.000456	0.000457	0.000623	0.000625	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Belarus	0.000998	0.000993	0.002012	0.002028	0.000563	0.000561	0.001222	0.001231	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Switzerland	0.000471	0.000469	0.000472	0.000492	0.000265	0.000264	0.000290	0.000299	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Czechia	0.000507	0.000504	0.000541	0.000574	0.000276	0.000276	0.000320	0.000338	ARIMA(1,1,0) with drift	ARIMA(1,1,0) with drift
East Germany	0.000836	0.000669	0.000732	0.000927	0.000438	0.000346	0.000425	0.000492	ARIMA(0,1,0) with drift	ARIMA(0,1,0) with drift
West Germany	0.000414	0.000405	0.000554	0.000683	0.000208	0.000203	0.000298	0.000353	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Denmark	0.000643	0.000646	0.001103	0.001079	0.000363	0.000364	0.000615	0.000602	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Spain	0.000446	0.000437	0.000556	0.000626	0.000239	0.000236	0.000297	0.000328	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Estonia	0.000971	0.000980	0.001097	0.001069	0.000576	0.000582	0.000693	0.000675	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Finland	0.000840	0.000840	0.000539	0.000539	0.000440	0.000440	0.000327	0.000327	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
France	0.000441	0.000295	0.000420	0.000806	0.000216	0.000167	0.000251	0.000422	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
England & Wales	0.000438	0.000438	0.000556	0.000556	0.000232	0.000232	0.000314	0.000314	ARIMA(3,1,0) with drift	ARIMA(3,1,0) with drift
Northern Ireland	0.000875	0.000875	0.001019	0.001080	0.000501	0.000500	0.000587	0.000623	ARIMA(2,1,0) with drift	ARIMA(1,1,0) with drift

Country	RMSE				MAE				ARIMA Model	
	Train		Test		Train		Test			
	α	CLR	α	CLR	α	CLR	α	CLR	α	CLR
Scotland	0.000740	0.000743	0.000592	0.000608	0.000418	0.000418	0.000365	0.000374	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Greece	0.000548	0.000551	0.000631	0.000646	0.000290	0.000291	0.000372	0.000379	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Hungary	0.000518	0.000519	0.000615	0.000638	0.000278	0.000280	0.000372	0.000381	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Ireland	0.000694	0.000694	0.001012	0.001012	0.000397	0.000397	0.000574	0.000574	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Iceland	0.002126	0.002133	0.001828	0.001819	0.001165	0.001168	0.001005	0.001031	ARIMA(0,1,2)	ARIMA(0,1,1) with drift
Italy	0.000445	0.000373	0.000627	0.000992	0.000224	0.000195	0.000333	0.000499	ARIMA(0,1,1) with drift	ARIMA(1,1,0) with drift
Lithuania	0.000914	0.000882	0.000847	0.000833	0.000536	0.000523	0.000549	0.000542	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Luxembourg	0.001470	0.001470	0.001621	0.001652	0.000837	0.000841	0.000916	0.000939	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Latvia	0.000833	0.000839	0.001110	0.001103	0.000504	0.000508	0.000729	0.000725	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Netherlands	0.000428	0.000429	0.000436	0.000449	0.000230	0.000229	0.000265	0.000271	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Norway	0.000643	0.000646	0.000512	0.000505	0.000353	0.000354	0.000310	0.000302	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Poland	0.000499	0.000499	0.000564	0.000564	0.000253	0.000253	0.000320	0.000320	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Portugal	0.000477	0.000477	0.000502	0.000505	0.000259	0.000259	0.000308	0.000310	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Slovakia	0.000739	0.000743	0.000972	0.000944	0.000394	0.000398	0.000560	0.000546	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Slovenia	0.000835	0.000828	0.000884	0.000970	0.000459	0.000454	0.000540	0.000587	ARIMA(0,1,1) with drift	ARIMA(0,1,1) with drift
Sweden	0.000410	0.000410	0.000372	0.000373	0.000231	0.000230	0.000226	0.000228	ARIMA(2,1,0) with drift	ARIMA(2,1,0) with drift

doi.org/10.32388/Z69TVG

Table 4. Mean forecast errors for female mortality in each country, along with their best ARIMA models.

Generally, the α-transformation results in lower forecast errors than the CLR transformation in many countries for both training and test sets. The large difference in RMSE and MAE between both transformations, specifically in East Germany, contributes to the slightly higher training error under α-transformation as depicted in Table 3. It is also worth noting that when $\alpha = 0$ as observed in Finland, Ireland, Poland, Portugal and England & Wales, which corresponds to the ILR transformation, the accuracy is similar to that of the CLR transformation. This is in accordance with Shang and Haberman's findings that both ILR and CLR perform similarly in terms of point forecast accuracy for Australian gender-specific $d_{t,x}$ within functional CoDA[25].

