Forecasting age distribution of life-table death counts via α-transformation

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

(Joint work with Steven Haberman, Bayes Business School)
City, University of London

August 24, 2023
Presentation at the NZ Population Conference 2023

Life-table death counts

1. Life-table death counts are non-negative & sum to a radix for a year

Life-table death counts

1. Life-table death counts are non-negative & sum to a radix for a year
2. Life-table death counts often have *zero values*

Life-table death counts

1. Life-table death counts are non-negative & sum to a radix for a year
2. Life-table death counts often have *zero values*
3. Compositional power α-transformation to model & forecast life-table death counts

Compositional data

1. Life-table death counts share resemblance with PDF

Compositional data

1. Life-table death counts share resemblance with PDF
2. Compositional data are a random vector of D compositions $[d_1, \dots, d_D]$ with non-negative values whose sum is a constant

Compositional data

1. Life-table death counts share resemblance with PDF
2. Compositional data are a random vector of D compositions $[d_1, \dots, d_D]$ with non-negative values whose sum is a constant
3. Sample space of compositional data is a simplex

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

Compositional data

1. Life-table death counts share resemblance with PDF
2. Compositional data are a random vector of D compositions $[d_1, \dots, d_D]$ with non-negative values whose sum is a constant
3. Sample space of compositional data is a simplex

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

4. Oeppen (2008) and Bergeron-Boucher et al. (2017) combine CoDa with PCA to model & forecast life-table death counts

clr transformation

1. Transform compositional data from simplex to Euclidean space

$$\boldsymbol{s} = \{s_x\}_{x=1,\dots,D} = \ln\left\{\frac{d_x}{\left(\prod_{x=1}^{D} d_x\right)^{1/D}}\right\}$$

clr transformation

1. Transform compositional data from simplex to Euclidean space

$$\boldsymbol{s} = \{s_x\}_{x=1,\dots,D} = \ln\left\{\frac{d_x}{\left(\prod_{x=1}^{D} d_x\right)^{1/D}}\right\}$$

2. Extending clr, isometric log-ratio (ilr) is

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

where $\boldsymbol{H}$: $(D-1)\times D$: Helmert matrix

What's Helmert matrix

1. Helmert matrices are used to describe a coordinate transformation

What's Helmert matrix

1. Helmert matrices are used to describe a coordinate transformation
2. Helmert matrix of order D is a square matrix

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

What's Helmert matrix

1. Helmert matrices are used to describe a coordinate transformation
2. Helmert matrix of order D is a square matrix

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

3. To remove any redundant dimension

Issues with log-ratio transformation

1. No guarantee transformed data are multivariate normal distributed

Issues with log-ratio transformation

1. No guarantee transformed data are multivariate normal distributed
2. Cannot be applied to data containing zeros

Australian data

1. Australian age & sex-specific life-table death counts from 1921 to 2020

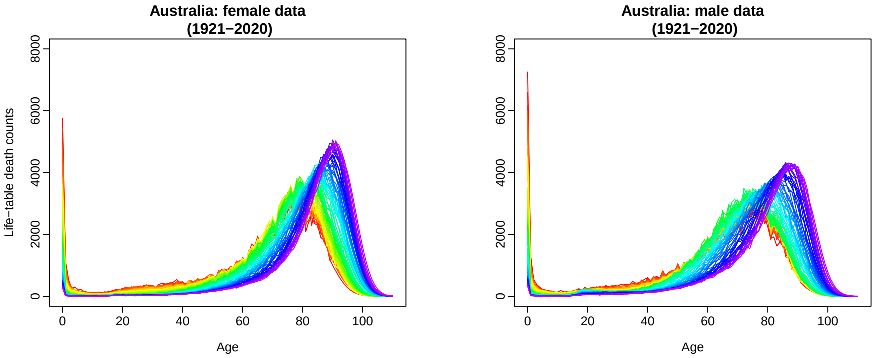
Australian data

1. Australian age & sex-specific life-table death counts from 1921 to 2020
2. Life-table radix is 100,000 at age 0 for each year

