



# City Research Online

## City St George's, University of London

**Citation:** Shang, H. L. & Haberman, S. (2025). Forecasting age distribution of life-table death counts via alpha-transformation. *Scandinavian Actuarial Journal*, 2025(4), pp. 387-403. doi: 10.1080/03461238.2024.2425723

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://city-ref-test.eprints-hosting.org/id/eprint/33958/>

**Link to published version:** <https://doi.org/10.1080/03461238.2024.2425723>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# Forecasting age distribution of life-table death counts via $\alpha$ -transformation

Han Lin Shang<sup>✉\*</sup>

Department of Actuarial Studies and Business Analytics  
Macquarie University

Steven Haberman<sup>✉</sup>

Bayes Business School  
City St George's, University of London

October 31, 2024

## Abstract

We introduce a compositional power transformation, known as an  $\alpha$ -transformation, to model and forecast a time series of life-table death counts, possibly with zero counts observed at older ages. As a generalisation of the isometric log-ratio transformation (i.e.,  $\alpha = 0$ ), the  $\alpha$  transformation relies on the tuning parameter  $\alpha$ , which can be determined in a data-driven manner. Using the Australian age-specific period life-table death counts from 1921 to 2020, the  $\alpha$  transformation can produce more accurate short-term point and interval forecasts than the log-ratio transformation. The improved forecast accuracy of life-table death counts is of great importance to demographers and government planners for estimating survival probabilities and life expectancy and actuaries for determining annuity prices and reserves for various initial ages and maturity terms.

*Keywords:* compositional data analysis; centre log-ratio; isometric log-ratio; principal component analysis; functional time-series forecasting

---

\*Corresponding address: Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW 2068, Australia; Telephone number: +61(2) 9850 4689; Email: hanlin.shang@mq.edu.au

# 1 Introduction

Actuaries and statistical demographers have been interested in developing models for mortality forecasting (see [Booth, 2006](#); [Booth and Tickle, 2008](#); [Basellini et al., 2023](#), for comprehensive reviews). From an actuarial perspective, mortality forecasts are important to manage adverse financial effects of mortality improvements over time on life or fixed-term annuities and sustainable pension system ([Pollard, 1987](#)). From a demographic perspective, mortality forecasts are an essential part of forward planning for national health and aged care systems and are an important component of population forecasts (see, e.g., [Shang et al., 2016](#); [Hyndman et al., 2021](#), among others).

In demography, three functions related to mortality are widely studied: central mortality rates, survival probabilities, and life-table death counts. Several scholars have proposed new approaches for forecasting age-specific mortality rates using statistical models (see, e.g., [Booth, 2006](#); [Booth and Tickle, 2008](#); [Cairns et al., 2008](#); [Shang et al., 2011](#), for reviews). Arguably, the most famous model is the [Lee and Carter \(1992\)](#) model. Let  $m_{x,t}$  be log central mortality rates for age  $x$  in year  $t$ . The Lee-Carter model is given by

$$m_{x,t} = a_x + b_x \kappa_t + \epsilon_{x,t}, \quad t = 1, 2, \dots, n, \quad x = 0, 1, \dots, p, \quad (1)$$

where  $a_x$  is the age pattern of the log central mortality rates averaged across years;  $b_x$  is the first principal component capturing relative change in the log central mortality rate at each age  $x$  with a constraint  $\sum_{x=0}^p b_x = 1$ ;  $\kappa_t$  is the first set of principal component scores measuring general level of the log central mortality rate at year  $t$  with a constraint  $\sum_{t=1}^n \kappa_t = 0$ ; and  $\epsilon_{x,t}$  is the model residual at age  $x$  and year  $t$ . Both constraints ensure the parameter identifiability in the Lee-Carter model. The principal component analysis is performed on the mean corrected matrix of log mortality rates. The principal component scores  $\kappa_t$  are extrapolated by a random walk with drift method, from which point forecasts are obtained by (1) with the fixed  $a_x$  and  $b_x$ . In demography, there exist several extensions and modifications of the Lee-Carter model (see, e.g., [Booth and Tickle, 2008](#); [Shang et al., 2011](#); [Basellini et al., 2023](#), for review).

Apart from modelling central mortality rates, one can also model a redistribution of life-table death counts over time (see, e.g., [Pascariu et al., 2019](#); [Basellini et al., 2020](#)). For instance, in many developed countries, deaths at younger ages have been gradually shifting toward older ages. The shift of the distribution symbolises longevity risk, which is a major issue for life insurers and pension funds, especially in the selling and risk management of annuity products (see [Brouhns et al., 2002](#); [Denuit et al., 2007](#), for a discussion). In addition to providing an informative description of the mortality experience of a population, the life-table death counts yield readily available information on “central longevity indicators”, i.e., mean, median, and mode age at death

(Canudas-Romo, 2010), and lifespan variability (Robine, 2001; Vaupel et al., 2011; van Raalte and Caswell, 2013; van Raalte et al., 2014). For many developed countries, a decrease in variability over time may be observed through the standard deviation of life-table ages at death, Gini coefficient (Wilmoth and Horiuchi, 1999; van Raalte and Caswell, 2013; Shang et al., 2022) or Drewnowski's index (Aburto et al., 2022). Life-table death counts provide important insights into longevity risk and lifespan variability that cannot be easily quantified from either central mortality rates or survival probabilities.

Life-table death counts relate strongly with the probability density function. For each year, the life-table death counts are non-negative and sum to a radix of  $10^5$ . Because of the two constraints, the age-specific life-table death counts can be viewed as compositional data. The sample space of compositional data is a simplex

$$S^{D-1} = \left\{ (d_1, d_2, \dots, d_D)^\top, \quad d_x > 0, \quad \sum_{x=1}^D d_x = c \right\},$$

where  $d_x$  denotes life-table death count for age  $x$ ,  $D$ -part compositional data are mapped from the simplex into a  $(D - 1)$ -dimensional real space,  $c$  is a fixed constant, and  $^\top$  denotes vector transpose.

Compositional data are defined as a random vector of  $D$  compositions  $[d_1, d_2, \dots, d_D]$  with non-negative values whose sum is a given constant, typically one (portions), 100 (%) and  $10^6$  for parts per million in geochemical compositions. These data arise in many scientific fields, such as geology (geochemical compositions), economics (income/expenditure distribution), medicine (body composition), food industry (goods composition), chemistry (chemical composition), ecology (abundance of different species), and demography. In statistics, Scealy et al. (2015) apply compositional data analysis (CoDa) to study the concentration of chemical compositions in sediment or rock samples. In economics, Scealy and Welsh (2017) apply CoDa to analyse the total weekly household expenditure on food and housing costs. Kokoszka et al. (2019) apply CoDa to model and forecast density functions of financial assets. In demography, Oeppen (2008), Bergeron-Boucher et al. (2017), Shang and Haberman (2020), and Shang et al. (2022) introduce a principal component approach to model and forecast life-table death counts within a compositional data analytic framework.

In the CoDa, Aitchison and Shen (1980) and Aitchison (1982, 1986) transform the compositional data from the simplex to Euclidean space using a centre log-ratio (clr) transformation:

$$s = \{s_x\}_{x=1, \dots, D} = \ln \left\{ \frac{d_x}{g(d)} \right\}_{x=1, \dots, D},$$

where  $g(d) = (\prod_{x=1}^D d_x)^{1/D}$ . Extending from the clr transformation, Egozcue et al. (2003) propose the isometric log-ratio (ilr) transformation, and it has been promoted as the more theoretically

correct method in CoDa ([Greenacre and Grunsky, 2019](#)). The ilr transformation is defined as:

$$z = \mathbf{H}s$$

where  $\mathbf{H}$  is a  $(D - 1) \times D$  Helmert sub-matrix (see, e.g., [Lancaster, 1965](#); [Dryden and Mardia, 1998](#)). In geographic mapping, Helmert sub-matrices are used to describe a coordinate transformation. The Helmert sub-matrix of order  $D$  is a square matrix defined as

$$\mathbf{H}_D = \begin{bmatrix} \frac{1}{\sqrt{D}} & \frac{1}{\sqrt{D}} & \frac{1}{\sqrt{D}} & \cdots & \frac{1}{\sqrt{D}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{D(D-1)}} & \frac{1}{\sqrt{D(D-1)}} & \frac{1}{\sqrt{D(D-1)}} & \cdots & -\frac{(D-1)}{\sqrt{D(D-1)}} \end{bmatrix}.$$

$\mathbf{H} = (\mathbf{H}_D)_{(D-1) \times D}$  is introduced to remove any redundant dimension presented due to the compositional constraint. The  $\mathbf{H}$  matrix is an orthogonal matrix without the first row.

An issue with any log-ratio transformation, including the clr and ilr, is that there is no guarantee that the transformed data are multivariate normal distributed ([Tsagris and Stewart, 2020](#)). This distributional assumption plays a crucial role in the validity of the log-ratio analysis ([Aitchison and Shen, 1980](#)). Another issue is that the log-ratio transformation cannot be applied to data containing zeros. The presence of zero counts makes the log transformation invalid. In statistical demography, there exist some missing-value imputation methods for handling zeros, but these methods are often ad-hoc by omitting, aggregating, or adding small arbitrary values to zero values (see, e.g., [Fry et al., 2000](#); [Martín-Fernández et al., 2003](#)). This is not ideal, however, [Greenacre \(2021\)](#) compared four different algorithms to substitute zeros and showed that the resulting conclusions can be strongly sensitive to the method of zero substitution. In a similar spirit to the Box-Cox transformation, we present a power transformation, known as  $\alpha$  transformation, which can handle zero counts.

