Weighted Compositional Data Analysis for Modeling and Forecasting Life-Table Death Counts

NOVEMBER| 2023

Mortality and Longevity

Aging and Retirement

LIVING TO 100

SOA INTERNATIONAL SYMPOSIUM

SOA Research INSTITUTE

Weighted Compositional Data Analysis for Modeling and Forecasting Life-Table Death Counts

AUTHOR
Han Lin Shang, Ph.D.
Professor of Business Analytics and Director of Research
Department of Actuarial Studies and Business Analytics
Macquarie University

Steven Haberman, Ph.D.
Bayes Business School
City, University of London

SPONSORS
Mortality and Longevity Strategic Research Program Steering Committee

Aging and Retirement Strategic Research Program Steering Committee

Weighted compositional data analysis for modeling and forecasting life-table death counts

Han Lin Shang [*]
Department of Actuarial Studies and Business Analytics
Macquarie University

Steven Haberman
Bayes Business School
City, University of London

February 8, 2023

Abstract

Age-specific life-table death counts observed over time are examples of densities. Non-negativity and summability are two constraints that prevent the direct implementation of standard linear statistical methods. Compositional data analysis presents a one-to-one mapping from constrained to unconstrained space to rectify the constraints. We introduce a weighted compositional data analysis for modeling and forecasting life-table death counts. Our extension assigns higher weights to more recent data and provides a modeling scheme that is easily adapted to allow for constraints. We illustrate our method using age-specific Swedish life-table death counts from 1751 to 2020 and show that the weighted compositional data analytic method improves short-term forecast accuracy compared to their unweighted counterparts.

Keywords: age distribution of death counts; geometrically decaying weights; centered log-ratio transformation; weighted principal component analysis

[*]Postal address: Department of Actuarial Studies and Business Analytics, Level 7, 4 Eastern Road, Macquarie University, Sydney, NSW 2109, Australia; Telephone: +61(2) 9850 4689; Email: hanlin.shang@mq.edu.au.

1 Introduction

Actuaries and demographers have long been interested in developing mortality modeling and forecasting methods. In the literature on human mortality, three functions are widely considered: hazard, survival, and probability density functions. Although these functions are complementary (Preston et al. 2001, Dickson et al. 2009), most attention was given to new approaches for forecasting age-specific hazard function (see, e.g., Booth 2006, Booth & Tickle 2008, for reviews). Instead of modeling central mortality rates, we consider modeling the life-table death distribution (Basellini et al. 2020). Observed over a period, we could model and forecast a redistribution of the density of life-table death counts, where deaths at younger ages are shifted gradually towards older ages. The period life-table death counts represent the mortality conditions, which reflect a longevity trend in recent years. Apart from providing an informative description of the mortality experience of a population, the life-table death counts yield readily available information on 'central longevity indicators' (see, e.g., Cheung et al. 2005, Canudas-Romo 2010), and lifespan variability (see, e.g., Robine 2001, Vaupel et al. 2011, Horiuchi et al. 2013, van Raalte & Caswell 2013, van Raalte et al. 2014, Aburto & van Raalte 2018, Aburto et al. 2020).

In demography, Oeppen (2008), Bergeron-Boucher et al. (2017) and Bergeron-Boucher et al. (2018) treat life-table death counts as compositional data and use compositional data analysis (CoDa) to model and forecast age distribution of death counts. The data are constrained to vary between two limits (e.g., 0 and a constant upper bound), which in turn imposes constraints upon the variance-covariance structure of the original data. To remove the non-negativity and summability constraints, a transformation can be deployed before applying standard linear techniques to the transformed data. Arguably, the most popular transformation is the centered log-ratio transformation (Aitchison & Shen 1980, Aitchison 1982, 1986). Shang & Haberman (2020) applied the centered log-ratio transformation within the CoDa framework to model and forecast life-table death counts.

We extend the CoDa by assigning a set of geometrically decaying weights to estimate the geometric mean function and estimated principal components. As described in Section 3, our extension assigns higher weights to relatively more recent data to improve short-term forecast accuracy (see also Hyndman & Shang 2009). This is particularly important in demography, where we can have over 200 years of data, and data from the 18^{th} and 19^{th} centuries may not be so helpful in determining the recent trend for forecasting.

Using Swedish life-table death counts from 1751 to 2020 in Section 2, we highlight the difference between the weighted and standard compositional data analytic methods in Sections 3 and 4 and evaluate and compare forecast accuracy in Section 5. The conclusion is presented in Section 6, along with some ideas on how the methodology can be further extended.

2 Swedish age distribution of death counts

We consider age- and sex-specific life-table death counts from 1751 to 2020 in Sweden obtained from the Human Mortality Database (2021). We study life-table death counts, where the life table radix (i.e., a population experiencing 100,000 births annually) is fixed at 100,000 at age 0 for each year. For the life-table death counts, there are 111 ages, and these are ages $0, 1, \dots, 109, 110+$. Due to rounding, there are zero counts for age 110+ at some years, which may create an issue when taking log-ratio transformation. To rectify this problem, we prefer to use the probability of dying and the life-table radix to recalculate our estimated death counts (up to 6 decimal places). In doing so, we obtain more precise death counts than the ones reported in the Human Mortality Database (2021). To some extent, the probability of dying relies on smooth rates (see the Human Mortality Database 2021, protocol for detail).

To understand the features of the data, Figure 1 presents rainbow plots of the female and male age-specific period life-table death counts in Sweden from 1751 to 2020 in a single year.

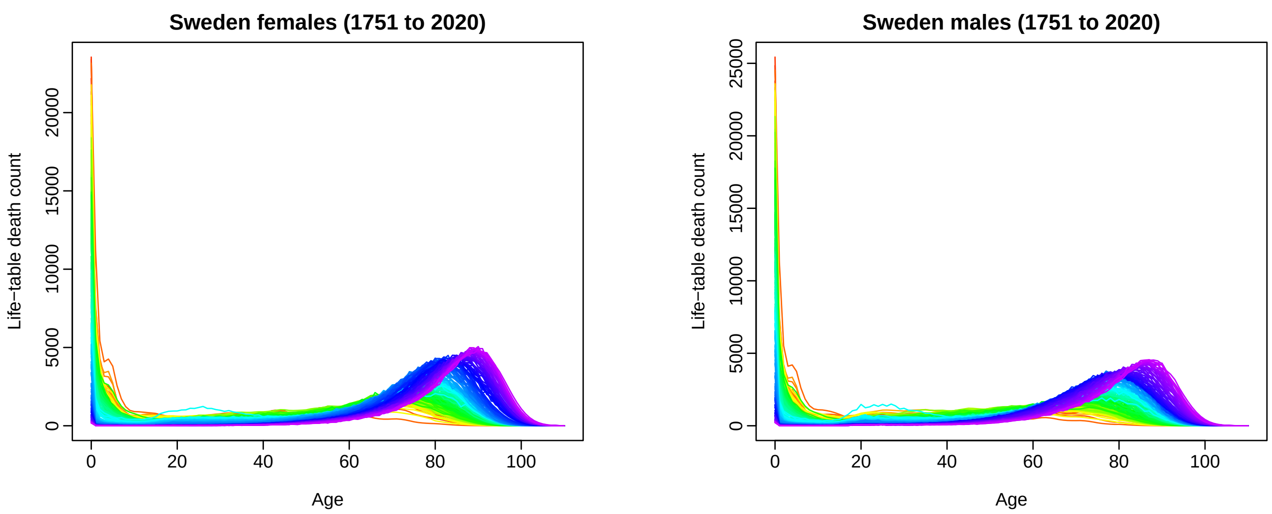


Figure 1 Rainbow plots of age-specific period life-table death count from 1751 to 2020 in a single-year group in Sweden. Curves are ordered chronologically according to the colors of the rainbow. The oldest years are displayed in red, with the most recent years shown in violet.