Australian data

1. Australian age & sex-specific life-table death counts from 1921 to 2020
2. Life-table radix is 100,000 at age 0 for each year
3. 111 ages, ages $0, 1, \dots, 109, 110+$

Rainbow plot



1. Decreasing trend in infant mortality
2. A negatively skewed distribution for life-table death counts
3. Shift of the distribution symbolises longevity risk

α transformation

1. α-transformation of a compositional vector $\boldsymbol{d}_t \in \mathcal{S}^D$

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

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

α transformation

1. α-transformation of a compositional vector $\boldsymbol{d}_t \in \mathcal{S}^D$

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

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

- $\alpha \in [0,1]$; $\alpha = 0$, becomes ilr; $\alpha = 1$, becomes Euclidean data analysis

α transformation

1. α-transformation of a compositional vector $\boldsymbol{d}_t \in \mathcal{S}^D$

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

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

- $\alpha \in [0,1]$; $\alpha = 0$, becomes ilr; $\alpha = 1$, becomes Euclidean data analysis
- $\mathbb{1}_D$: D-dimensional vector of ones

α transformation

1. α-transformation of a compositional vector $\boldsymbol{d}_t \in \mathcal{S}^D$

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

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

- $\alpha \in [0, 1]$; $\alpha = 0$, becomes ilr; $\alpha = 1$, becomes Euclidean data analysis
- $\mathbb{1}_D$: D-dimensional vector of ones
- $\boldsymbol{H}$: $(D-1) \times D$ Helmert matrix

α transformation

1. α-transformation of a compositional vector $\boldsymbol{d}_t \in \mathcal{S}^D$

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

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

   - $\alpha \in [0,1]$; $\alpha = 0$, becomes ilr; $\alpha = 1$, becomes Euclidean data analysis
   - $\mathbb{1}_D$: D-dimensional vector of ones
   - $\boldsymbol{H}$: $(D-1) \times D$ Helmert matrix

2. α transformation is invertible

$$\boldsymbol{v}_t^\alpha = D\boldsymbol{d}_t^\alpha = \alpha \times \boldsymbol{H}^\top \boldsymbol{z}_t^\alpha + \mathbb{1}_D$$

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

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^{\alpha} = (\boldsymbol{z}_1^{\alpha}, \dots, \boldsymbol{z}_n^{\alpha})$.

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^{\alpha} = (\boldsymbol{z}_1^{\alpha}, \dots, \boldsymbol{z}_n^{\alpha})$.
2. PCA to decompose $\boldsymbol{z}^{\alpha}$ into

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

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^{\alpha} = (\boldsymbol{z}_1^{\alpha}, \dots, \boldsymbol{z}_n^{\alpha})$.
2. PCA to decompose $\boldsymbol{z}^{\alpha}$ into

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

- $\bar{\boldsymbol{z}}^{\alpha} = \frac{1}{n}\sum_{t=1}^{n} z_t^{\alpha}$: mean term

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^{\alpha} = (\boldsymbol{z}_1^{\alpha}, \dots, \boldsymbol{z}_n^{\alpha})$.
2. PCA to decompose $\boldsymbol{z}^{\alpha}$ into

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

- $\bar{\boldsymbol{z}}^{\alpha} = \frac{1}{n} \sum_{t=1}^{n} z_t^{\alpha}$: mean term
- $\widehat{\beta}_{t,k}$: k^{th} set of estimated principal component scores for time t

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^\alpha = (\boldsymbol{z}_1^\alpha, \dots, \boldsymbol{z}_n^\alpha)$.
2. PCA to decompose $\boldsymbol{z}^\alpha$ into