The paper is organised as follows. In Section 3, we show that the log-ratio analysis can be seen as a special case of  $\alpha$  transformation, where the optimal value of  $\alpha$  can be determined in a data-driven way, and this is discussed in Section 3.3. Via the  $\alpha$  transformation, the functional time-series forecasting method can be applied to model and forecast the unconstrained data. We introduce the Australian age-specific life-table death counts in Section 2, which we use to illustrate the methodology. In Section 4, we describe the model fitting and forecasting and introduce the measures of point forecast error in Section 4.2, which we use in Section 4.3 to evaluate and compare the point forecast accuracy. Using the coverage probability difference and mean interval score in Section 4.4, we evaluate and compare the interval forecast accuracy in Section 4.5. The conclusion is presented in Section 5, along with some ideas on how the methodology may be further extended.

## 2 Age- and sex-specific life-table death counts

We study Australian age- and sex-specific life-table death counts from 1921 to 2020, obtained from the [Human Mortality Database \(2023\)](#). The life-table radix is fixed at  $10^5$  at age 0 and gradually decreases to 0 at age 110+ for every calendar year. Due to rounding, there may be zero counts for older ages at some years. One potential remedy is to use the probability of dying (i.e.,  $q_x$ ) and the life-table radix to recalculate our estimated age-specific death counts (see, e.g., [Shang and Haberman, 2020](#)). The estimated death counts are more detailed and smoother than those reported in the [Human Mortality Database \(2023\)](#).

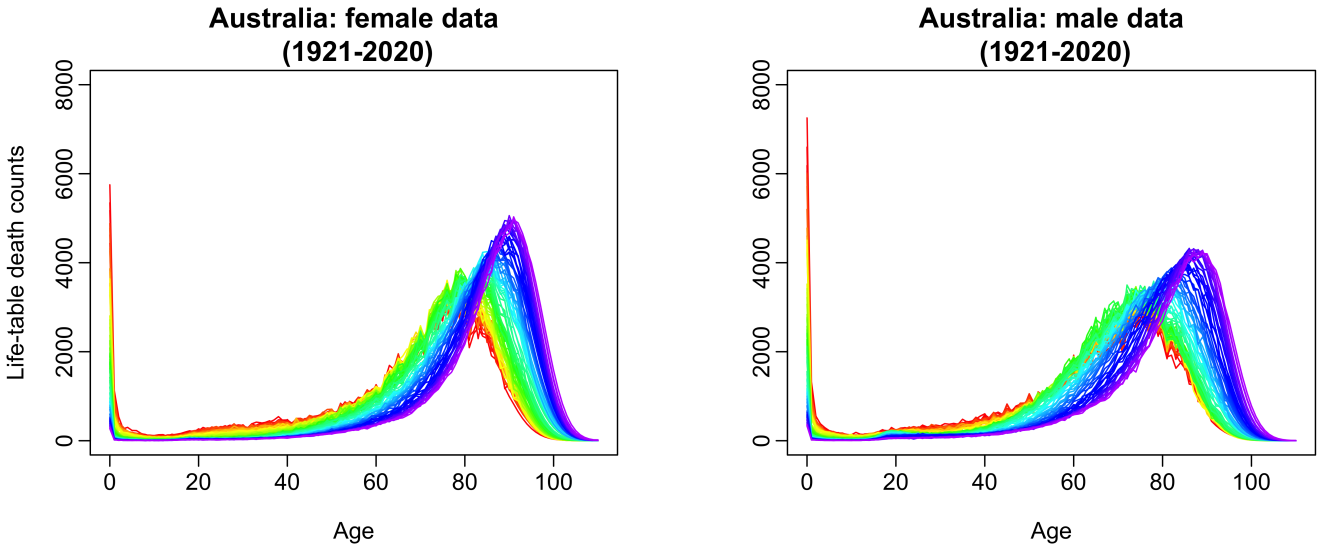


Figure 1: Rainbow plots of age-specific life-table death count from 1921 to 2020 in a single-year group. The oldest years are shown in red, with the most recent in violet. Curves are ordered chronologically according to the colours of the rainbow.

In Figure 1, we present rainbow plots of the female and male age-specific life-table death counts in Australia for a single year of age groups. Both panels of Figure 1 demonstrate a decreasing trend in infant mortality and a negatively skewed distribution for the life-table death counts, where the peaks shift to older ages for females and males. The gradual change in distribution confirms the longevity risk experienced in recent years.

## 3 The $\alpha$ transformation

One common approach in compositional data analysis involves transforming compositional data to address non-normality and heteroscedasticity, with the  $\alpha$  transformation being one such technique. Akin to Box-Cox transformation, the  $\alpha$  transformation of a compositional vector  $d_t \in \mathbb{S}^D$  is defined

by

$$\mathbf{z}_t^\alpha = \mathbf{H} \cdot \left( \frac{D\mathbf{d}_t^\alpha - \mathbb{1}_D}{\alpha} \right) \quad (2)$$

where  $\alpha \in [0, 1]$  and  $\mathbf{d}_t^\alpha$  is given as

$$\mathbf{d}_t^\alpha = \left( \frac{d_{t,1}^\alpha}{\sum_{j=1}^D d_{t,j}^\alpha}, \dots, \frac{d_{t,D}^\alpha}{\sum_{j=1}^D d_{t,j}^\alpha} \right)^\top,$$

and  $\mathbb{1}_D$  is the  $D$ -dimensional vector of ones, and  $\mathbf{H}$  is  $(D-1) \times D$  Helmert matrix. The compositional power transformation is invertible:

$$\boldsymbol{\nu}_t^\alpha = D\mathbf{d}_t^\alpha = \alpha \times \mathbf{H}^\top \mathbf{z}_t^\alpha + \mathbb{1}_D, \quad (3)$$

where

$$\mathbf{d}_t = \left( \frac{v_{t,1}^{1/\alpha}}{\sum_{j=1}^D v_{t,j}^{1/\alpha}}, \dots, \frac{v_{t,D}^{1/\alpha}}{\sum_{j=1}^D v_{t,j}^{1/\alpha}} \right).$$

As noted in Tsagris et al. (2016), when  $\alpha = 0$ , it corresponds to the isometric log-ratio transformation, whereas  $\alpha = 1$ , it corresponds to Euclidean data analysis (eda) ignoring the compositional constraint (see, e.g., Baxter, 1995, 2001; Baxter et al., 2005). When  $\alpha = 0$  and one does not pre-multiply Helmert matrix  $\mathbf{H}$ , the formulation reduces to the clr.

### 3.1 Modelling and forecasting a time series of density-valued curves

Through the  $\alpha$  transformation, we obtain a time series of unconstrained matrix  $\mathbf{z}^\alpha = (\mathbf{z}_1^\alpha, \dots, \mathbf{z}_n^\alpha)$  where  $\mathbf{z}_1^\alpha = (z_{1,1}^\alpha, \dots, z_{1,D}^\alpha)^\top$ . We apply a principal component analysis to decompose  $\mathbf{z}^\alpha$ . Under a finite second moment, we compute the sample covariance of  $\mathbf{z}^\alpha$ . Via Mercer's lemma, the sample covariance, denoted by  $\hat{\mathbf{C}}$ , can be estimated by

$$\hat{\mathbf{C}} = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\boldsymbol{\phi}}_k \hat{\boldsymbol{\phi}}_k^\top,$$

where  $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots$  are non-increasing eigenvalues. By Karhunen-Loève expansion, for a given year  $t$ ,

$$\mathbf{z}_t^\alpha = \bar{\mathbf{z}}^\alpha + \sum_{k=1}^K \hat{\beta}_{t,k} \hat{\boldsymbol{\phi}}_k + \mathbf{e}_t,$$

where  $\bar{\mathbf{z}}^\alpha = \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^\alpha$  is the mean term,  $\hat{\beta}_{t,k}$  denotes the  $k^{\text{th}}$  set of the estimated principal component score for time  $t$ ,  $\hat{\boldsymbol{\phi}}_k$  denotes the  $k^{\text{th}}$  set of the estimated principal component, and  $\mathbf{e}_t$

denotes the model error term encompassing the remaining principal component pairs.

To determine the optimal number of principal components, we select  $K$  via an eigenvalue ratio criterion based on Li et al. (2020).

$$K = \arg \min_{1 \leq k \leq K_{\max}} \left\{ \frac{\hat{\lambda}_{k+1}}{\hat{\lambda}_k} \times \mathbb{1} \left\{ \frac{\hat{\lambda}_k}{\hat{\lambda}_1} \geq \theta \right\} + \mathbb{1} \left\{ \frac{\hat{\lambda}_k}{\hat{\lambda}_1} < \theta \right\} \right\}, \quad (4)$$

where  $\mathbb{1}\{\cdot\}$  denotes the binary indicator function,  $\theta = \frac{1}{\ln[\max(\hat{\lambda}_1, n)]}$  is a small positive value, and instead of searching through  $n$  sets of principal component pairs, we restrict our searching range by setting  $K_{\max} = \#\{k | \hat{\lambda}_k \geq \frac{1}{n} \sum_{k=1}^n \hat{\lambda}_k, k \geq 1\}$  (see also Li et al., 2020). Equation (4) can be viewed as a ridge-type eigenvalue ratio criterion. The functional time-series forecasting method is robust to over-estimating  $K$ , but under-estimating  $K$  can result in inferior accuracy. As in Hyndman et al. (2013), we also consider  $K = 6$ .