Both sub-figures demonstrate a decreasing trend in infant death counts, and a typical negatively skewed distribution for the life-table death counts, where the peaks shift to higher ages for both females and males. This shift is a primary driver of the longevity risk, which is a major issue for insurers and pension funds, especially in the selling and risk management of annuity products (see Denuit et al. 2007, for a discussion).

The re-distribution of life-table death counts indicates lifespan variability across age. A decrease in variability over time can be observed. This variability can be measured, for example, with the interquartile range of life-table death counts, Drewnowski's index, or the Gini coefficient (for comprehensive reviews, see Wilmoth & Horiuchi 1999, Shkolnikov et al. 2003, van Raalte & Caswell 2013, Debón et al. 2017, Aburto et al. 2022). Via the Gini coefficient, Figure 2 presents an example where the life-table death counts provide essential insights on longevity and lifespan variability that cannot be grasped directly from an examination of the central mortality rate or the survival function. We also include graphs of the trend in life expectancy at age zero, denoted by e(0), over time.

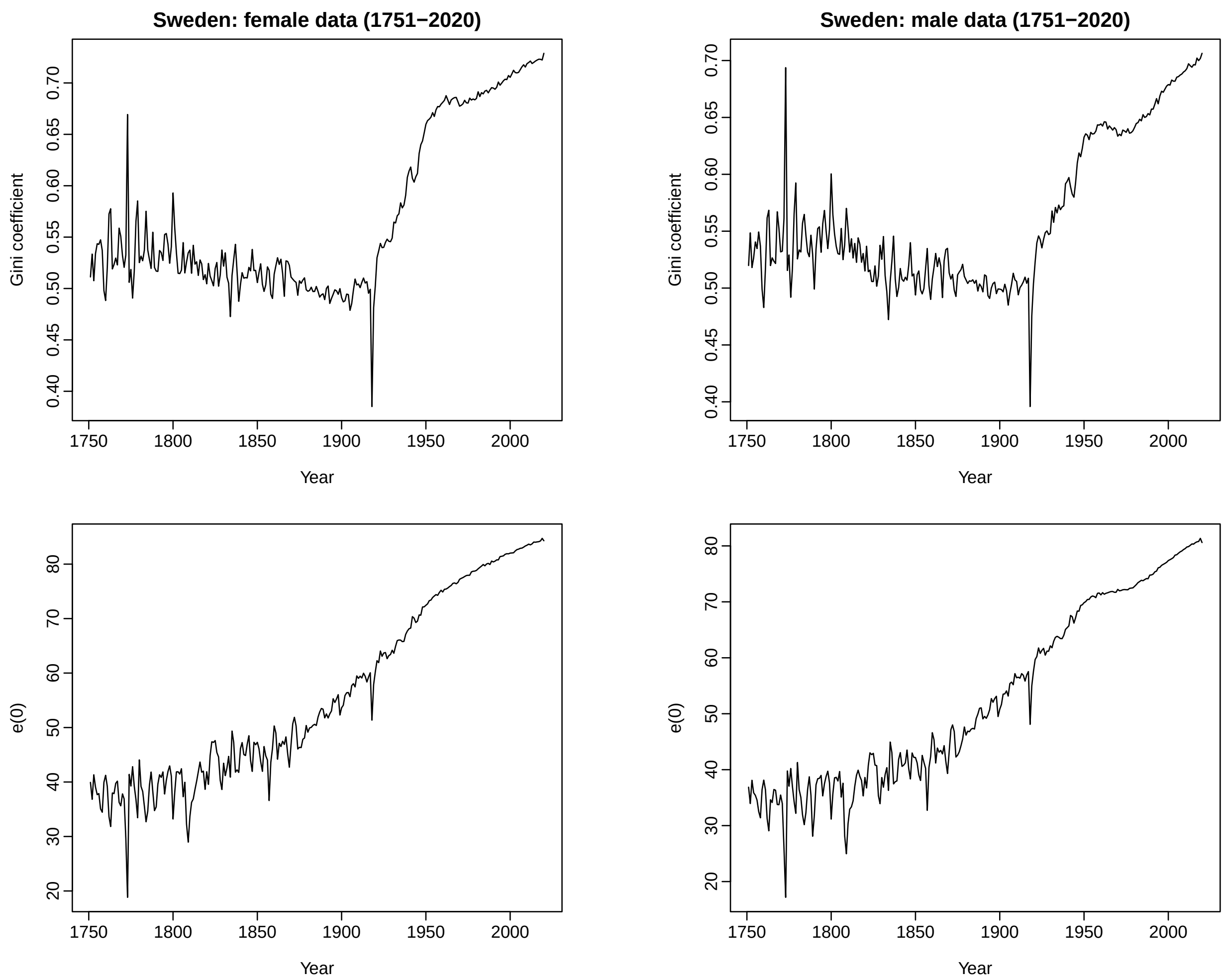


Figure 2 Gini coefficients and e(0) for Swedish period female and male life-table death counts from 1751 to 2020.

The Gini coefficient can be viewed as a measure of age-at-death inequality. When the coefficient equals 0, it expresses maximal age-at-death inequality. Inversely, when the coefficient equals 1, equality occurs for all ages. From Figure 2, the effects of the cholera epidemic in 1834 are apparent for the Swedish female and male data (Larsson 2020). In 1918, there was a sudden drop in the Gini coefficient related to the Spanish flu.

3 Geometrically weighted compositional data analytic approach

Density functions are non-negative functions that integrate into one. They share features with compositional data (see, e.g., Aitchison 1986, Pawlowsky-Glahn et al. 2015). Compositional data arise in many scientific fields, such as geology (geochemical elements), economics (income/expenditure distribution), medicine (body composition), the food industry (food composition), chemistry (chemical composition), agriculture (nutrient balance bionomics), environmental science (soil contamination), ecology (abundance of different species) and demography (life-table death counts).

In statistics, Scealy et al. (2017) use CoDa to study the concentration of chemical elements in sediment. Scealy & Welsh (2017) apply CoDa to analyze total weekly expenditure on food and housing costs for households. Delicado (2011), Kokoszka et al. (2019) and Shang & Haberman (2020) use CoDa to analyze density functions and implement principal component analysis on the unconstrained space. In demography, Oeppen (2008) and Bergeron-Boucher et al. (2017) introduce a principal component analysis approach to forecast life-table death counts within a CoDa framework.