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

- $\bar{\boldsymbol{z}}^\alpha = \frac{1}{n}\sum_{t=1}^{n} z_t^\alpha$: mean term
- $\widehat{\beta}_{t,k}$: k^{th} set of estimated principal component scores for time t
- $\widehat{\phi}_k$: k^{th} set of estimated principal component

Modeling & forecasting a time series of curves

1. A time series of unconstrained matrix $\boldsymbol{z}^{\alpha} = (\boldsymbol{z}_1^{\alpha}, \dots, \boldsymbol{z}_n^{\alpha})$.
2. PCA to decompose $\boldsymbol{z}^{\alpha}$ into

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

- $\bar{\boldsymbol{z}}^{\alpha} = \frac{1}{n} \sum_{t=1}^{n} z_t^{\alpha}$: mean term
- $\widehat{\beta}_{t,k}$: k^{th} set of estimated principal component scores for time t
- $\widehat{\phi}_k$: k^{th} set of estimated principal component
- $\boldsymbol{e}_t$: model error term

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \mathop{\arg\min}_{1\leq k\leq K_{\max}} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right\} + \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right\} \right\}$$

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \underset{1\leq k\leq K_{\max}}{\arg\min}\left\{\frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k}\times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1}\geq\theta\right\}+\mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1}<\theta\right\}\right\}$$

- $\mathbb{1}\{\cdot\}$: binary indicator

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \underset{1\leq k\leq K_{\max}}{\arg\min} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right\} + \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right\} \right\}$$

- $\mathbb{1}\{\cdot\}$: binary indicator
- $\theta = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \mathop{\arg\min}_{1\leq k\leq K_{\max}} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right\} + \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right\} \right\}$$

- $\mathbb{1}\{\cdot\}$: binary indicator
- $\theta = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$
- $K_{\max} = \#\{k|\widehat{\lambda}_k \geq \frac{1}{n}\sum_{k=1}^{n}\widehat{\lambda}_k, k \geq 1\}$

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \underset{1\leq k\leq K_{\max}}{\arg\min} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right\} + \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right\} \right\}$$

   - $\mathbb{1}\{\cdot\}$: binary indicator
   - $\theta = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$
   - $K_{\max} = \#\{k|\widehat{\lambda}_k \geq \frac{1}{n}\sum_{k=1}^{n}\widehat{\lambda}_k, k \geq 1\}$

2. Fit a univariate time series forecasting method, such as ARIMA, to obtain h-step-ahead forecast

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

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \underset{1\leq k\leq K_{\max}}{\arg\min}\left\{\frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k}\times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1}\geq\theta\right\}+\mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1}<\theta\right\}\right\}$$

   - $\mathbb{1}\{\cdot\}$: binary indicator
   - $\theta = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$
   - $K_{\max} = \#\{k|\widehat{\lambda}_k \geq \frac{1}{n}\sum_{k=1}^{n}\widehat{\lambda}_k, k\geq 1\}$

2. Fit a univariate time series forecasting method, such as ARIMA, to obtain h-step-ahead forecast

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

   - $\widehat{\boldsymbol{\Phi}} = (\widehat{\Phi}_1, \dots, \widehat{\Phi}_K)^{\top}$

Eigenvalue ratio criterion

1. Determine optimal number of components

$$K = \mathop{\arg\min}_{1\leq k\leq K_{\max}} \left\{ \frac{\widehat{\lambda}_{k+1}}{\widehat{\lambda}_k} \times \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} \geq \theta\right\} + \mathbb{1}\left\{\frac{\widehat{\lambda}_k}{\widehat{\lambda}_1} < \theta\right\} \right\}$$

- $\mathbb{1}\{\cdot\}$: binary indicator
- $\theta = \frac{1}{\ln[\max(\widehat{\lambda}_1, n)]}$
- $K_{\max} = \#\{k|\widehat{\lambda}_k \geq \frac{1}{n}\sum_{k=1}^{n}\widehat{\lambda}_k, k \geq 1\}$