Having determined the optimal number of principal components  $K$ , we apply an autoregressive integrated moving average (ARIMA) model to obtain the  $h$ -step-ahead forecast. Conditional on the historical data and estimated mean and principal components, the  $h$ -step-ahead forecast of  $z_{n+h}^\alpha$  can be expressed as

$$\hat{z}_{n+h|n}^\alpha = E[z_{n+h}^\alpha | \hat{\Phi}, z^\alpha] = \bar{z}^\alpha + \sum_{k=1}^K \hat{\beta}_{n+h|n,k} \hat{\phi}_k,$$

where  $\hat{\Phi} = (\hat{\phi}_1, \dots, \hat{\phi}_K)^\top$ , and  $\hat{\beta}_{n+h|n,k}$  denotes the  $h$ -step-ahead forecast of the  $k^{\text{th}}$  principal component scores. By taking the inverse  $\alpha$ -transformation in (3), we obtain the  $h$ -step-ahead forecast, denoted by  $\hat{\mathbf{d}}_{n+h|n} = (\hat{d}_{n+h|n,1}, \dots, \hat{d}_{n+h|n,D})$ .

We use the automatic algorithm of Hyndman and Khandakar (2008) to choose the optimal autoregressive order  $p$ , moving average order  $q$ , and difference order  $d$  based on the optimal AIC with a correction for a small finite sample size (see, e.g., Hurvich and Tsai, 1993). After identifying the optimal ARIMA model, the maximum likelihood method can be used to estimate the regression coefficients associated with the ARIMA model.

### 3.2 Construction of pointwise prediction intervals for the $\alpha$ transformation

Following Hyndman and Shang (2009), we incorporate two sources of uncertainty: truncation errors in the principal component decomposition and forecast errors in the forecast principal component scores. To obtain bootstrap forecasts of the scores, we assess the variability between the forecast and observed principal component scores. With a univariate time-series model, we can obtain multi-step-ahead forecasts for the scores,  $\{\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \dots, \hat{\beta}_{n,k}\}$  for  $k = 1, 2, \dots, K$ . Let the

$h$ -step-ahead forecast errors be given by  $\zeta_{t,h,k} = \hat{\beta}_{t,k} - \hat{\beta}_{t|t-h,k}$  for  $t = h + 1, \dots, n$ . These can then be sampled with replacement to give a bootstrap sample of  $\beta_{n+h,k}$ :

$$\hat{\beta}_{n+h|n,k}^{(b)} = \hat{\beta}_{n+h|n,k} + \hat{\zeta}_{*,h,k}^{(b)}, \quad b = 1, \dots, B,$$

where  $B = 1,000$  symbolises the number of bootstrap replications and  $\hat{\zeta}_{*,h,k}^{(b)}$  are sampled with replacement from  $\hat{\zeta}_{t,h,k}$ .

Assuming the first  $K$  principal components approximate the data relatively well, the model residual should contribute nothing but random noise. Consequently, we can bootstrap the model residuals by sampling with replacement from  $\{e_1, e_2, \dots, e_n\}$ .

Adding two components of variability, we obtain  $B$  variants for  $\hat{z}_{n+h|n}^{\alpha,(b)}$ :

$$\hat{z}_{n+h|n}^{\alpha,(b)} = \bar{z}^{\alpha} + \sum_{k=1}^K \hat{\beta}_{n+h|n,k}^{(b)} \hat{\phi}_k + e_{n+h}^{(b)},$$

where  $\hat{\beta}_{n+h|n,k}^{(b)}$  denotes the forecast of bootstrapped principal component scores. With the bootstrapped  $\{\hat{z}_{n+h|n}^{\alpha,(1)}, \hat{z}_{n+h|n}^{\alpha,(2)}, \dots, \hat{z}_{n+h|n}^{\alpha,(B)}\}$ , we take the inverse  $\alpha$  transformation to obtain bootstrap forecasts for age distribution of death counts. At the  $100(1 - \gamma)\%$  nominal coverage probability, the pointwise prediction intervals are obtained by  $\gamma/2$  and  $1 - \gamma/2$  quantiles based on  $\{\hat{d}_{n+h|n}^{(1)}, \hat{d}_{n+h|n}^{(2)}, \dots, \hat{d}_{n+h|n}^{(B)}\}$ , where  $\gamma$  represents a level of significance, customarily  $\gamma = 0.2$  or  $0.05$ .

### 3.3 Selection of $\alpha$ parameter

We split the data with a sample size of  $n$  into a training sample and a testing set. Within the training sample, we further divide it into a training set and a validation set. With a maximum forecast horizon  $H$ , the testing set contains the elements  $(n - H + 1) : n$ , the validation set contains  $(n - 2H + 1) : (n - H)$ , and the initial training set contains  $1 : (n - 2H)$ . To select the weight parameter, we aim to find  $\alpha$  that minimises a point forecast error in Section 4.2 and an interval forecast error in Section 4.4. In Figure 2, we present a visual display of the sample splitting.

|                |                          |                   |
|----------------|--------------------------|-------------------|
| $1 : (n - 2H)$ | $(n - 2H + 1) : (n - H)$ | $(n - H + 1) : n$ |
| Training       | Validation               | Testing           |

Figure 2: Illustration of the cross-validation method. A model is constructed using data in the training set to forecast data in the validation set. The model's predictive ability is evaluated based on either point or interval forecast error. The optimal value of  $\alpha$  is determined based on the minimal point or interval forecast error in the validation set. For sensitivity analysis, we consider  $H = 10$  or  $20$ .

The forecast accuracy of  $\alpha$  transformation depends on the optimal selection of the  $\alpha$  value for a given forecast horizon. In Figure 3a, we present the  $\alpha$ -transformed data of female life-table death counts, where  $\alpha = 0.3544$  is selected based on the Kullback-Leibler divergence for  $H = 10$ . For comparison, we also show the transformed data where  $\alpha = 0$  and  $1$ . As one alters the value of  $\alpha$ , it impacts the shape of the transformed time series of densities.

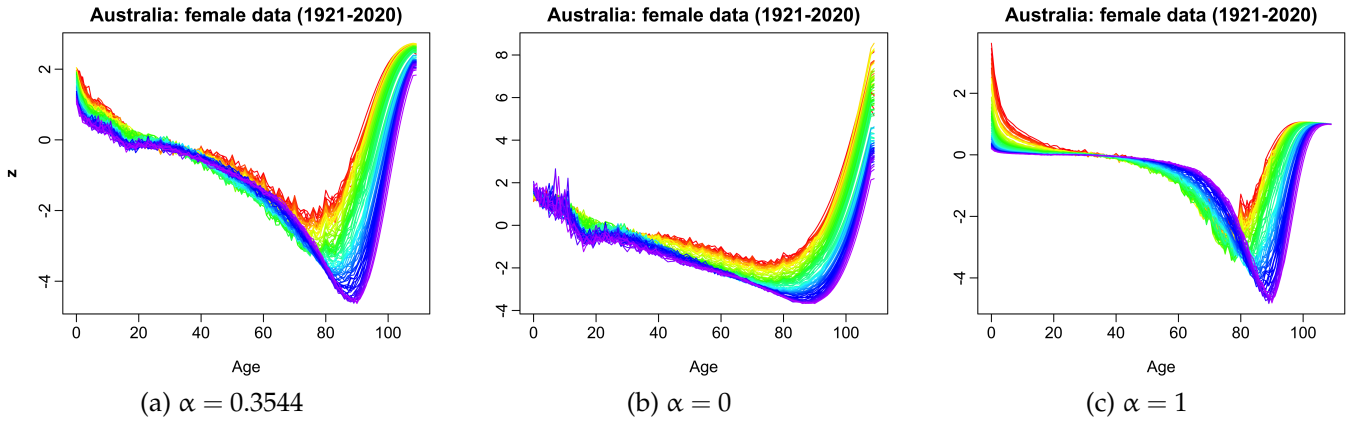


Figure 3: The  $\alpha$ -transformed data of the Australian female life-table death counts. Based on the Kullback-Leibler divergence described in Section 4.2,  $\alpha = 0.3544$  is optimal. When  $\alpha = 0$ , it corresponds to the ilr method; when  $\alpha = 1$ , it corresponds to the eda method.

## 4 Model fitting and forecasting

Based on the historical death counts from 1921 to 2020 (i.e., 100 observations), we apply the eigenvalue ratio criterion in (4) to estimate the number of retained principal components, which is  $K = 2$ . For modelling each set of the scores, we apply the automatic algorithm of Hyndman and Khandakar (2008) to select the optimal orders of the ARIMA model shown in the second

row of Figure 4. From the forecast principal component scores, one can observe an increasing trend indicating a continuation of the recent trend, i.e., compression of life-table death counts (see also Shang and Haberman, 2020). By multiplying the forecast principal component scores with the estimated functional principal components and adding the mean function, we produce 10-steps-ahead forecasts of life-table death counts between 2021 and 2030, shown in Figure 4d.