For a given year t, compositional data are defined as a random vector of I non-negative components, $[\mathcal{X}_t(u_1), \dots, \mathcal{X}_t(u_I)]$, whose sum is a specified constant, such as one (portion), 100 (percentage), 10^5 (radix) in life-table death counts, and 10^6 (parts per million) in geochemical trace element compositions (Aitchison 1986, p.1). Between the non-negativity and summability constraints, the sample space of compositional data is a simplex

$$\mathcal{S}^I = \left\{ [\mathcal{X}_t(u_1), \dots, \mathcal{X}_t(u_I)]^\top, \quad \mathcal{X}_t(u_i) > 0, \quad \sum_{i=1}^{I} \mathcal{X}_t(u_i) = c \right\}, \qquad t = 1, \dots, n,$$

where u denotes a continuum, such as age, $\mathcal{S}$ denotes a simplex, c is a fixed constant, $^\top$ denotes vector transpose, and the simplex sample space is a $I-1$ dimensional subset of real-valued space

R^I. The restriction of shares to the unit simplex leads to the lack of an interpretable covariance structure, which has been recognized by researchers in many fields (see, e.g., Aitchison 1986, Barceló et al. 1996, Fry et al. 1996).

In the CoDa framework, intuition involves rectifying the summability and non-negativity constraints via a transformation of the raw data. The transformation preserves one-to-one mapping between constrained and unconstrained spaces. In the unconstrained space, standard statistical techniques can be directly applied. Among many suitable transformations, the centered log-ratio transformation is widely used (Aitchison & Shen 1980, Aitchison 1982, 1986). The algorithm for implementing the CoDa method consists of the following steps:

1) *De-centering the data.* Compute the geometric mean function with geometrically decaying weights. The mean function $\mu(u)$ can be estimated by a weighted average

$$\alpha_n(u) = \exp\left\{\sum_{t=1}^{n} w_t \ln[\mathcal{X}_t(u)]\right\},$$

where $w_t = \kappa(1-\kappa)^{n-t}$ is a set of geometrically decaying weights with $0 < \kappa < 1$ and $\sum_{t=1}^{n} w_t = 1$, and $\ln(\cdot)$ denotes natural logarithm. These weights are to improve the estimation of the mean function.

In Figure 3, we display a set of geometrically decaying weights when the weight parameter $\kappa = 0.05$ or 0.95, respectively. When $\kappa = 0.05$, forecasts depend on more distant-past observations. When $\kappa = 0.95$, forecasts rely on the most recent observations.

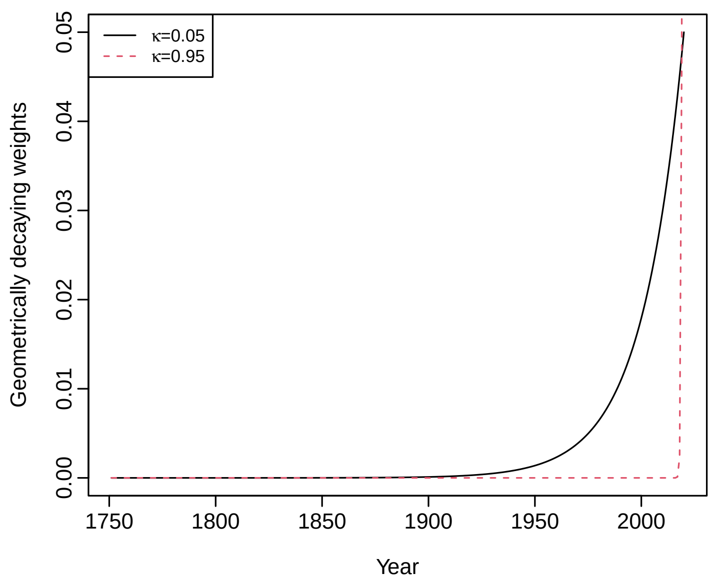


Figure 3 Geometrically decaying weights when $\kappa = 0.05$ and 0.95, respectively.

We treat age as a continuum $u \in [0, 110]$ although age is observed at discrete points, and set

$$s_t(u) = \frac{\mathcal{X}_t(u)}{\alpha_n(u)}.$$

2) Apply the centered log-ratio transformation given by

$$\beta_t(u) = \ln[s_t(u)]. \tag{1}$$

The log transformation in (1) removes the constraints on $\mathcal{X}_t(u)$. For a given population, $\boldsymbol{\beta}(u) = [\beta_1(u), \dots, \beta_n(u)]$ can be viewed as an unconstrained functional time series. To study its patterns, we estimate its variance and consider a variance-decomposition approach.

3) *Implement a weighted functional principal component analysis.* Functional principal component analysis has been extensively studied as a powerful dimension reduction tool to summarize infinite-dimensional functional objects by a few sets of orthonormal functional principal components and their associated scores. Survey articles on functional principal component analysis are presented in Shang (2014) and Wang et al. (2016).

We extend the functional principal component analysis by incorporating geometrically decaying weights (see, e.g., Hyndman & Shang 2009, Lam & Wang 2022). Compute eigenvalues and eigenfunctions of the weighted unconstrained functional time series: Let $\boldsymbol{w} = \text{diagonal}(w_1, \dots, w_n)$. The weighted curve time series is given as $\boldsymbol{\beta}^*(u) = \boldsymbol{w}\boldsymbol{\beta}(u)$. These weights are to improve the estimations of the functional principal components and their associated scores.

Applying eigen-decomposition to $\boldsymbol{\beta}^*(u)$ gives

$$\beta_t^*(u) = \sum_{k=1}^{n} \widehat{\gamma}_{t,k}\widehat{\phi}_k(u) = \sum_{k=1}^{K} \widehat{\gamma}_{t,k}\widehat{\phi}_k(u) + \omega_t(u),$$

where $\widehat{\phi}_k(u)$ is the k^{th} orthonormal eigen-function (i.e., functional principal components), $\widehat{\gamma}_{t,k} = \langle \beta_t(u), \widehat{\phi}_k(u) \rangle$ is the k^{th} principal component score at time t, and $\omega_t(u)$ denotes model residual function for age u in year t.

We determine K via an eigenvalue ratio criterion (see Li et al. 2020). The estimated value of K is chosen as the integer minimizing ratios of two adjacent empirical eigenvalues given by

$$K = \underset{1\leq k\leq k_{\max}}{\operatorname{argmin}} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left(\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right) + \mathbb{1}\left(\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right) \right\},$$

where $k_{\max}$ is a pre-specified positive integer, $\theta = 1/\ln[\max(\widehat{\lambda}_1, n)]$ is a pre-specified small positive number penalizing those relatively smaller empirical eigenvalues, and $\mathbb{1}(\cdot)$ is the binary indicator function. Without any prior information, we set $k_{\max} = \#\{k|\widehat{\lambda}_k \geq \sum_{k=1}^{n}\widehat{\lambda}_k/n, k \geq 1\}$ (Ahn & Horenstein 2013).