2. Fit a univariate time series forecasting method, such as ARIMA, to obtain h-step-ahead forecast

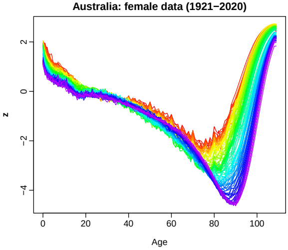
$$\widehat{\boldsymbol{z}}^{\alpha}_{n+h|n} = E[\boldsymbol{z}^{\alpha}_{n+h}|\widehat{\boldsymbol{\Phi}}, \boldsymbol{z}^{\alpha}] = \overline{z}^{\alpha} + \sum_{k=1}^{K} \widehat{\beta}_{n+h|n,k}\widehat{\phi}_k$$

- $\widehat{\boldsymbol{\Phi}} = (\widehat{\Phi}_1, \dots, \widehat{\Phi}_K)^\top$
- $\widehat{\beta}_{n+h|n,k}$: h-step-ahead forecast of k^{th} principal component scores

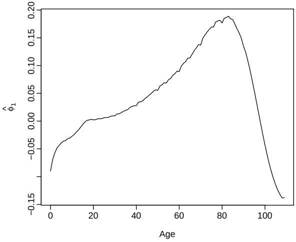
Selection of α parameter

Split data into a training set, a validation set, a testing set

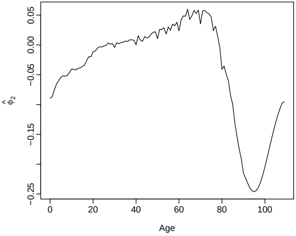
$1:(n-20)$	$(n-19):(n-10)$	$(n-9):n$
Training	Validation	Testing



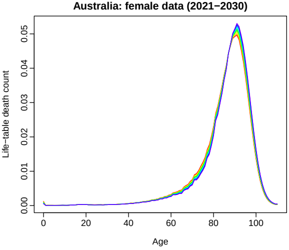
(a) *Unconstrained female data*



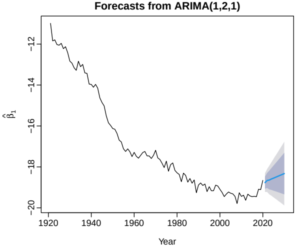
(b) *First principal component*



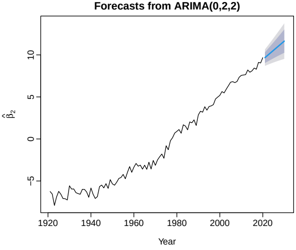
(c) *Second principal component*



(d) *Forecast LTDC*



(e) *Forecast first set of scores*



(f) *Forecast second set of scores*

Parameter estimation

1. Via eigenvalue ratio criterion, $K = 2$

Parameter estimation

1. Via eigenvalue ratio criterion, $K = 2$
2. Optimal ARIMA model is ARIMA$(1, 2, 1)$ using AIC$_c$

Parameter estimation

1. Via eigenvalue ratio criterion, $K = 2$
2. Optimal ARIMA model is ARIMA$(1, 2, 1)$ using AIC$_c$
3. Multiplying forecast scores with estimated functional principal components to obtain forecast curves

Goodness-of-fit (R^2 criterion)

1. In-sample goodness-of-fit via an R^2 criterion

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

Goodness-of-fit (R^2 criterion)

1. In-sample goodness-of-fit via an R^2 criterion

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

- $d_{t,x}$: observed age-specific life-table death counts

Goodness-of-fit (R^2 criterion)

1. In-sample goodness-of-fit via an R^2 criterion

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

- $d_{t,x}$: observed age-specific life-table death counts
- $\widehat{d}_{t,x}$: fitted value

Goodness-of-fit (R^2 criterion)

1. In-sample goodness-of-fit via an R^2 criterion

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

   - $d_{t,x}$: observed age-specific life-table death counts
   - $\widehat{d}_{t,x}$: fitted value