3.3.3. Mean errors over years

Figure 2 visualizes the trend of mean forecast errors for female mortality over years in the training and test sets. A tabulated version can be found in the supplementary materials. While the α-transformation gives a higher forecast errors at the beginning, it performs consistently better than the CLR transformation in the holdout set. In addition, it is worth highlighting that the extremely small errors up to 10^{-18} in the last fitting year, i.e. year 2010, is due to the jump-off adjustment. This occurs because the last observed $d_{t,x}$ are used as jump-off rates to avoid jump-off bias[37][38].

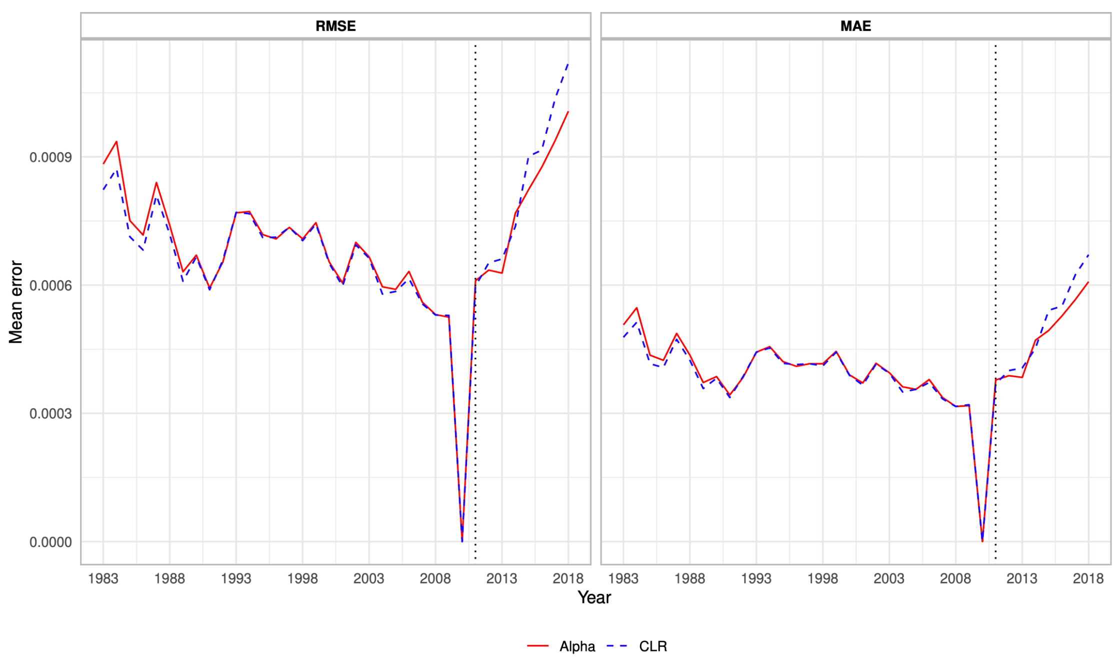


Figure 2. Mean forecast errors for female mortality over years across countries.

doi.org/10.32388/Z69TVG

When the widely used Lee-Carter model was first introduced, Lee and Carter[4] noted that using fitted values as jump-off rates would not perfectly match the data in the jump-off year, leading to a discontinuity between observed and forecast logs of mortality rates. Although they claimed that such discontinuity affects only rates that are absolutely low with little impact on life expectancy, they suggested that the jump-off error could be fixed by setting a_x equal to the most recently observed age-specific $\log m_{t,x}$ with a drawback that such approach might deteriorate the goodness of fit for the rest of the fitting period.

Later, Lee and Miller[37] found that the jump-off bias was actually apparent. This could likely be due to the shortcoming of using a one-principal component (PC), i.e. $K = 1$ approximation in the LC model[39]. Although forecasting of only a single time series is required following the use of the first PC, having subtracted the age-specific means α_x, this method comes along with more approximation error compared to taking more PCs. Hence, bias adjustment would be required.

Bell[39] and Lee and Miller[37] discovered that a jump-off adjustment using observed jump-off rates can eliminate the bias and achieve a more accurate forecast in forecasting period, as also proven in[38]. Using the similar approach, errors between $d_{t,x}$ and $\hat{d}_{t,x}$ are nearly zero since the fitted values are corrected to match the observed values exactly[37].