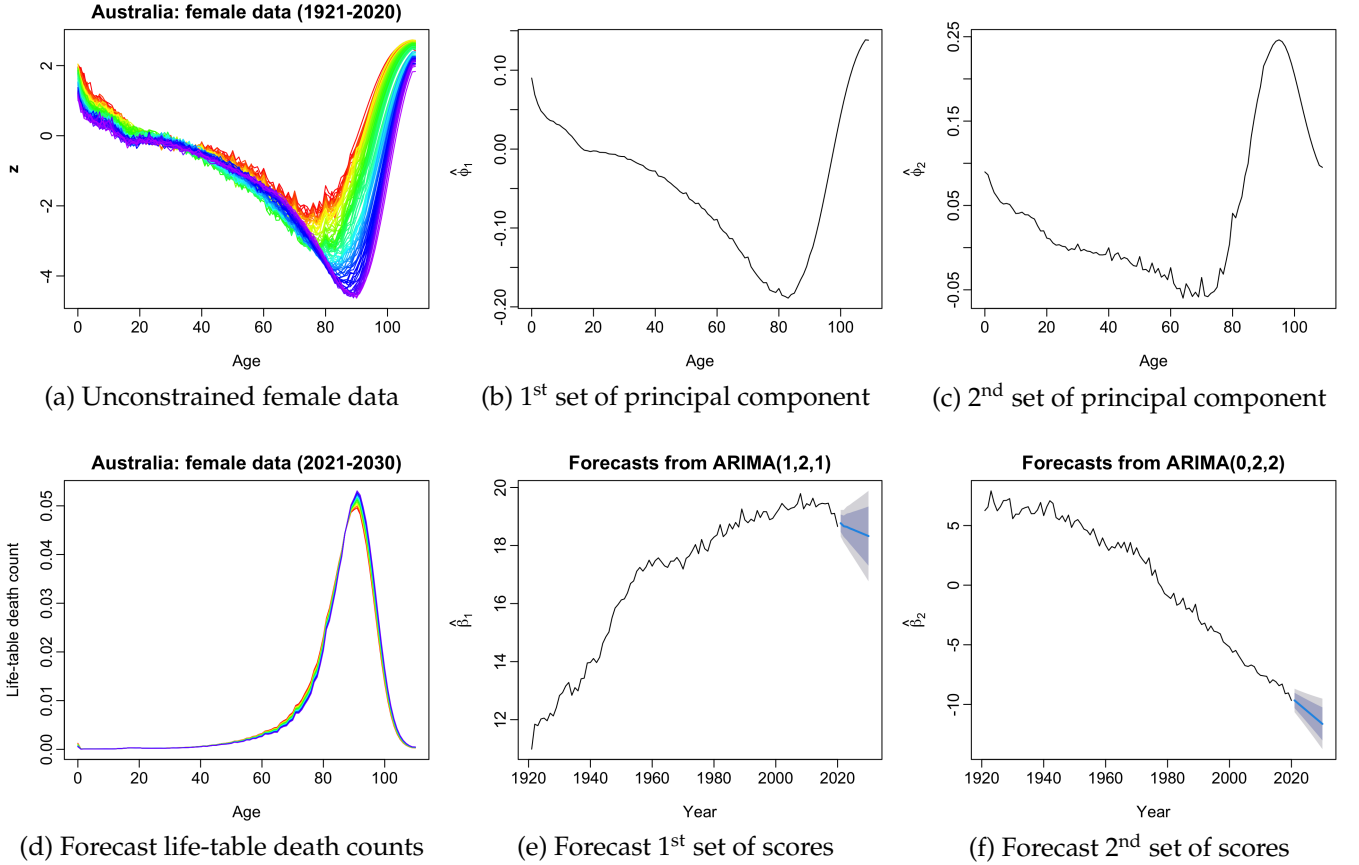


Figure 4: Elements of the  $\alpha$ -transformation for analysing the female age-specific life-table death counts in Australia. As determined by the eigenvalue criterion, we present the first two principal components and their associated principal component scores.

In Figure 5, we display the  $\alpha$ -transformed data of male life-table death counts, where  $\alpha = 0.0528$  on the basis of the Kullback-Leibler divergence with  $H = 10$ . In contrast to the female life-table death counts, the estimated principal components exhibit different shape patterns. Similar to the female life-table death counts, the forecast principal component scores show an increasing trend in the future, indicating a continuation of the recent compression of life-table death counts. By multiplying the forecast principal component scores with the estimated functional principal components and adding the mean function, we produce 10-steps-ahead forecasts of life-table death counts shown in Figure 5d. The forecast age distribution of death counts is negatively skewed.

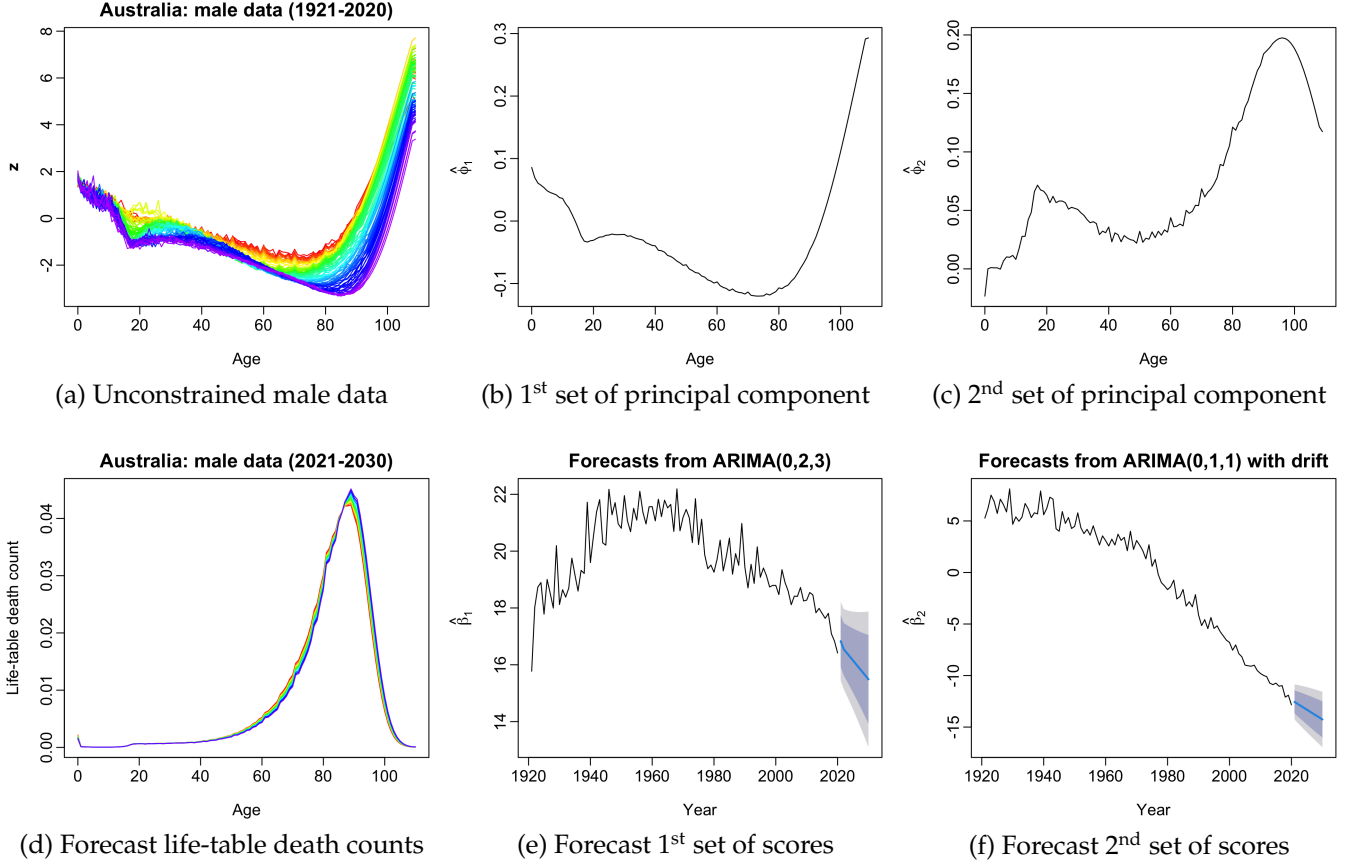


Figure 5: Elements of the  $\alpha$ -transformation for analysing the male age-specific life-table death counts in Australia. We present the first two principal components and their associated principal component scores.

Apart from the graphical display, we can measure the in-sample goodness-of-fit of the  $\alpha$  transformation to the observed life-table death counts via an  $R^2$  criterion and root mean squared error (RMSE). They can be defined as

$$R^2 = 1 - \frac{\sum_{x=1}^{111} \sum_{t=1}^{100} (d_{t,x} - \hat{d}_{t,x})^2}{\sum_{x=1}^{111} \sum_{t=1}^{100} (d_{x,t} - \bar{d}_x)^2}$$

$$RMSE = \sqrt{\frac{1}{111 \times 100} \sum_{x=1}^{111} \sum_{t=1}^{100} (d_{t,x} - \hat{d}_{t,x})^2}$$

where  $d_{t,x}$  denotes the observed age-specific life-table death count for age  $x$  in year  $t$ , and  $\hat{d}_{t,x}$  denotes its estimate. In Table 1, we compare the values of  $R^2$  and RMSE between the ilr method and  $\alpha$  transformation. For both datasets, the eigenvalue ratio criterion selects the number of components  $K = 2$ . The  $\alpha$  transformation comparably improves the goodness of fit.