2. Compared to ilr,

Method	Female	Male
ilr	0.9953	0.9911
α transformation	0.9968	0.9915

Point forecast error measure

1. Kullback-Leibler divergence is symmetric & non-negative:

$$\begin{aligned}\mathrm{KLD}(h) &= D_{\mathsf{KL}}\left(\boldsymbol{d}_{m+\xi}||\widehat{\boldsymbol{d}}_{m+\xi}\right) + D_{\mathsf{KL}}\left(\widehat{\boldsymbol{d}}_{m+\xi}||\boldsymbol{d}_{m+\xi}\right) \\ &= \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{x=1}^{111} d_{m+\xi,x}\cdot\left(\ln d_{m+\xi,x} - \ln \widehat{d}_{m+\xi,x}\right) \\ &\quad + \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{x=1}^{111} \widehat{d}_{m+\xi,x}\cdot\left(\ln \widehat{d}_{m+\xi,x} - \ln d_{m+\xi,x}\right)\end{aligned}$$

m: end of training or validation set

Point forecast error measure

1. Kullback-Leibler divergence is symmetric & non-negative:

$$\text{KLD}(h) = D_{\text{KL}}\left(\boldsymbol{d}_{m+\xi}||\widehat{\boldsymbol{d}}_{m+\xi}\right) + D_{\text{KL}}\left(\widehat{\boldsymbol{d}}_{m+\xi}||\boldsymbol{d}_{m+\xi}\right)$$
$$= \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{x=1}^{111} d_{m+\xi,x}\cdot\left(\ln d_{m+\xi,x} - \ln \widehat{d}_{m+\xi,x}\right)$$
$$+ \frac{1}{111\times(11-h)}\sum_{\xi=h}^{10}\sum_{x=1}^{111} \widehat{d}_{m+\xi,x}\cdot\left(\ln \widehat{d}_{m+\xi,x} - \ln d_{m+\xi,x}\right)$$

m: end of training or validation set

2. Jensen-Shannon divergence:

$$\text{JSD}(h) = \frac{1}{2}D_{\text{KL}}\left(\boldsymbol{d}_{m+\xi}||\boldsymbol{\delta}_{n+\xi}\right) + \frac{1}{2}D_{\text{KL}}\left(\widehat{\boldsymbol{d}}_{m+\xi}||\boldsymbol{\delta}_{n+\xi}\right)$$

where $\boldsymbol{\delta}_{m+\xi}$ measures a common quantity between $\boldsymbol{d}_{m+\xi}$ & $\widehat{\boldsymbol{d}}_{m+\xi}$;

$$\boldsymbol{\delta}_{m+\xi} = \sqrt{\boldsymbol{d}_{m+\xi}\widehat{\boldsymbol{d}}_{m+\xi}}$$

KLD results

	Female			Male		
h	α-transform	ilr	eda	α-transform	ilr	eda
1	0.0038	0.0078	0.0828	0.0083	0.0107	0.0064
2	0.0040	0.0082	0.0842	0.0106	0.0121	0.0102
3	0.0047	0.0087	0.0853	0.0109	0.0137	0.0175
4	0.0051	0.0087	0.0868	0.0118	0.0151	0.0268
5	0.0054	0.0093	0.0889	0.0152	0.0172	0.0376
6	0.0060	0.0100	0.0928	0.0174	0.0197	0.0511
7	0.0063	0.0101	0.0959	0.0179	0.0230	0.0555
8	0.0082	0.0116	0.0966	0.0246	0.0277	0.0548
9	0.0083	0.0115	0.1043	0.0266	0.0298	0.0786
10	0.0107	0.0132	0.1076	0.0309	0.0345	0.1407
Mean	0.0062	0.0099	0.0925	0.0174	0.0204	0.0479

JSD(geo)