3.3.4. Mean errors by age

Figure 3 illustrates the mean forecast errors for female mortality by age across countries measured using RMSE and MAE. These results are further summarized in tabular form in the supplementary materials. Both the α-transformed and CLR-transformed data result in comparable forecast accuracy up to age 84, after which α-transformation begins to exhibit slightly higher errors. However, this situation is short-lived as the α-transformation shows a clear advantage over the CLR transformation at older ages, particularly between ages 91 and 100. Beyond this age interval, different types of transformation appear to have a minimal impact on model predictive performance. This result aligns with previous findings that the α-transformation can be a competitive alternative to the widely used CLR approach for converting compositional mortality data into real space in mortality forecasting.

doi.org/10.32388/Z69TVG

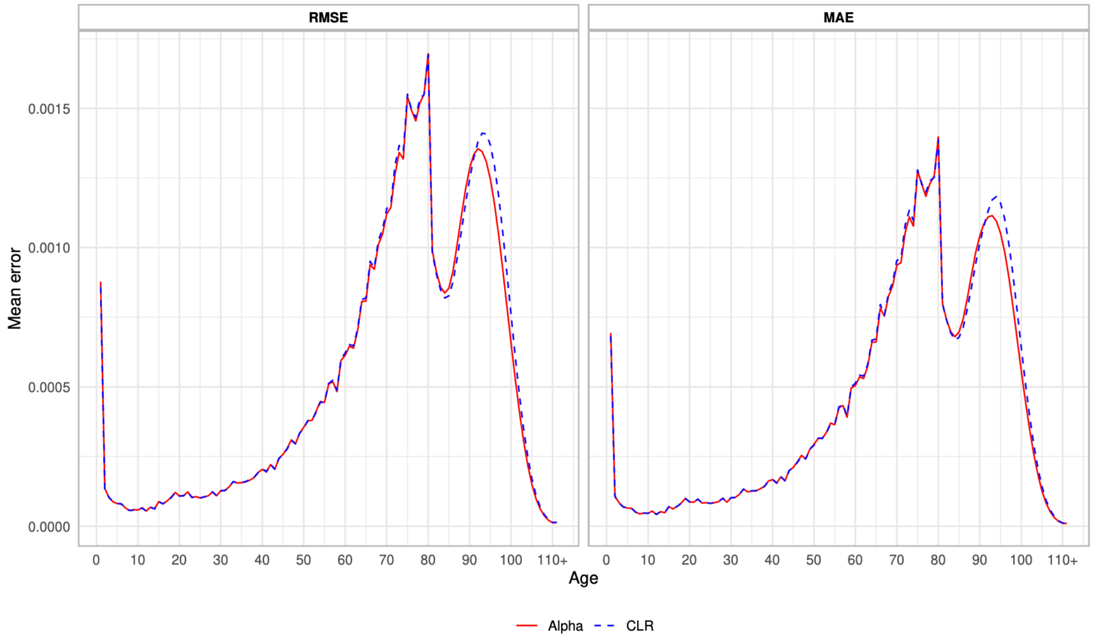


Figure 3. Mean forecast errors for female mortality by age across countries.

4. Conclusion

In this study, the potential of the α-transformation as an alternative to the commonly used CLR transformation in compositional mortality forecasting has been investigated. Using age-specific life table death counts for each gender across the selected countries/regions, a mortality modeling and forecasting procedure within a non-functional CoDA framework akin to the Lee-Carter model that involves α-transformation has been proposed. A comparative analysis is then carried out between the two transformations from a few perspectives, including overall mean errors, mean errors by country, mean errors over years and mean errors by age.

The results show that the model fitted to the α-transformed data has a comparable and superior performance in most of the selected countries as compared to the CLR-transformed data, especially at older ages. In particular, these findings are consistent with previous studies that involve the application of α-transformation within the functional CoDA framework[24][25]. The improved accuracy can be mainly attributed to the flexibility of the αparameter, which is fine-tuned for better data adaptation, as opposed to the rigid log-ratio transformation.

Future research may explore the application of α-transformation in coherent mortality forecasting to account for a common trend that affects the country-specific mortality pattern. In addition, investigating the predictive performance of weighted compositional forecasting models using α-transformed versus log-ratio transformed life table death counts may provide some valuable insights in this field.

Statements and Declarations

Conflicts of interest

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Universiti Malaya Research Excellence Grant (UMREG 2.0) 2024 [UMREG044-2024].

Data availability

All data used in this manuscript are available from the Human Mortality Database (www.mortality.org).