4) *Transform back to the compositional data.* We take the inverse centered log-ratio transformation given by

$$\widehat{\boldsymbol{s}}_{n+h|n}(u) = \left[\frac{\exp^{\widehat{\beta}_{n+h|n}(u_1)}}{\sum_{i=1}^{111}\exp^{\widehat{\beta}_{n+h|n}(u_i)}}, \frac{\exp^{\widehat{\beta}_{n+h|n}(u_2)}}{\sum_{i=1}^{111}\exp^{\widehat{\beta}_{n+h|n}(u_i)}}, \cdots, \frac{\exp^{\widehat{\beta}_{n+h|n}(u_{111})}}{\sum_{i=1}^{111}\exp^{\widehat{\beta}_{n+h|n}(u_i)}}\right],$$

where $\widehat{\beta}_{n+h|n}(u_i)$ denotes the forecasts in Step 3).

5) We add back the geometric means to obtain the forecasts of the life-table death counts $\boldsymbol{\mathcal{X}}_{n+h}(u)$,

$$\widehat{\boldsymbol{\mathcal{X}}}_{n+h|n}(u) = \left[\frac{\widehat{s}_{n+h|n}(u_1)\times\alpha_1}{\sum_{i=1}^{111}\widehat{s}_{n+h|n}(u_i)\times\alpha_i}, \frac{\widehat{s}_{n+h|n}(u_2)\times\alpha_2}{\sum_{i=1}^{111}\widehat{s}_{n+h|n}(u_i)\times\alpha_i}, \cdots, \frac{\widehat{s}_{n+h|n}(u_{111})\times\alpha_{111}}{\sum_{i=1}^{111}\widehat{s}_{n+h|n}(u_i)\times\alpha_i}\right]$$

where α_i is the weighted geometric mean given in Step 1).

4 Selection of the geometrically decaying weight parameter

4.1 Point forecast error criteria

Since the age-specific life-table death counts can be considered a probability density function, we measure goodness-of-fit through several density evaluation metrics. These metrics can evaluate if our model accurately forecasts holdout data in the form of a probability density function. The metrics considered include discrete Kullback-Leibler divergence (Kullback & Leibler 1951) and the square root of the Jensen-Shannon divergence (Shannon 1948, Fuglede & Topsoe 2004).

The Kullback-Leibler divergence is intended to measure the information loss by replacing unknown density with approximation. For two probability density functions, denoted by

$\mathcal{X}_{n+\xi}(u)$ and $\widehat{\mathcal{X}}_{n+\xi|n}(u)$, the discrete Kullback-Leibler divergence for each horizon h is defined as

$$
\begin{aligned}
\text{KLD}(h) =& D_{\text{KL}}[\mathcal{X}_{n+\xi}(u_i)||\widehat{\mathcal{X}}_{n+\xi|n}(u_i)] + D_{\text{KL}}[\widehat{\mathcal{X}}_{n+\xi|n}(u_i)||\mathcal{X}_{n+\xi}(u_i)] \\
=& \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{i=1}^{111}\mathcal{X}_{n+\xi}(u_i)\cdot[\ln\mathcal{X}_{n+\xi}(u_i) - \ln\widehat{\mathcal{X}}_{n+\xi|n}(u_i)]+ \\
& \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{i=1}^{111}\widehat{\mathcal{X}}_{n+\xi|n}(u_i)\cdot[\ln\widehat{\mathcal{X}}_{n+\xi|n}(u_i) - \ln\mathcal{X}_{n+\xi}(u_i)],
\end{aligned}
$$

which is symmetric and non-negative. The forecast horizon can take integer values from one to ten. When $h = 1$, there are 10 years of data in the validation or testing samples. When $h = 10$, there is only one year of data.

An alternative is given by the Jensen-Shannon divergence, defined by

$$
\text{JSD}(h) = \frac{1}{2}\text{D}_{\text{KL}}[\mathcal{X}_{n+\xi}(u_i)||\delta_{n+\xi}(u_i)] + \frac{1}{2}\text{D}_{\text{KL}}[\widehat{\mathcal{X}}_{n+\xi|n}(u_i)||\delta_{n+\xi}(u_i)],
$$

where $\delta_{n+\xi}(u_i)$ measures a common quantity between $\mathcal{X}_{n+\xi}(u_i)$ and $\widehat{\mathcal{X}}_{n+\xi|n}(u_i)$. We consider simple or geometric mean

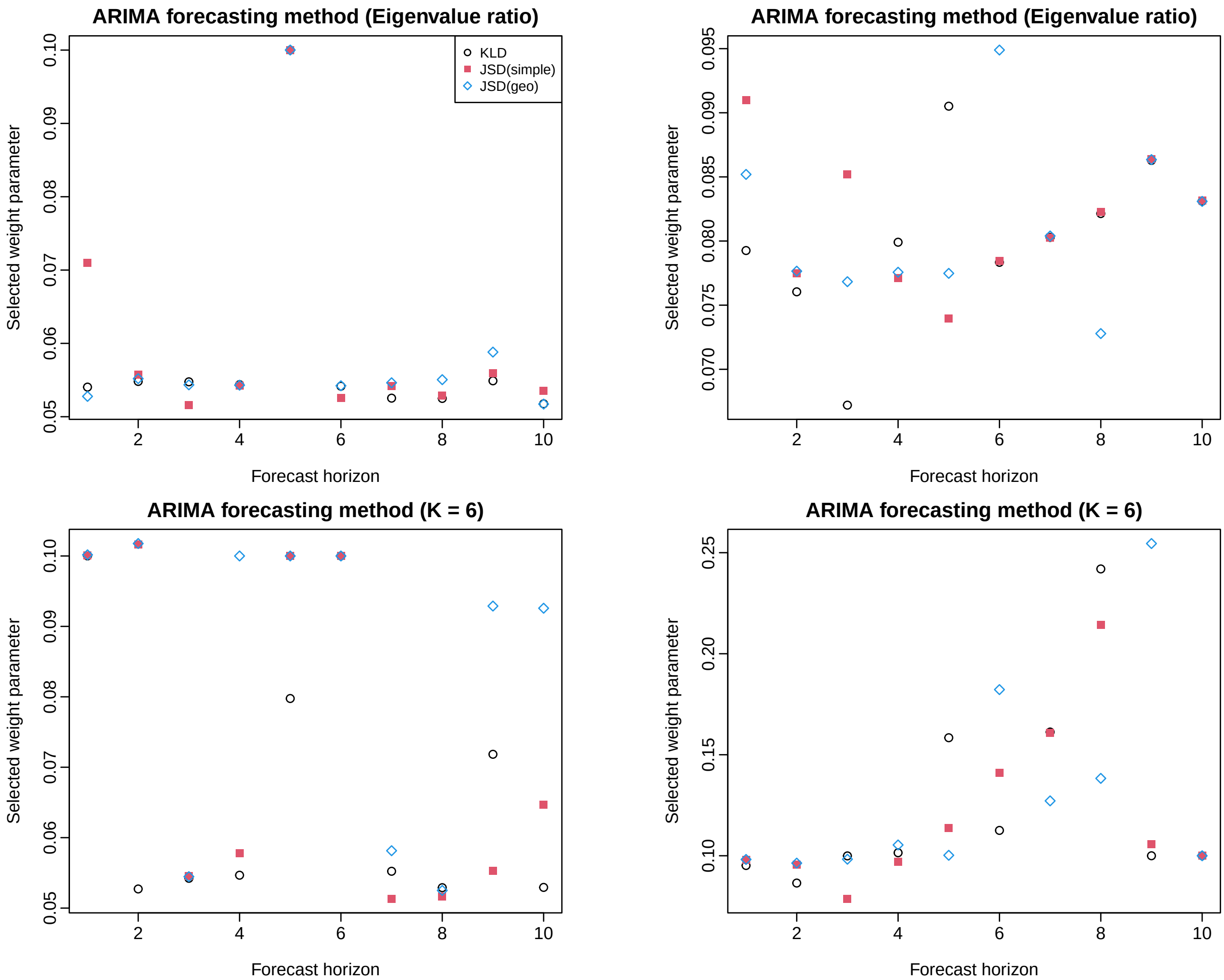
$$
\begin{aligned}
\delta_{n+\xi}(u_i) &= \frac{1}{2}[\mathcal{X}_{n+\xi}(u_i) + \widehat{\mathcal{X}}_{n+\xi|n}(u_i)] \\
\delta_{n+\xi}(u_i) &= \sqrt{\mathcal{X}_{n+\xi}(u_i)\widehat{\mathcal{X}}_{n+\xi|n}(u_i)}.
\end{aligned}
$$