Table 1: The values of  $R^2$  and in-sample RMSE between the ilr method and  $\alpha$  transformation for the Australian female and male life-table death counts.

| Method                  | $R^2$  |        | RMSE    |         |
|-------------------------|--------|--------|---------|---------|
|                         | Female | Male   | Female  | Male    |
| ilr                     | 0.9953 | 0.9911 | 0.00111 | 0.00141 |
| $\alpha$ transformation | 0.9968 | 0.9915 | 0.00091 | 0.00138 |

## 4.1 Model comparison

The ilr and eda are special cases of the  $\alpha$  transformation. In Section 4.1.1, we also include the clr method without pre-multiplying the Helmert sub-matrix in (2). In Section 4.1.2, we compare our methods with the maximum entropy model of Pascariu et al. (2019) based on a set of finite moments.

### 4.1.1 Centered log-ratio method

There are two versions of the clr: 1) The eigenvalue ratio criterion is a data-driven way of selecting the number of retained components, and it allows the possibility of having more components to improve the in-sample goodness-of-fit. Instead of the random walk with drift, we prefer the ARIMA model as the forecasting method because it can handle nonstationary series. 2) As advocated in Hyndman et al. (2013), there is a little price to pay for potentially overfitting by choosing a large number of  $K$ , such as  $K = 6$ .

### 4.1.2 Maximum entropy mortality (MEM) model

There are several attempts to model the age distribution of death counts by parametric mortality laws. Dellaportas et al. (2001) used the Heligman and Pollard's (1980) model to fit age-specific life-table death counts with Bayesian methods, and Mazzuco et al. (2018) models mortality by fitting a half-normal and a skew-bimodal-normal distribution to the observed age-at-death density function. While these approaches are parametric, Pascariu et al. (2019) consider a nonparametric alternative based on statistical moments. The forecast life-table death counts for a population can be determined by extrapolating finite statistical moments derived from the observed data.

The extrapolation is achieved through multivariate time series models, such as multivariate random walk with drift. The age distribution of death counts is then estimated in time from the predicted moments. In practice, the first four moments are sufficient to characterise a distribution.

## 4.2 Point forecast error measures

Since life-table death counts can be considered a probability density function, we study some density evaluation measures. These measures include the discrete version of the Kullback-Leibler divergence (Kullback and Leibler, 1951) and the square root of the Jensen-Shannon divergence (Shannon, 1948). For two probability density functions, denoted by  $\mathbf{d}_{m+\xi}$  and  $\hat{\mathbf{d}}_{m+\xi}$ , the *symmetric* version of the Kullback-Leibler divergence is defined as

$$\begin{aligned} \text{KLD}(h) &= D_{\text{KL}}(\mathbf{d}_{m+\xi} \parallel \hat{\mathbf{d}}_{m+\xi}) + D_{\text{KL}}(\hat{\mathbf{d}}_{m+\xi} \parallel \mathbf{d}_{m+\xi}) \\ &= \frac{1}{111 \times (H+1-h)} \sum_{\xi=h}^H \sum_{x=1}^{111} \mathbf{d}_{m+\xi,x} \cdot (\ln \mathbf{d}_{m+\xi,x} - \ln \hat{\mathbf{d}}_{m+\xi,x}) \\ &\quad + \frac{1}{111 \times (H+1-h)} \sum_{\xi=h}^H \sum_{x=1}^{111} \hat{\mathbf{d}}_{m+\xi,x} \cdot (\ln \hat{\mathbf{d}}_{m+\xi,x} - \ln \mathbf{d}_{m+\xi,x}), \end{aligned}$$

which is non-negative, for  $h = 1, 2, \dots, H$ . Let  $m$  denote the number of data samples in the training set for selecting the optimal  $\alpha$  in the validation set. With a slight abuse of notation,  $m$  can also represent the data in the training sample when examining the out-of-sample forecast accuracy.

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

$$\text{JSD}(h) = \frac{1}{2} D_{\text{KL}}(\mathbf{d}_{m+\xi} \parallel \boldsymbol{\delta}_{m+\xi}) + \frac{1}{2} D_{\text{KL}}(\hat{\mathbf{d}}_{m+\xi} \parallel \boldsymbol{\delta}_{m+\xi}),$$

where  $\boldsymbol{\delta}_{m+\xi}$  measures a common quantity between  $\mathbf{d}_{m+\xi}$  and  $\hat{\mathbf{d}}_{m+\xi}$ . As in Shang and Haberman (2020), we consider two cases: the arithmetic mean and the geometric mean

$$\begin{aligned} \boldsymbol{\delta}_{m+\xi} &= \frac{1}{2} (\mathbf{d}_{m+\xi} + \hat{\mathbf{d}}_{m+\xi}), \\ \boldsymbol{\delta}_{m+\xi} &= \sqrt{\mathbf{d}_{m+\xi} \hat{\mathbf{d}}_{m+\xi}}. \end{aligned}$$

Let  $\text{JSD}^a(h)$  be the Jensen-Shannon divergence with the arithmetic mean, while  $\text{JSD}^g(h)$  be the Jensen-Shannon divergence with the geometric mean.

For  $h = 1, 2, \dots, H$ , we compute the averaged KLD,  $\text{JSD}^a$  and  $\text{JSD}^g$  defined as

$$\begin{aligned} \text{KLD} &= \frac{1}{H} \sum_{h=1}^H \text{KLD}(h), \\ \text{JSD}^a &= \frac{1}{H} \sum_{h=1}^H \text{JSD}^a(h), \end{aligned}$$

$$\text{JSD}^g = \frac{1}{H} \sum_{h=1}^H \text{JSD}^g(h).$$

### 4.3 Point forecast results

As explained in [Booth \(2006\)](#), there exist three approaches to forecasting demographic processes, namely extrapolation, expectation and theory-based structural modelling involving exogenous variables. As our method belongs to a time-series extrapolation approach, we focus on a short-term forecast horizon, such as  $H = 10$  or  $20$ . We implement an *expanding window* approach to evaluate and compare point and interval forecasts. The expanding window is a forecasting technique where we iteratively increase the size of the training sample to make our predictions.

Based on the KLD,  $\text{JSD}^a$  and  $\text{JSD}^g$ , we evaluate and compare the point forecast accuracy of the  $\alpha$  transformation and its two special cases (ilr and eda). We consider forecasting each set of the estimated principal component scores by the selected ARIMA model based on the correct Akaike information criterion for each horizon  $h = 1, 2, \dots, H$ .

Table 2: For two different forecast horizons  $H = 10$  and  $20$ , we present a comparison of the point forecast accuracy, measured by the KLD,  $\text{JSD}^a$  and  $\text{JSD}^g$ , among the  $\alpha$  transformation, isometric log-ratio (ilr), and Euclidean data analysis (eda), using the holdout sample of the Australian female and male life-table death counts. For comparison, we include the clr method and the MEM of [Pascariu et al. \(2019\)](#). Also, we apply the Lee-Carter model to forecast age-specific smoothed mortality rates, from which we obtain the forecast life-table death counts. The smallest errors among the methods are highlighted in bold for  $H = 10$  and  $20$ .

| Sex | Criterion      | H  | $\alpha$ | Eigenvalue ratio |        |        | $\alpha$      | K = 6         |        |               | MEM           | LC     |
|-----|----------------|----|----------|------------------|--------|--------|---------------|---------------|--------|---------------|---------------|--------|
|     |                |    |          | ilr              | eda    | clr    |               | ilr           | eda    | clr           |               |        |
| F   | KLD            | 10 | 0.0062   | 0.0099           | 0.0925 | 0.0109 | <b>0.0036</b> | <b>0.0036</b> | 0.0104 | 0.0037        | 0.0053        | 0.0050 |
|     |                | 20 | 0.0101   | 0.0134           | 0.0729 | 0.0150 | <b>0.0045</b> | 0.0056        | 0.0241 | 0.0059        | 0.0123        | 0.0077 |
|     | $\text{JSD}^a$ | 10 | 0.0016   | 0.0025           | 0.0127 | 0.0027 | <b>0.0009</b> | <b>0.0009</b> | 0.0023 | 0.0009        | 0.0013        | 0.0012 |
|     |                | 20 | 0.0025   | 0.0033           | 0.0114 | 0.0037 | <b>0.0014</b> | <b>0.0014</b> | 0.0055 | 0.0015        | 0.0030        | 0.0019 |
|     | $\text{JSD}^g$ | 10 | 0.0016   | 0.0025           | 0.0131 | 0.0027 | <b>0.0009</b> | <b>0.0009</b> | 0.0024 | <b>0.0009</b> | 0.0013        | 0.0012 |
|     |                | 20 | 0.0025   | 0.0033           | 0.0120 | 0.0037 | <b>0.0014</b> | <b>0.0014</b> | 0.0060 | 0.0015        | 0.0031        | 0.0019 |
| M   | KLD            | 10 | 0.0174   | 0.0204           | 0.0479 | 0.0197 | 0.0077        | 0.0091        | 0.0298 | 0.0082        | <b>0.0046</b> | 0.0114 |
|     |                | 20 | 0.0418   | 0.0453           | 0.3564 | 0.0431 | 0.0192        | 0.0213        | 0.3125 | 0.0292        | <b>0.0084</b> | 0.0253 |
|     | $\text{JSD}^a$ | 10 | 0.0042   | 0.0050           | 0.0104 | 0.0049 | 0.0020        | 0.0022        | 0.0067 | 0.0020        | <b>0.0011</b> | 0.0028 |
|     |                | 20 | 0.0103   | 0.0112           | 0.0235 | 0.0106 | 0.0048        | 0.0052        | 0.0201 | 0.0072        | <b>0.0021</b> | 0.0063 |
|     | $\text{JSD}^g$ | 10 | 0.0042   | 0.0051           | 0.0124 | 0.0049 | 0.0020        | 0.0023        | 0.0076 | 0.0020        | <b>0.0011</b> | 0.0028 |
|     |                | 20 | 0.0104   | 0.0113           | 0.0297 | 0.0107 | 0.0049        | 0.0053        | 0.0249 | 0.0073        | <b>0.0021</b> | 0.0063 |