References

1. ^ *Booth H, Ticker L (2008). "Mortality modelling and forecasting: A review of methods". Annals of Actuarial Science.* **3** *(1-2): 3–43. doi:10.1017/S1748499500000440.*
2. a, b *Basellini U, Camarda CG, Booth H (2023). "Thirty years on: A review of the Lee-Carter method for forecasting mortality". International Journal of forecasting.* **39** *(3): 1033–1049. doi:10.1016/j.ijforecast.2022.11.002.*
3. ^ *Hyndman RJ, Ullah MS (2007). "Robust forecasting of mortality and fertility rates: A functional data approach". Comput. Statist. Data Anal..* **51** *(10): 4942–4956. doi:10.1016/j.csda.2006.07.028.*
4. a, b, c *Lee RD, Carter LR (1992). "Modeling and forecasting U. S. mortality". J. Amer. Statist. Assoc..* **87** *(419): 659–671. doi:10.2307/2290201.*
5. a, b *Li N, Lee R (2005). "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method". Demography.* **42** *(3): 575–594. doi:10.1353/dem.2005.0021.*
6. a, b, c, d, e, f *Oeppen J. "Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data". In: European Population Conference 2008, 9--12 July 2008, Barcelona, Spain. European Association for Population Studies (EAPS); 2008. p. 1--25. Available from: https://api.semanticscholar.org/CorpusID:1030770.*
7. ^ *Atance D, Claramunt MM, Varea X, Aburto JM (2024). "Convergence and divergence in mortality: A global study from 1990 to 2030". PLoS ONE.* **19** *(1). doi:10.1371/journal.pone.0295842. e0295842.*
8. ^ *Cairns AJG, Blake D, Dowd K (2006). "A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration". The Journal of Risk and Insurance.* **73** *(4): 687–718. doi:10.1111/j.1539-6975.2006.00195.x.*
9. ^ *King G, Soneji S (2011). "The future of death in America". Demographic Res..* **25***: 1–38. doi:10.4054/DemRes.2011.25.1.*
10. a, b, c, d, e, f, g, h, i, j, k, l, m, n *Bergeron-Boucher M-P, Canudas-Romo V, Oeppen J, Vaupel JW (2017). "Coherent forecasts of mortality with compositional data analysis". Demographic Res..* **37** *(17): 527–566. doi:10.4054/DemRes.2017.37.17.*
11. a, b *Kjærgaard S, Ergemen YE, Kallestrup-Lamb M, Oeppen J, Lindahl-Jacobsen R (2019). "Forecasting causes of death by using compositional data analysis: The case of cancer deaths". J. R. Stat. Soc. Ser. C. Appl. Stat..* **68** *(5): 1351–1370. doi:10.1111/rssc.12357.*

12. a, b, c, d, e *Shang HL, Haberman S (2024). "Weighted Compositional Functional Data Analysis for Modeling and Forecasting Life-table Death Counts". J. Forecast.. **43** (8): 3051–3071. doi:10.1002/for.3171.*

13. ^ *Jong PD, Marshall C (2007). "Mortality projection based on the Wang transform". Astin Bull.. **37** (1): 149–161. doi:10.2143/AST.37.1.2020803.*

14. ^ *Pascariu MD, Canudas-Romo V, Vaupel JW (2018). "The double-gap life expectancy forecasting model". Insurance: Mathematics and Economics. **78**: 339–350. doi:10.1016/j.insmatheco.2017.09.011.*

15. a, b *Bergeron-Boucher MP, Kjærgaard S, Oeppen J, Vaupel JW (2019). "The impact of the choice of life table statistics when forecasting mortality". Demographic Res.. **41**: 1235–1268. Available from: http://www.demographic-research.org/Volumes/Vol41/43/.*

16. ^ *Canudas-Romo V. Three measures of longevity: Time trends and record values. Demography. **47** (2): 299–312. doi:10.1353/dem.0.0098.*

17. ^ *Aburto JM, Villavicencio F, Basellini U, Kjærgaard S, Vaupel JW (2020). "Dynamics of life expectancy and life span equality". Proc. Natl. Acad. Sci. U.S.A.. **117** (10): 5250–5259. doi:10.1073/pnas.1915884117.*

18. a, b, c, d, e, f, g *Aitchison J. The Statistical Analysis of Compositional Data. London, UK: Chapman & Hall; 1986.*

19. a, b, c, d, e, f, g, h *Tsagris MT, Preston S, Wood ATA. "A data-based power transformation for compositional data". In: Egozcue JJ, Tolosana-Delgado R, Ortego MI, editors. 4th International Workshop on Compositional Data Analysis; 1-5 June 2011; Girona, Spain. Barcelona, Spain: International Center for Numerical Methods in Engineering (CIMNE); 2011. p. 1-8. Available from: http://hdl.handle.net/2117/366618.*