We denote $\text{JSD}^s(h)$ for the Jensen-Shannon divergence with the simple mean, and $\text{JSD}^g(h)$ for the Jensen-Shannon divergence with the geometric mean.

4.2 Expanding window scheme

An expanding window scheme of a time series model is commonly used to assess model and parameter stability over time and the reliability of predictions. The expanding window analysis determines the variability of the model's parameters by computing parameter estimates and their resultant forecasts over an expanding window. We divide the entire data into a training sample consisting of data from 1751 to 2000, a validation sample consisting of data from 2001 to 2010, and a testing sample consisting of data from 2011 to 2020. The selections of validation and testing samples are arbitrary, each has 10 years of data.

Using the first 250 observations from 1751 to 2000 in the life-table death counts, we produce one- to 10-step-ahead forecasts. Through an expanding window approach, we re-estimate the parameters using the first 251 observations from 1751 to 2001. Forecasts from the estimated models are then produced for one- to nine-step-ahead forecasts. We iterated this process by increasing the sample size by one year until we reached the data in 2010. This process produces 10 one-step-ahead forecasts, 9 two-step-ahead forecasts, . . . , and one 10-step-ahead forecast. We evaluate and compare these forecasts with the validation samples from 2001 to 2010 to determine the optimal weight parameter for each forecast horizon. In Figure 4, we tabulate the estimated geometrically decaying weight parameter in the weighted CoDa method under various evaluation metrics.



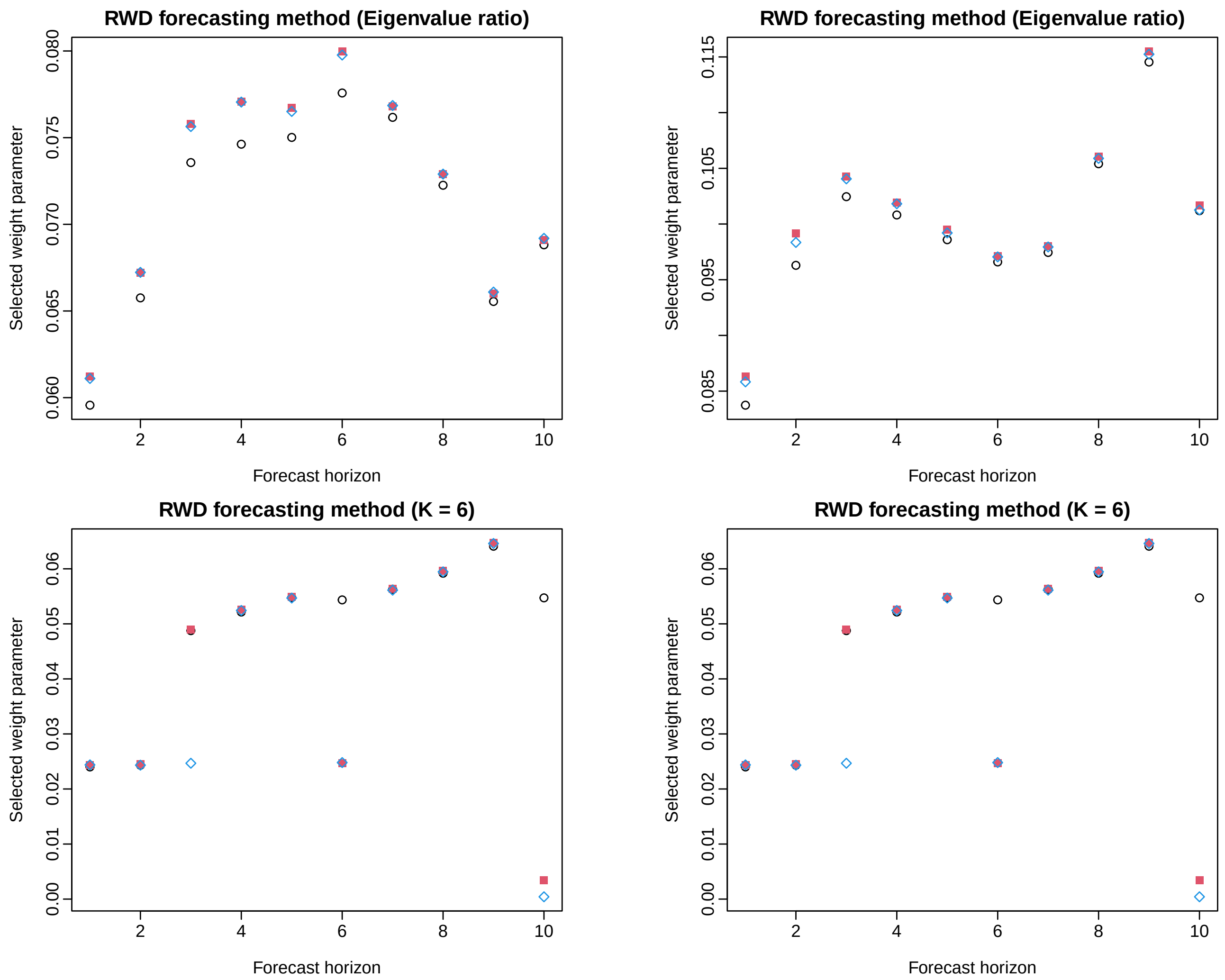


Figure 4 Estimated geometrically decaying weight parameter in the weighted CoDa method under various evaluation metrics.

4.3 Model fitting

We apply the standard and weighted CoDa methods to the female and male data. Via the eigenvalue ratio criterion, the retained numbers of components are determined as one. From the observed life-table death counts from 1751 to 2019 (i.e., 278 observations), we present the weighted geometric means of the female and male life-table death counts in Figures 5a and 5b. Via functional principal component analysis, we display the first functional principal component in Figures 5c and 5d. The estimated functions for the standard CoDa method are shown in red, while those for the weighted CoDa method are displayed in blue. For producing one-step-ahead forecasts, the weight parameter $\kappa = 0.054$ and 0.079 based on the KLD. From the one-step-ahead point forecasts of life-table death counts in 2020, the weighted CoDa method produces one-step-ahead forecasts that are closer to the holdout data in Figures 5e and 5f.