The detailed results are presented in Table 2 for the five methods, where we average over the ten forecast horizons. As the forecast horizon increases, the forecast errors generally become larger. This phenomenon reflects the increasing uncertainties associated with the model and estimated parameters. The  $\alpha$ -transformation method performs the best among the five methods for the female life-table death counts. In contrast, the MEM produces the most accurate forecasts for the male life-table death counts. The  $\alpha$ -transformation and ilr methods tend to produce more accurate forecasts than the eda method as it respects the non-negativity and summability constraints. The  $\alpha$ -transformation and ilr methods can also adapt to temporal changes in the age distribution of the life-table death counts over the years. Between the ilr and clr, they perform similarly in terms of point forecast accuracy. Between the two methods of selecting K, having a larger number of components not only improves model fitting but also improves point forecast accuracy.

For comparison, we also consider applying the Lee-Carter model to forecast age-specific mortality rates directly and then transforming the forecasts into the forecasted life-table death counts. Computationally, the conversion between rate and life-table death counts can be achieved via the *LifeTable* function in the MortalityLaws package (Pascariu, 2024). Since our objective is to model life-table death counts at ages 0, 1, ..., 109, 110+, it is necessary to smooth the original data via penalized regression spline with monotonic constraint. Computationally, the smoothing step can be achieved via the *smooth.demogdata* function in demography package (Hyndman, 2023). Although this comparison does not impact our recommendation, it presents an interesting research direction to explore various mortality instruments.

In Table 3, we present the point forecast results using a *rolling window* scheme. With the rolling-window scheme, the training sample considers the more recent data and the size of the training sample remains the same. For the female series, the  $\alpha$ -transformation, ilr and clr perform similarly, outperforming the forecasts obtained from the MEM. For the male series, the MEM provides the best point forecast accuracy. These findings are aligned with our previous findings using the expanding window scheme.

Table 3: For two different forecast horizons  $H = 10$  and  $20$ , we present a comparison of the point forecast accuracy using the rolling-window scheme, measured by the KLD,  $JSD^a$  and  $JSD^g$ , among the  $\alpha$  transformation, isometric log-ratio (ilr), and Euclidean data analysis (eda), using the holdout sample of the Australian female and male life-table death counts. We also include the clr method and the MEM of Pascariu et al. (2019) for comparison. The smallest errors among the methods are highlighted in bold for  $H = 10$  and  $20$ .

| Criterion | H  | $\alpha$      | K = 6 (Female) |        |               |        | $\alpha$ | K = 6 (Male) |        |        |               |
|-----------|----|---------------|----------------|--------|---------------|--------|----------|--------------|--------|--------|---------------|
|           |    |               | ilr            | eda    | clr           | MEM    |          | ilr          | eda    | clr    | MEM           |
| KLD       | 10 | <b>0.0033</b> | <b>0.0033</b>  | 0.0094 | <b>0.0033</b> | 0.0048 | 0.0079   | 0.0094       | 0.0287 | 0.0094 | <b>0.0043</b> |

Continued on next page

| Criterion        | H  | $\alpha$      | K = 6 (Female) |        |               |        | $\alpha$ | K = 6 (Male) |        |        |               |
|------------------|----|---------------|----------------|--------|---------------|--------|----------|--------------|--------|--------|---------------|
|                  |    |               | ilr            | eda    | clr           | MEM    |          | ilr          | eda    | clr    | MEM           |
|                  | 20 | <b>0.0050</b> | <b>0.0050</b>  | 0.0227 | <b>0.0050</b> | 0.0102 | 0.0203   | 0.0208       | 0.2694 | 0.0208 | <b>0.0075</b> |
| JSD <sup>a</sup> | 10 | <b>0.0008</b> | <b>0.0008</b>  | 0.0022 | <b>0.0008</b> | 0.0012 | 0.0020   | 0.0023       | 0.0064 | 0.0023 | <b>0.0010</b> |
|                  | 20 | <b>0.0012</b> | <b>0.0012</b>  | 0.0052 | <b>0.0012</b> | 0.0025 | 0.0049   | 0.0051       | 0.0184 | 0.0051 | <b>0.0018</b> |
| JSD <sup>g</sup> | 10 | <b>0.0008</b> | <b>0.0008</b>  | 0.0023 | <b>0.0008</b> | 0.0012 | 0.0020   | 0.0023       | 0.0072 | 0.0023 | <b>0.0011</b> |
|                  | 20 | <b>0.0012</b> | <b>0.0012</b>  | 0.0056 | <b>0.0012</b> | 0.0025 | 0.0050   | 0.0052       | 0.0229 | 0.0052 | <b>0.0019</b> |

#### 4.4 Interval forecast error measures

We consider the absolute difference between empirical and nominal coverage probabilities and the mean interval score of [Gneiting and Raftery \(2007\)](#) to evaluate the interval forecast accuracy. For each year in the forecasting period, the  $h$ -step-ahead prediction intervals are computed at the  $100(1 - \gamma)\%$  nominal coverage probability. For constructing  $100(1 - \gamma)\%$  prediction intervals, let us denote  $\hat{d}_{m+\xi,x}^{lb}$  and  $\hat{d}_{m+\xi,x}^{ub}$  as the lower and upper bounds, respectively. We compute the empirical coverage probability (ECP), defined as

$$ECP_{\gamma}(h) = 1 - \frac{1}{111 \times (H + 1 - h)} \sum_{\xi=h}^H \sum_{x=1}^{111} \left[ \mathbb{1}\{d_{m+\xi,x} < \hat{d}_{m+\xi,x}^{lb}\} + \mathbb{1}\{d_{m+\xi,x} > \hat{d}_{m+\xi,x}^{ub}\} \right].$$

While the ECP reveals over- or under-estimation of the nominal coverage probability, it is not an accuracy criterion due to the possible cancellation. As an alternative, the CPD is defined as

$$CPD_{\gamma}(h) = |ECP_{\gamma}(h) - (1 - \gamma)|.$$

The smaller the value of CPD is, the more accurate interval forecast accuracy the method produces. Despite the ECP and CPD being measures of interval forecast accuracy, neither consider the sharpness of the prediction intervals, i.e., the distance between the lower and upper bounds. To address this problem, [Gneiting and Raftery \(2007\)](#) introduce a scoring rule for the interval forecasts at time point  $d_{n+\xi,x}$ :

$$S_{\gamma,\xi}(\hat{d}_{m+\xi,x}^{lb}, \hat{d}_{m+\xi,x}^{ub}, d_{m+\xi,x}) = (\hat{d}_{m+\xi,x}^{ub} - \hat{d}_{m+\xi,x}^{lb}) + \frac{2}{\gamma} (\hat{d}_{m+\xi,x}^{lb} - d_{m+\xi,x}) \mathbb{1}\{d_{m+\xi,x} < \hat{d}_{m+\xi,x}^{lb}\} \\ + \frac{2}{\gamma} (d_{m+\xi,x} - \hat{d}_{m+\xi,x}^{ub}) \mathbb{1}\{d_{m+\xi,x} > \hat{d}_{m+\xi,x}^{ub}\}.$$

The interval score rewards a narrow prediction interval if and only if the true observation lies within the prediction interval. The optimal interval score is achieved when  $d_{m+\xi,x}$  lies between  $\hat{d}_{m+\xi,x}^{lb}$  and  $\hat{d}_{m+\xi,x}^{ub}$  and the distance between  $\hat{d}_{m+\xi,x}^{lb}$  and  $\hat{d}_{m+\xi,x}^{ub}$  is minimal for age  $x$ .

For different ages and years in the forecasting period, the mean interval score is defined by

$$S_\gamma(h) = \frac{1}{111 \times (H + 1 - h)} \sum_{\xi=h}^H \sum_{x=1}^{111} S_{\gamma,\xi}(\hat{d}_{m+\xi,x}^{lb}, \hat{d}_{m+\xi,x}^{ub}; d_{m+\xi,x}),$$

where  $S_{\gamma,\xi}(\hat{d}_{m+\xi,x}^{lb}, \hat{d}_{m+\xi,x}^{ub}; d_{m+\xi,x})$  denotes the interval score at the  $\xi^{\text{th}}$  curve in the validation or testing set.