20. a, b, c, d, e, f, g *Clarotto L, Allard D, Menafoglio A (2022). "A new class of α-transformations for the spatial analysis of compositional data". Spat. Stat.. **47**: 100570. doi:10.1016/j.spasta.2021.100570.*

21. a, b, c, d *Tsagris M. "Regression analysis with compositional data containing zero values". Chil. J. Stat.. **6** (2): 47–57. https://www.soche.cl/chjs/volumes/06/02/Tsagris(2015).pdf.*

22. ^ *Tsagris M, Alenazi A, Stewart C (2023). "Flexible non-parametric regression models for compositional response data with zeros". Stat. Comput.. **33** (5): 106. doi:10.1007/s11222-023-10277-5.*

23. ^ *Tsagris M, Preston S, Wood ATA (2016). "Improved classification for compositional data using the α-transformation". J. Classification. **33**: 243–261. doi:10.1007/s00357-016-9207-5.*

24. a, b, c, d, e *Giacomello S. Functional α-transformations for the statistical analysis of density functions [Master's thesis]. Politecnico di Milano; 2024. Available from: https://hdl.handle.net/10589/222832.*

25. a, b, c, d, e, f, g *Shang HL, Haberman S (2024). "Forecasting age distribution of life-table death counts via \u03b1-transformation". Scand. Actuar. J.. pp. 1–17. doi:10.1080/03461238.2024.2425723.*

26. ^ *HMD. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France), 2024; dataset available at www.mortality.org.*

27. a, b, c, d, e *Wilmoth JR, Andreev K, Jdanov D, Glei DA, Riffe T, Boe C, Bubenheim M, Philipov D, Shkolnikov V, Vachon P, Winant C, Barbieri M. Methods protocol for the Human Mortality Database. 6th ed. 2021. Available from: https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf.*

28. ^Thatcher AR, Kannisto V, Vaupel JW. The Force of Mortality at Ages 80 to 120. Odense: Odense University Press; 1998. Vol. 5. Available from: https://www.demogr.mpg.de/papers/books/monograph5/start.htm.
29. ^Aitchison J. Principles of compositional data analysis. In: Multivariate Analysis and Its Applications. IMS Lecture Notes Monograph Series Vol. 24. Hayward, CA: Institute of Mathematical Statistics; 1994. p. 73-81. doi:10.1214/lnms/1215463786. MR 1479457.
30. ^Andreev EM, Kingkade WW (2015). "Average age at death in infancy and infant mortality level: Reconsidering the Coale-Demeny formulas at current levels of low mortality". Demographic Res.. 33 (13): 363–390. doi:10.4054/DemRes.2015.33.13.
31. ^Pascariu MD. MortalityLaws: Parametric mortality models, life tables and HMD. R package version 2.1.0, 2024; software available at https://doi.org/10.32614/CRAN.package.MortalityLaws.
32. a, bHyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne, Australia: OTexts; 2021. Available from: https://otexts.com/fpp3/.
33. ^Hyndman RJ, Khandakar Y (2008). "Automatic time series forecasting: The forecast package for R". Journal of Statistical Software. 27 (3): 1–22. doi:10.18637/jss.v027.i03.
34. ^R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2024; software available at https://www.R-project.org/.
35. a, bRobertson HT, Allison DB (2012). "A novel generalized normal distribution for human longevity and other negatively skewed data". PLoS One. 7 (5): e37025. doi:10.1371/journal.pone.0037025.
36. ^Abouzahr C, Mikkelsen L, Rampatige R, Lopez A. "Mortality statistics: A tool to enchance understanding and improve quality". Pac Health Dialog. 18 (1): 247–270 (2012). PMID 23240364.
37. a, b, c, dLee R, Miller T (2001). "Evaluating the performance of the Lee-Carter method for forecasting mortality". Demography. 38 (4): 537–549. doi:10.1353/dem.2001.0036.
38. a, bStoeldraijer L, van Duin C, van Wissen L, Janssen F (2018). "Comparing strategies for matching mortality forecasts to the most recently observed data: Exploring the trade-off between accuracy and robustness". Genus. 74. doi:10.1186/s41118-018-0041-y. 16.
39. a, bBell WR (1997). "Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates". J. Off. Stat.. 13 (3): 279--303.

Declarations

Funding: This work was supported by the Universiti Malaya Research Excellence Grant (UMREG 2.0) 2024 [UMREG044-2024].

Potential competing interests: No potential competing interests to declare.