	Female			Male		
h	α-transform	ilr	eda	α-transform	ilr	eda
1	0.0009	0.0019	0.0099	0.0021	0.0027	0.0015
2	0.0010	0.0020	0.0104	0.0024	0.0030	0.0025
3	0.0012	0.0022	0.0108	0.0027	0.0034	0.0044
4	0.0015	0.0022	0.0115	0.0029	0.0038	0.0069
5	0.0013	0.0023	0.0123	0.0033	0.0043	0.0096
6	0.0016	0.0025	0.0133	0.0038	0.0049	0.0131
7	0.0014	0.0025	0.0141	0.0045	0.0057	0.0143
8	0.0024	0.0029	0.0145	0.0061	0.0069	0.0140
9	0.0020	0.0028	0.0165	0.0066	0.0074	0.0203
10	0.0029	0.0033	0.0176	0.0077	0.0086	0.0375
Mean	0.0016	0.0025	0.0131	0.0042	0.0051	0.0124

Interval score

Scoring rule for interval forecasts at time point $d_{n+\xi,x}$:

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

- γ: level of significance, customarily $\gamma = 0.2$
- For different ages & years in the forecasting period

$$\overline{S}_{\gamma}(h) = \frac{1}{111 \times (11-h)} \sum_{\xi=h}^{10} \sum_{x=1}^{111} S_{\gamma,\xi}\left(\widehat{d}^{\text{lb}}_{m+\xi,x}, \widehat{d}^{\text{ub}}_{m+\xi,x}; d_{m+\xi,x}\right)$$

CPD=|Nominal coverage - Empirical coverage|

	Female			Male		
h	α-transform	ilr	eda	α-transform	ilr	eda
1	0.0623	0.0759	0.2297	0.0427	0.0759	0.1258
2	0.0655	0.0859	0.2284	0.0819	0.0939	0.1205
3	0.0723	0.0869	0.2189	0.0597	0.0649	0.1315
4	0.0855	0.0958	0.2389	0.0432	0.1060	0.1295
5	0.0934	0.1339	0.2279	0.0357	0.1009	0.1318
6	0.0901	0.1369	0.2432	0.0811	0.1369	0.1445
7	0.0559	0.1640	0.2437	0.1302	0.1392	0.1518
8	0.1129	0.1820	0.2264	0.0829	0.1249	0.2006
9	0.0559	0.1730	0.2279	0.0829	0.1505	0.2910
10	0.1820	0.1820	0.2144	0.1189	0.1369	0.5658
Mean	**0.0876**	0.1316	0.2300	**0.0759**	0.1130	0.1993

Mean interval score ($\overline{S}_\gamma$)

	Female			Male		
h	α-transform	ilr	eda	α-transform	ilr	eda
1	0.0025	0.0031	0.0051	0.0033	0.0036	0.0024
2	0.0025	0.0031	0.0055	0.0035	0.0036	0.0029
3	0.0028	0.0032	0.0059	0.0037	0.0037	0.0036
4	0.0027	0.0031	0.0063	0.0035	0.0036	0.0041
5	0.0026	0.0032	0.0068	0.0035	0.0036	0.0047
6	0.0029	0.0035	0.0074	0.0037	0.0037	0.0055
7	0.0025	0.0037	0.0079	0.0038	0.0039	0.0064
8	0.0030	0.0036	0.0089	0.0039	0.0040	0.0081
9	0.0028	0.0039	0.0093	0.0043	0.0043	0.0098
10	0.0029	0.0042	0.0103	0.0041	0.0041	0.0137
Mean	**0.0027**	0.0035	0.0073	**0.0037**	0.0038	0.0061

Future research

1. A robust α transformation

Future research

1. A robust α transformation
2. Extend α transformation to model & forecast life-table death counts for multiple populations

Future research

1. A robust α transformation
2. Extend α transformation to model & forecast life-table death counts for multiple populations
3. Consider cohort life-table death counts for specific groups of individuals

Thank you

Paper: `https://arxiv.org/abs/2305.19749`