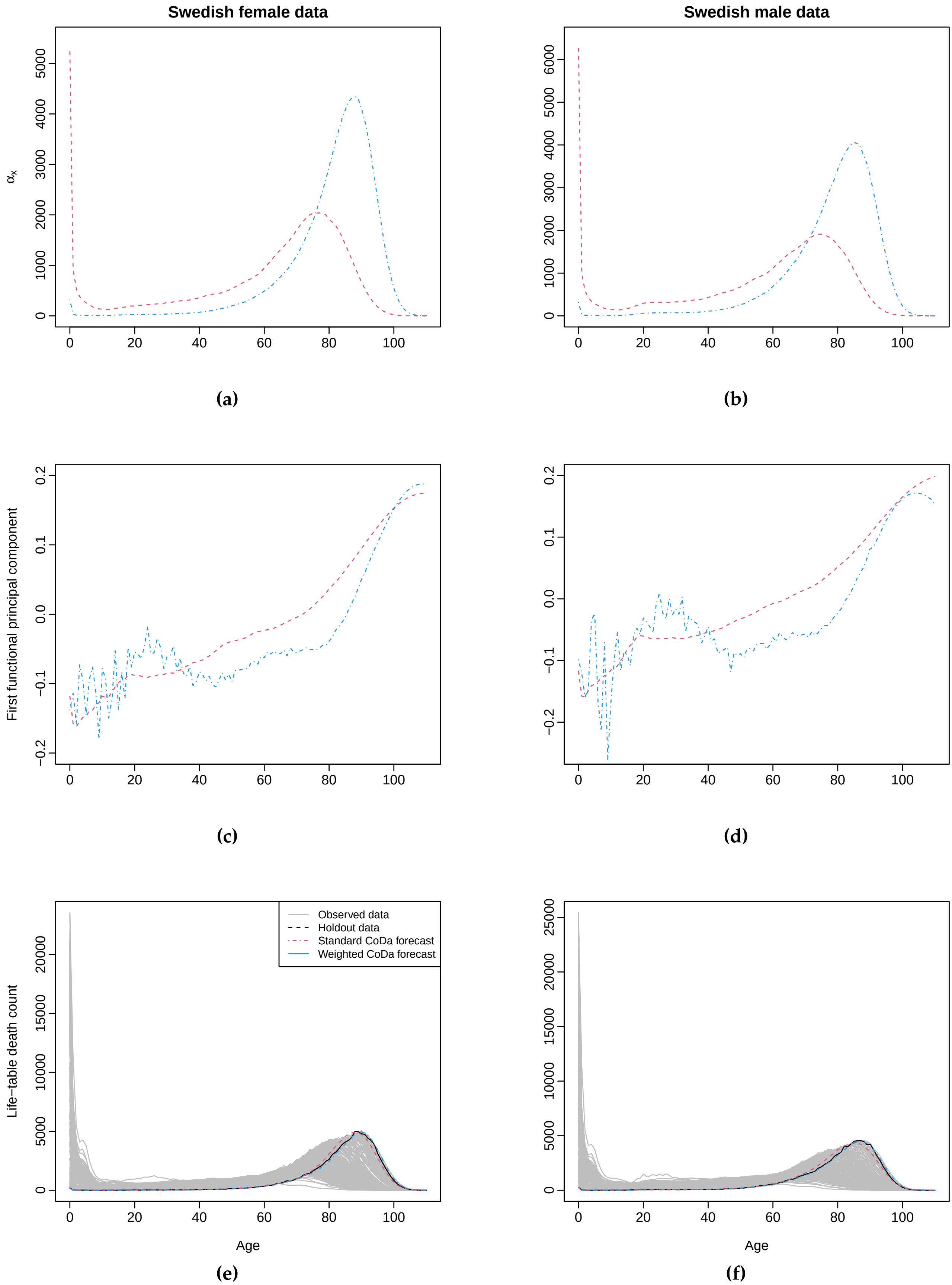


Figure 5 Elements of the CoDa methods for modeling and forecasting the Swedish female and male life-table death counts. The estimated functions for the standard CoDa method are shown in red, while the estimated functions for the weighted CoDa method are displayed in blue. The weight parameter values are selected as 0.054 and 0.079 for the female and male data, respectively.

5 Comparison of point forecast accuracy

Using the first 260 observations from 1751 to 2010, we produce one- to 10-step-ahead forecasts via an expanding window approach. We evaluate and compare forecast accuracy by comparing the forecasts with the holdout data from 2011 to 2020.

In Tables 1 and 2, we present three evaluation metrics between standard and geometrically decaying weighted CoDa methods. For the weighted CoDa methods, the estimated weight parameters for different horizons are displayed in Figure 4. We consider the autoregressive integrated moving average (ARIMA) and random-walk with drift (RWD) methods for forecasting principal component scores. We present two criteria for selecting the number of retained principal components. The first criterion uses the eigenvalue ratio, while the second one chooses $K = 6$ (see also Hyndman et al. 2013). For forecasting the Swedish female life-table death counts, we observe

1) the weighted CoDa method produces more accurate forecasts than the ones from the standard CoDa method;

2) fixing $K = 6$ reduces the point forecast error for the standard CoDa method, but not so for our weighted CoDa method; and

3) the weighted CoDa works well with both ARIMA and RWD methods.

Table 1 Comparison of point forecast errors ($\times 100$) between the CoDa and its weighted variant for forecasting the age-specific Swedish female data. For selecting the number of retained components K, we consider the eigenvalue ratio criterion and $K = 6$. For the ARIMA method, the optimal parameters are selected automatically based on the corrected Akaike information criterion (AICc).

	Eigenvalue ratio criterion						$K = 6$					
	CoDa			Weighted CoDa			CoDa			Weighted CoDa		
h	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g
<u>ARIMA</u>												
1	1.287	0.316	0.318	0.162	0.037	0.037	0.272	0.065	0.065	0.180	0.041	0.042
2	1.358	0.334	0.335	0.161	0.037	0.037	0.326	0.078	0.079	0.162	0.043	0.043
3	1.458	0.359	0.360	0.170	0.039	0.039	0.397	0.096	0.097	0.167	0.038	0.039
4	1.537	0.378	0.380	0.183	0.042	0.043	0.456	0.111	0.111	0.184	0.042	0.049
5	1.549	0.381	0.383	0.308	0.072	0.073	0.532	0.129	0.130	0.205	0.053	0.054
6	1.609	0.396	0.398	0.211	0.050	0.050	0.593	0.145	0.146	0.251	0.059	0.060
7	1.672	0.412	0.414	0.225	0.053	0.054	0.653	0.160	0.161	0.226	0.053	0.054
8	1.801	0.443	0.446	0.251	0.060	0.059	0.803	0.197	0.198	0.246	0.059	0.059
9	1.876	0.462	0.465	0.309	0.074	0.074	0.871	0.214	0.215	0.319	0.074	0.078