For  $h = 1, 2, \dots, H$ , we compute the averaged CPD and interval scores:

$$\begin{aligned} \text{CPD}_\gamma &= \frac{1}{H} \sum_{h=1}^H \text{CPD}_\gamma(h), \\ S_\gamma &= \frac{1}{H} \sum_{h=1}^H S_\gamma(h). \end{aligned}$$

## 4.5 Interval forecast results

Using the *expanding window* approach, we also evaluate and compare the interval forecast accuracy of the  $\alpha$  transformation, ilr and eda, based on coverage probability difference (CPD) and  $S_\gamma$ . The detailed results are presented in Table 4 for the three methods. Between the  $\alpha$  transformation and ilr, the former is more flexible and can provide improved forecast accuracy subject to a finer tuning of the  $\alpha$  parameter. For the female data, the  $\alpha$  transformation generally produces the smallest CPD values and mean interval scores for  $\alpha = 0.05$  and  $0.2$ . For the male data, the ilr method generally produces the smallest CPD values and mean interval scores. This result further highlights the importance of selecting an optimal  $\alpha$  value and its difficulty associated with the relatively volatile male data. Because the eda ignores the compositional constraints, it shows the worst interval forecast accuracy. When the number of components is  $K = 6$ , it improves the interval forecast accuracy, especially for the eda method. We also computed the pointwise prediction intervals obtained from the MEM. Because each of the lower and upper bounds sums up to the radix, this constraint can result in cross-over. Hence, it produces inferior interval forecast accuracy. [Because of this cross-over effect, the conversion from mortality rate forecasts to forecasted life-table death count will also have inferior forecast accuracy.](#)

Table 4: Comparison of the interval forecast accuracy, as measured by the  $CPD_\gamma$  and  $S_\gamma$ , among the  $\alpha$  transformation, isometric log-ratio (ilr) and Euclidean data analysis (eda), using the holdout sample of the Australian female and male life-table death counts.

| Sex | Criterion    | H  | Eigenvalue ratio |        |        | K = 6  |        |        | MEM    |
|-----|--------------|----|------------------|--------|--------|--------|--------|--------|--------|
|     |              |    | alpha            | ilr    | eda    | alpha  | ilr    | eda    |        |
| F   | $CPD_{0.2}$  | 10 | 0.0486           | 0.0765 | 0.2652 | 0.0497 | 0.1065 | 0.0598 | 0.8171 |
|     |              | 20 | 0.0744           | 0.0969 | 0.1490 | 0.0553 | 0.1007 | 0.0857 | 0.8199 |
|     | $S_{0.2}$    | 10 | 0.0029           | 0.0034 | 0.0074 | 0.0022 | 0.0027 | 0.0034 | 0.0100 |
|     |              | 20 | 0.0033           | 0.0037 | 0.0069 | 0.0024 | 0.0032 | 0.0045 | 0.0149 |
|     | $CPD_{0.05}$ | 10 | 0.0218           | 0.0341 | 0.1165 | 0.0367 | 0.0367 | 0.0456 | 0.8702 |
|     |              | 20 | 0.0334           | 0.0394 | 0.0761 | 0.0308 | 0.0435 | 0.0433 | 0.8513 |
|     | $S_{0.05}$   | 10 | 0.0034           | 0.0046 | 0.0130 | 0.0036 | 0.0038 | 0.0060 | 0.0543 |
|     |              | 20 | 0.0042           | 0.0054 | 0.0113 | 0.0034 | 0.0047 | 0.0065 | 0.0828 |
| M   | $CPD_{0.2}$  | 10 | 0.1311           | 0.1099 | 0.2562 | 0.1028 | 0.0854 | 0.1187 | 0.8351 |
|     |              | 20 | 0.3417           | 0.1390 | 0.4503 | 0.1511 | 0.0620 | 0.2653 | 0.8443 |
|     | $S_{0.2}$    | 10 | 0.0053           | 0.0052 | 0.0068 | 0.0032 | 0.0030 | 0.0048 | 0.0096 |
|     |              | 20 | 0.0106           | 0.0089 | 0.0119 | 0.0052 | 0.0040 | 0.0090 | 0.0144 |
|     | $CPD_{0.05}$ | 10 | 0.0841           | 0.0421 | 0.2030 | 0.0418 | 0.0397 | 0.0612 | 0.9145 |
|     |              | 20 | 0.2008           | 0.1168 | 0.3350 | 0.0956 | 0.0353 | 0.1533 | 0.9239 |
|     | $S_{0.05}$   | 10 | 0.0066           | 0.0060 | 0.0128 | 0.0040 | 0.0039 | 0.0079 | 0.0512 |
|     |              | 20 | 0.0142           | 0.0136 | 0.0238 | 0.0086 | 0.0051 | 0.0167 | 0.0794 |

As a sensitivity analysis, we conduct the interval forecast accuracy comparison using the rolling window approach and the results are shown in Table 5. For modelling the female series, the alpha transformation produces the smallest interval forecast errors, measured by the CPD and interval score. However, for the male series, the ilr method is advantageous and is our recommended approach. A partial reason is that the selected  $\alpha$  parameter from the validation data set is sub-optimal for the testing data set. Between the eigenvalue ratio and  $K = 6$ , the latter one produces comparably smaller interval forecast errors.

Table 5: Comparison of the interval forecast accuracy, as measured by the  $CPD_{\gamma}$  and  $S_{\gamma}$ , among the  $\alpha$  transformation, isometric log-ratio (ilr) and Euclidean data analysis (eda), using the rolling window scheme.

| Sex | Criterion    | H  | Eigenvalue ratio |        |        | K = 6  |        |        | MEM    |
|-----|--------------|----|------------------|--------|--------|--------|--------|--------|--------|
|     |              |    | alpha            | ilr    | eda    | alpha  | ilr    | eda    |        |
| F   | $CPD_{0.2}$  | 10 | 0.0627           | 0.0785 | 0.2797 | 0.0453 | 0.0976 | 0.0667 | 0.8179 |
|     |              | 20 | 0.0910           | 0.0958 | 0.1752 | 0.0660 | 0.0757 | 0.0912 | 0.8277 |
|     | $S_{0.2}$    | 10 | 0.0029           | 0.0033 | 0.0073 | 0.0023 | 0.0026 | 0.0034 | 0.0092 |
|     |              | 20 | 0.0033           | 0.0036 | 0.0074 | 0.0025 | 0.0033 | 0.0045 | 0.0132 |
|     | $CPD_{0.05}$ | 10 | 0.0222           | 0.0319 | 0.1254 | 0.0367 | 0.0358 | 0.0450 | 0.8798 |
|     |              | 20 | 0.0325           | 0.0378 | 0.0870 | 0.0328 | 0.0396 | 0.0445 | 0.8833 |
|     | $S_{0.05}$   | 10 | 0.0037           | 0.0046 | 0.0127 | 0.0036 | 0.0038 | 0.0057 | 0.0498 |
|     |              | 20 | 0.0041           | 0.0052 | 0.0128 | 0.0035 | 0.0047 | 0.0067 | 0.0723 |
| M   | $CPD_{0.2}$  | 10 | 0.1655           | 0.1387 | 0.2441 | 0.1143 | 0.0954 | 0.0981 | 0.8307 |
|     |              | 20 | 0.3532           | 0.1642 | 0.4284 | 0.1600 | 0.0926 | 0.2144 | 0.8387 |
|     | $S_{0.2}$    | 10 | 0.0053           | 0.0053 | 0.0069 | 0.0033 | 0.0031 | 0.0048 | 0.0090 |
|     |              | 20 | 0.0093           | 0.0093 | 0.0151 | 0.0054 | 0.0043 | 0.0089 | 0.0130 |
|     | $CPD_{0.05}$ | 10 | 0.0718           | 0.0580 | 0.2155 | 0.0687 | 0.0446 | 0.0573 | 0.9152 |
|     |              | 20 | 0.2612           | 0.1302 | 0.3579 | 0.1162 | 0.0387 | 0.1388 | 0.9214 |
|     | $S_{0.05}$   | 10 | 0.0067           | 0.0064 | 0.0133 | 0.0039 | 0.0039 | 0.0080 | 0.0480 |
|     |              | 20 | 0.0156           | 0.0155 | 0.0313 | 0.0089 | 0.0053 | 0.0167 | 0.0715 |

## 5 Concluding comments

We consider the  $\alpha$  transformation to model and forecast age-specific life-table death counts in Australia. We evaluate the point and interval forecast accuracies among the  $\alpha$  transformation, isometric log-ratio analysis, and Euclidean data analysis for forecasting the age distribution of death counts. Based on the Kullback-Leibler and Jensen-Shannon divergences, the  $\alpha$  transformation is recommended for modelling the female data, while the MEM method is advocated for modelling

the male data. Subject to an optimal selection of  $\alpha$  parameter, the  $\alpha$  transformation is more flexible and can provide improved point forecast accuracy than the log-ratio analysis. Similar to the Box-Cox transformation, the  $\alpha$  transformation is a natural way of handling zeros in compositional data analysis. In terms of interval forecast accuracy, the  $\alpha$  transformation also provides the smallest interval forecast errors, measured by CPD and mean interval score for the female data. For the more volatile male data, the ilr method is advocated. For reproducibility, all the code is available on GitHub: [https://github.com/hanshang/alpha\\_transformation](https://github.com/hanshang/alpha_transformation).