Continued on next page

h	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g
10	1.184	0.291	0.292	0.218	0.051	0.051	0.529	0.129	0.130	0.223	0.049	0.056
Mean	1.533	0.377	0.379	0.220	0.051	0.052	0.543	0.133	0.133	0.216	0.051	0.053
RWD												
1	1.278	0.314	0.315	0.185	0.043	0.043	0.282	0.067	0.067	0.248	0.057	0.057
2	1.353	0.333	0.334	0.162	0.037	0.037	0.309	0.073	0.074	0.232	0.052	0.052
3	1.448	0.356	0.358	0.171	0.039	0.040	0.324	0.078	0.078	0.206	0.045	0.049
4	1.519	0.373	0.375	0.163	0.036	0.037	0.362	0.087	0.087	0.229	0.051	0.052
5	1.519	0.373	0.375	0.206	0.047	0.048	0.373	0.089	0.089	0.230	0.050	0.051
6	1.585	0.390	0.392	0.210	0.050	0.050	0.458	0.111	0.112	0.270	0.069	0.069
7	1.683	0.414	0.417	0.226	0.053	0.053	0.615	0.150	0.151	0.316	0.074	0.074
8	1.835	0.452	0.455	0.265	0.062	0.063	0.645	0.158	0.158	0.281	0.065	0.065
9	1.886	0.464	0.468	0.299	0.070	0.071	0.655	0.160	0.161	0.365	0.084	0.085
10	1.170	0.288	0.289	0.220	0.051	0.051	0.300	0.072	0.072	0.215	0.086	0.088
Mean	1.527	0.376	0.378	0.211	0.049	0.049	0.432	0.104	0.105	0.259	0.063	0.064

For forecasting the Swedish male life-table death counts, we observe

1) fixing $K = 6$ reduces the point forecast error for the standard CoDa method, but not so for our weighted CoDa method;

2) when the eigenvalue ratio criterion determines the number of components, the weighted CoDa method is preferred; when the number of components is fixed $K = 6$, the standard CoDa method produces the smallest errors; and

3) the weighted CoDa method works better with the RWD method than the ARIMA method.

Table 2 Comparison of point forecast errors ($\times$100) between the CoDa and its weighted variant for forecasting the age-specific Swedish male data. For selecting the number of retained components K, we consider the eigenvalue ratio criterion and $K = 6$. For the ARIMA method, the optimal parameters are selected automatically based on the AICc criterion.

	Eigenvalue ratio criterion						$K = 6$					
		CoDa			Weighted CoDa			CoDa			Weighted CoDa	
h	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g
ARIMA												
1	1.490	0.368	0.369	0.288	0.073	0.071	0.283	0.068	0.068	0.296	0.070	0.070
2	1.746	0.432	0.433	0.426	0.104	0.101	0.440	0.107	0.108	0.390	0.102	0.104
3	1.954	0.483	0.485	0.528	0.139	0.133	0.631	0.155	0.155	0.504	0.129	0.130
4	2.139	0.529	0.531	0.662	0.160	0.161	0.832	0.205	0.205	0.645	0.156	0.156
5	2.286	0.566	0.568	0.782	0.178	0.161	1.040	0.257	0.257	0.764	0.161	0.172
6	2.508	0.620	0.623	0.877	0.215	0.231	1.323	0.327	0.328	0.855	0.212	0.229
7	2.697	0.667	0.670	1.003	0.247	0.247	1.616	0.399	0.401	1.028	0.252	0.248
8	2.788	0.689	0.692	1.110	0.273	0.267	1.860	0.460	0.462	1.064	0.277	0.244

Continued on next page

h	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g	KLD	JSD^s	JSD^g
9	2.938	0.725	0.729	1.296	0.318	0.320	2.193	0.542	0.544	1.252	0.310	0.314
10	1.994	0.493	0.494	0.671	0.170	0.165	1.564	0.387	0.388	0.693	0.170	0.170
Mean	2.254	0.557	0.559	0.764	0.188	0.186	1.178	0.291	0.292	0.749	0.184	0.184
RWD												
1	1.246	0.308	0.309	0.242	0.057	0.058	0.231	0.055	0.055	0.233	0.054	0.053
2	1.392	0.344	0.345	0.228	0.054	0.054	0.241	0.057	0.058	0.220	0.051	0.051
3	1.510	0.374	0.375	0.214	0.050	0.051	0.203	0.048	0.048	0.260	0.059	0.059
4	1.632	0.403	0.405	0.301	0.071	0.072	0.262	0.062	0.063	0.331	0.077	0.078
5	1.719	0.425	0.427	0.265	0.063	0.063	0.269	0.064	0.065	0.322	0.077	0.074
6	1.908	0.472	0.474	0.410	0.098	0.098	0.355	0.085	0.086	0.480	0.114	0.114
7	2.112	0.523	0.525	0.470	0.114	0.114	0.420	0.102	0.102	0.506	0.121	0.122
8	2.115	0.524	0.526	0.434	0.105	0.105	0.401	0.097	0.098	0.518	0.124	0.125
9	2.077	0.514	0.516	0.454	0.109	0.109	0.438	0.106	0.106	0.653	0.155	0.156
10	1.250	0.309	0.309	0.187	0.044	0.044	0.223	0.053	0.053	0.288	0.067	0.067
Mean	1.696	0.419	0.421	0.321	0.076	0.077	0.304	0.073	0.073	0.381	0.090	0.090

6 Conclusion

We present an extension of the CoDa method by incorporating geometrically decaying weights to estimate mean function and functional principal components. We select the estimated weight parameter to minimize various point forecast errors using a set of validation data. We compare the point forecast errors between the standard and weighted CoDa method in the testing data with the estimated weight parameter. The weighted CoDa method improves accuracy more than its standard counterpart when the eigenvalue ratio criterion selects the number of retained components for forecasting the Swedish female and male life-table death counts. We also compute the forecast errors by fixing the number of retained components to six. Including more components helps reduce the forecast errors for the standard CoDa method, but not so for our weighted CoDa extension. From the viewpoint of model parsimony and forecast accuracy, we recommend implementing the weighted CoDa method.

There are a few ways in which the present paper can be further extended, and we briefly mention two: 1) We could also compare interval forecast accuracy between the standard and weighted CoDa methods. 2) Centered log-ratio transformation is one of many possible transformations. We may also study additive log-ratio transformation in Aitchison (1986) or square-root transformation in Scealy & Welsh (2011).

Acknowledgment

The authors thank insightful comments from participants of the Living to 100 Symposia in 2023, 30th colloquium on Pensions and Retirement research, and 5th International Conference on Econometrics and Statistics (EcoSta 2022), as well as participants of a departmental seminar at the Research School of Finance, Actuarial Studies and Statistics at the Australian National University in August 2022.

References

Aburto, J. M., Basellini, U., Baudisch, A. & Villavicencio, F. (2022), Drewnowski's index to measure lifespan variation: Revisiting the Gini coefficient of the life table, Working paper, arXiv.
URL: https://arxiv.org/pdf/2111.11256.pdf

Aburto, J. M. & van Raalte, A. A. (2018), 'Lifespan dispersion in times of life expectancy fluctuation: The case of central and eastern Europe', *Demography* **55**, 2071–2096.

Aburto, J. M., Villavicencio, F., Basellini, U., Kjaergaard, S. & Vaupel, J. W. (2020), 'Dynamics of life expectancy and life span equality', *Proceedings of the National Academy of Sciences of the United States of America* **117**(10), 5250–5259.

Ahn, S. C. & Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society: Series B* **44**(2), 139–177.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman & Hall, London.

Aitchison, J. & Shen, S. M. (1980), 'Logistic-normal distributions: Some properties and uses', *Biometrika* **67**(2), 261–272.

Barceló, C., Pawlowsky, V. & Grunsky, E. (1996), 'Some aspects of transformations of compositional data and the identification of outliers', *Mathematical Geology* **28**, 501–518.

Basellini, U., Kjaergaard, S. & Camarda, C. G. (2020), 'An age-at-death distribution approach to forecast cohort mortality', *Insurance: Mathematics & Economics* **91**, 129–143.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J. & Vaupel, J. W. (2017), 'Coherent forecasts of mortality with compositional data analysis', *Demographic Research* **37**, 527–566.

Bergeron-Boucher, M.-P., Simonacci, V., Oeppen, J. & Gallo, M. (2018), 'Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: An application to Canadian provinces and territories', *The North American Actuarial Journal* **22**(1), 92–118.

Booth, H. (2006), 'Demographic forecasting: 1980 to 2005 in review', *International Journal of Forecasting* **22**(3), 547–581.

Booth, H. & Tickle, L. (2008), 'Mortality modelling and forecasting: A review of methods', *Annals of Actuarial Science* **3**(1-2), 3–43.

Canudas-Romo, V. (2010), 'Three measures of longevity: Time trends and record values', *Demography* **47**(2), 299–312.

Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C. & Caselli, G. (2005), 'Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension', *Demography* **42**(2), 243–258.

Debón, A., Chaves, L., Haberman, S. & Villa, F. (2017), 'Characterization of between-group inequality of longevity in EU countries', *Insurance: Mathematics and Economics* **75**, 151–165.

Delicado, P. (2011), 'Dimensionality reduction when data are density functions', *Computational Statistics and Data Analysis* **55**(1), 401–420.

Denuit, M., Devolder, P. & Goderniaux, A.-C. (2007), 'Securitization of longevity risk: Pricing survivor bonds with Wang transform in the Lee-Carter framework', *The Journal of Risk and Insurance* **74**(1), 87–113.

Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2009), *Actuarial Mathematics for Life Contingent Risks*, Cambridge University Press, Cambridge.

Fry, J. M., Fry, T. R. L. & McLaren, K. R. (1996), 'The stochastic specification of demand share equations: Restricting budget shares to the unit simplex', *Journal of Econometrics* **73**, 377–385.

Fuglede, B. & Topsoe, F. (2004), Jensen-Shannon divergence and Hilbert space embedding, *in* 'Proceedings of International Symposium on Information Theory'.

Horiuchi, S., Ouellette, N., Cheung, S. L. K. & Robine, J.-M. (2013), 'Modal age at death: Lifespan indicator in the era of longevity extension', *Vienna Yearbook of Population Research* **11**, 37–69.

Human Mortality Database (2021), *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. Accessed on August 26, 2021.
URL: http://www.mortality.org

Hyndman, R. J., Booth, H. & Yasmeen, F. (2013), 'Coherent mortality forecasting: the product-ratio method with functional time series models', *Demography* **50**(1), 261–283.

Hyndman, R. J. & Shang, H. L. (2009), 'Forecasting functional time series (with discussions)', *Journal of the Korean Statistical Society* **38**(3), 199–221.

Kokoszka, P., Miao, H., Petersen, A. & Shang, H. L. (2019), 'Forecasting of density functions with an application to cross-sectional and intraday returns', *International Journal of Forecasting* **35**(4), 1304–1317.

Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *Annals of Mathematical Statistics* **22**(1), 79–86.

Lam, K. K. & Wang, B. (2022), 'Multipopulation mortality modelling and forecasting: the weighted multivariate functional principal component approaches', *Journal of Applied Statistics* **in press**.

Larsson, D. (2020), 'Diseases in early modern Sweden', *Scandinavian Journal of History* **45**(4), 407–432.

Li, D., Robinson, P. M. & Shang, H. L. (2020), 'Long-range dependent curve time series', *Journal of the American Statistical Association: Theory and Methods* **115**(530), 957–971.

Oeppen, J. (2008), Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, *in* 'European Population Conference', Barcelona, Spain.
URL: http://epc2008.princeton.edu/papers/80611

Pawlowsky-Glahn, V., Egozcue, J. & Tolosana-Delgado, R. (2015), *Modeling and Analysis of Compositional Data*, John Wiley & Sons, Ltd, Chichester.

Preston, S., Heuveline, P. & Guillot, M. (2001), *Demography: Measuring and Modeling Population Processes*, Blackwell Publishers, Oxford, U.K.

Robine, J.-M. (2001), ‘Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: The case of France’, *Population: An English Selection* **13**(1), 173–193.

Scealy, J. L., de Caritat, P., Grunsky, E. C., Tsagris, M. T. & Welsh, A. H. (2017), ‘Robust principal component analysis for power transformed compositional data’, *Journal of the American Statistical Association: Theory and Methods* **110**(509), 136–148.

Scealy, J. L. & Welsh, A. H. (2011), ‘Regression for compositional data by using distributions defined on the hypersphere’, *Journal of the Royal Statistical Society: Series B* **73**(3), 351–375.

Scealy, J. L. & Welsh, A. H. (2017), ‘A directional mixed effects model for compositional expenditure data’, *Journal of the American Statistical Association: Applications and Case Studies* **112**(517), 24–36.

Shang, H. L. (2014), ‘A survey of functional principal component analysis’, *ASTA Advances in Statistical Analysis* **98**, 121–142.

Shang, H. L. & Haberman, S. (2020), ‘Forecasting age distribution of death counts: An application to annuity pricing’, *Annals of Actuarial Science* **14**(1), 150–169.

Shannon, C. E. (1948), ‘A mathematical theory of communication’, *Bell Labs Technical Journal* **27**(3), 379–423.

Shkolnikov, V. M., Andreev, E. E. & Begun, A. Z. (2003), ‘Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples’, *Demographic Research* **8**, 305–358.

van Raalte, A. A. & Caswell, H. (2013), ‘Perturbation analysis of indices of lifespan variability’, *Demography* **50**(5), 1615–1640.

van Raalte, A. A., Martikainen, P. & Myrskylä, M. (2014), ‘Lifespan variation by occupational class: Compression or stagnation over time?’, *Demography* **51**, 73–95.

Vaupel, J. W., Zhang, Z. & van Raalte, A. A. (2011), ‘Life expectancy and disparity: An international comparison of life table data’, *BMJ Open* **1**(1), e000128.

Wang, J.-L., Chiou, J.-M. & Müller, H.-G. (2016), ‘Functional Data Analysis’, *Annual Review of Statistics and Its Application* **3**, 257–295.

Wilmoth, J. R. & Horiuchi, S. (1999), 'Rectangularization revisited: Variability of age at death within human populations', *Demography* **36**(4), 475–495.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org