There are a few ways in which this paper could be extended, and we briefly discuss five:

- 1) In the demographic literature, there are other models for forecasting the age distribution of death, such as the segmented transformation age-at-death distribution from [Basellini and Camarda \(2019\)](#).
- 2) A robust  $\alpha$  transformation may be proposed in the presence of outlying years.
- 3) One can extend  $\alpha$  transformation to jointly model and forecast the age distribution of death counts for multiple populations.
- 4) We considered period life table, which is a useful measure of mortality rates experienced over a given period. As an alternative, one can consider cohort life-table death counts. A cohort life table displays the probability of a person from a given cohort dying at each age over the course of their lifetime. In this setting, a cohort refers to a group of individuals with the same year of birth.
- 5) One could consider a probabilistic framework that accounts for compositional structure of the data, such as the Dirichlet distribution (see, e.g., [Tsagris and Stewart, 2018](#); [Tang et al., 2022](#); [Graziani and Nigri, 2023](#)). The use of the Dirichlet distribution is to avoid any transformation and directly model life-table death counts that are easier to interpret.

## Acknowledgements

The authors would like to acknowledge the insightful comments from two reviewers that led to a much-improved paper, and fruitful discussions with Professor Denis Allard, who introduced us to  $\alpha$  transformation at the KAUST2022 workshop on statistics. The authors thank helpful comments received from the participants at the 2023 SOA Living to 100 symposium, the Population Association of New Zealand conference held at the University of Auckland, and a seminar at the Research School of Finance, Actuarial Studies and Statistics at the Australian National University. The first author acknowledges the fundings from the Australian Research Council under Discovery Project Grant DP230102250 and Future Fellow FT240100338.

## Declarations

**Disclosure statement:** The authors report there are no competing interests to declare.

**Data availability statement:** The datasets are available from [https://github.com/hanshang/alpha\\_transformation](https://github.com/hanshang/alpha_transformation).

## References

- Aburto, J. M., Basellini, U., Baudisch, A., and Villavicencio, F. (2022). Drewnowski's index to measure lifespan variation: Revisiting the Gini coefficient of the life table. *Theoretical Population Biology*, 148:1–10.
- 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. and Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. *Biometrika*, 67(2):261–272.
- Basellini, U. and Camarda, C. G. (2019). Modelling and forecasting adult age-at-death distributions. *Population Studies*, 73(1):119–138.
- Basellini, U., Camarda, C. G., and Booth, H. (2023). Thirty years on: A review of the Lee-Carter method for forecasting mortality. *International Journal of Forecasting*, 39(3):1033–1049.
- Basellini, U., Kjaergaard, S., and Camarda, C. G. (2020). An age-at-death distribution approach to forecast cohort mortality. *Insurance: Mathematics and Economics*, 91:129–143.
- Baxter, M. J. (1995). Standardization and transformation in principal component analysis, with applications to archaeometry. *Journal of the Royal Statistical Society: Series C*, 44(4):513–527.
- Baxter, M. J. (2001). Statistical modelling of artefact compositional data. *Archaeometry*, 43(1):131–147.
- Baxter, M. J., Beardah, C. C., Cool, H. E. M., and Jackson, C. M. (2005). Compositional data analysis of some alkaline glasses. *Mathematical Geology*, 37(2):183–196.
- Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., and Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37:527–566.

- Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22:547–581.
- Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(1-2):3–43.
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). Measuring the longevity risk in mortality projections. *Bulletin of the Swiss Association of Actuaries*, 2:105–130.
- Cairns, A. J. G., Blake, D., and Dowd, K. (2008). Modeling and management of mortality risk: A review. *Scandinavian Actuarial Journal*, 2-3:79–113.
- Canudas-Romo, V. (2010). Three measures of longevity: Time trends and record values. *Demography*, 47(2):299–312.
- Dellaportas, P., Smith, A. F., and Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of the Royal Statistical Society: Series A*, 164(2):275–291.
- Denuit, M., Devolder, P., and 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.
- Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. John Wiley, Chichester.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35:279–300.
- Fry, J. M., Fry, T. R. L., and McLaren, K. R. (2000). Compositional data analysis and zeros in micro data. *Applied Economics*, 32(8):953–959.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association: Review Article*, 102(477):359–378.
- Graziani, R. and Nigri, A. (2023). An age-period-cohort model in a Dirichlet framework: A coherent causes of death estimation. Working paper, SocArXiv. URL: <https://osf.io/preprints/socarxiv/856yw>.
- Greenacre, M. (2021). Compositional data analysis. *Annual Review of Statistics and its Application*, 8:271–299.
- Greenacre, M. and Grunsky, E. (2019). The isometric logratio transformation in compositional data analysis: A practical evaluation. Working paper 1627, Universitat Pompeu Fabra. URL: <https://econ-papers.upf.edu/papers/1627.pdf>.

- Heligman, L. and Pollard, J. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1):49–80.
- Human Mortality Database (2023). *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. Accessed at October 5, 2023. URL: <http://www.mortality.org>.
- Hurvich, C. M. and Tsai, C.-L. (1993). A corrected akaike information criterion for vector autoregressive model selection. *Journal of Time Series Analysis*, 14(3):271–279.
- Hyndman, R. (2023). *demography: Forecasting Mortality, Fertility, Migration and Population Data*. R package version 2.0.
- Hyndman, R. J., Booth, H., and Yasmeeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1):261–283.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3).
- Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series (with discussion). *Journal of the Korean Statistical Society*, 38(3):199–221.
- Hyndman, R. J., Zeng, Y., and Shang, H. L. (2021). Foreacsting the old-age dependency ratio to determine a sustainable pension age. *Australian and New Zealand Journal of Statistics*, 63(2):241–256.
- Kokoszka, P., Miao, H., Petersen, A., and 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. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- Lancaster, H. O. (1965). The Helmert matrices. *The American Mathematical Monthly*, 72(1):4–12.
- Lee, R. D. and Carter, L. (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association: Applications & Case Studies*, 87(419):659–671.
- Li, D., Robinson, P. M., and Shang, H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association: Theory and Methods*, 115(530):957–971.

- Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, 35(3):253–278.
- Mazzucco, S., Scarpa, B., and Zanotto, L. (2018). A mortality model based on a mixture distribution function. *Population Studies*, 72(2):1–10.
- Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data. In *European Population Conference*. European Association for Population Studies. URL: <http://hdl.handle.net/10256/742>.
- Pascariu, M. D. (2024). *MortalityLaws: Parametric Mortality Models, Life Tables and HMD*. R package version 2.1.0.
- Pascariu, M. D., Lenart, A., and Canudas-Romo, V. (2019). The maximum entropy mortality model: Forecasting mortality using statistical moments. *Scandinavian Actuarial Journal*, 2019(8):661–685.
- Pollard, J. H. (1987). Projection of age-specific mortality rates. In *Population Bulletin of the United Nations*, 21/22, pages 55–69. Population Commission. URL: <https://digitallibrary.un.org/record/42274?ln=en>.
- 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., Tsaigris, M. T., and Welsh, A. H. (2015). 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. and 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., Booth, H., and Hyndman, R. J. (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research*, 25(5):173–214.
- Shang, H. L. and Haberman, S. (2020). Forecasting age distribution of death counts: An application to annuity pricing. *Annals of Actuarial Science*, 14(1):150–169.
- Shang, H. L., Haberman, S., and Xu, R. (2022). Multi-population modelling and forecasting life-table death counts. *Insurance: Mathematics and Economics*, 106:239–253.

- Shang, H. L., Smith, P. W. F., Bijak, J., and Wiśniowski, A. (2016). A multilevel functional data method for forecasting population, with an application to the United Kingdom. *International Journal of Forecasting*, 32(3):629–649.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Labs Technical Journal*, 27(3):379–423.
- Tang, M.-L., Wu, Q., Yang, S., and Tian, G.-L. (2022). Dirichlet composition distribution for compositional data with zero components: An application to fluorescence in situ hybridization (fish) detection of chromosome. *Biometrical Journal*, 64(4):714–732.
- Tsagris, M., Preston, S., and Wood, A. T. A. (2016). Improved classification for compositional data using the alpha-transformation. *Journal of Classification*, 33:243–261.
- Tsagris, M. and Stewart, C. (2018). A Dirichlet regression model for compositional data with zeros. *Lobachevskii Journal of Mathematics*, 39:398–412.
- Tsagris, M. and Stewart, C. (2020). A folded model for compositional data analysis. *Australian & New Zealand Journal of Statistics*, 62(2):249–277.
- van Raalte, A. A. and Caswell, H. (2013). Perturbation analysis of indices of lifespan variability. *Demography*, 50(5):1615–1640.
- van Raalte, A. A., Martikainen, P., and Myrskylä, M. (2014). Lifespan variation by occupational class: Compression or stagnation over time? *Demography*, 51:73–95.
- Vaupel, J. W., Zhang, Z., and van Raalte, A. A. (2011). Life expectancy and disparity: An international comparison of life table data. *BMJ Open*, 1(1):e000128.
- Wilmoth, J. R. and Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography*, 36(4):475–495.