# Reconciling forecasts of age distribution of death counts: An application to annuity pricing

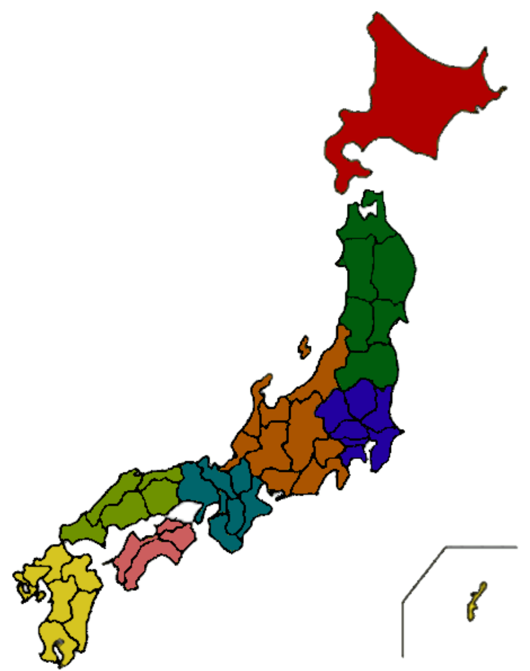
Han Lin Shang
Research School of Finance, Actuarial Studies and Statistics
Australian National University

Steven Haberman
Cass Business School
City, University of London

July 27, 2017

# Japanese age-specific lifetable

1. Japanese national and sub-national age-specific life-table death counts from 1975 to 2014 from *Japanese Mortality Database*
2. Period life-table radix is fixed at 100,000 at age 0 for each year group
3. 8 five-year year groups, 1975-1979, 1980-1984, $\cdots$, 2010-2014
4. 24 age groups, age 0, 1-4, 5-9, 10-14, $\cdots$, 105-109, 110+
5. Due to zero counts for age 110+ for some years, merge this age group with age group 105-109



# Japanese group structure

Have national and sub-national mortality rates, data structure is displayed below where each row denotes a level of disaggregation

| Group level | Number of series |
|---|---|
| Japan | 1 |
| Sex | 2 |
| Region | 8 |
| Region × Sex | 16 |
| Prefecture | 47 |
| Prefecture × Sex | 94 |
| Total | 168 |

# Compositional data-analytic approach

1. Compositional data are defined as a random vector of $K$ positive components $\boldsymbol{D} = [d_1, \dots, d_K]$ with strictly positive values whose sum is a given constant

# Compositional data-analytic approach

1. Compositional data are defined as a random vector of $K$ positive components $\boldsymbol{D} = [d_1, \dots, d_K]$ with strictly positive values whose sum is a given constant
2. Sample space of compositional data is the simplex

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

where $c$ is a fixed constant (such as, radix in period life table), $^\top$ denote vector transpose, simplex sample space is $K-1$ dimensional subset of $R^{K-1}$

# CoDa in action

1. Begin from a data matrix $\boldsymbol{D}$ of size $n \times K$ of life-table deaths $(d_{t,x})$ with $n$ rows representing the number of years and $K$ columns representing the age $x$. Sum of each row adds up to life-table radix, such as 100,000

# CoDa in action

1. Begin from a data matrix $\boldsymbol{D}$ of size $n \times K$ of life-table deaths $(d_{t,x})$ with $n$ rows representing the number of years and $K$ columns representing the age $x$. Sum of each row adds up to life-table radix, such as 100,000
2. Compute geometric mean at each age, given by

$$\alpha_x = \exp^{\frac{1}{n}\sum_{t=1}^{n}\ln(d_{t,x})}, \qquad x = 1, \dots, K$$

For a given year $t$, divide $(d_{t,1}, \dots, d_{t,K})$ by corresponding geometric means $(\alpha_1, \dots, \alpha_K)$,

$$C\left[\frac{d_{t,1}}{\alpha_1}, \frac{d_{t,2}}{\alpha_2}, \dots, \frac{d_{t,K}}{\alpha_K}\right]$$

# CoDa in action

$C[\cdot]$ represents a closure operation, performing standardization

$$f_{t,x} = \frac{\frac{d_{t,x}}{\alpha_x}}{\frac{d_{t,1}}{\alpha_1} + \frac{d_{t,2}}{\alpha_2} + \cdots + \frac{d_{t,K}}{\alpha_K}}, \qquad x = 1, \dots, K$$

where $f_{t,x}$ is a non-negative value

# CoDa in action

3. Log-ratio transformation: Aitchison (1982, 1986) showed that compositional data are represented in a restricted space where components can only vary between 0 and positive constant, proposed centered log-ratio transformation

$$h_{t,x} = \ln\left(\frac{f_{t,x}}{g_t}\right)$$

where $g_t$ are the geometric means over age at time $t$

$$g_t = \exp^{\frac{1}{K}\sum_{x=1}^{K}\ln(f_{t,x})}.$$

Transformed data matrix is $\boldsymbol{H}$ with elements $h_{t,x} \in R$ real-valued ☺

# CoDa in action

4. Principal component analysis: applied to the matrix $\boldsymbol{H}_x = \{h_{t,1}, \dots, h_{t,K}\}$ to obtain the estimated principal components and their associated scores,

$$h_{t,x} = \sum_{\ell=1}^{\min(n,K)} \beta_{t,\ell}\phi_{\ell,x} \approx \sum_{\ell=1}^{L} \beta_{t,\ell}\phi_{\ell,x}$$

# CoDa in action

4. Principal component analysis: applied to the matrix $\boldsymbol{H}_x = \{h_{t,1},\dots,h_{t,K}\}$ to obtain the estimated principal components and their associated scores,

$$h_{t,x} = \sum_{\ell=1}^{\min(n,K)} \beta_{t,\ell}\phi_{\ell,x} \approx \sum_{\ell=1}^{L} \beta_{t,\ell}\phi_{\ell,x}$$

- $\{\phi_{1,x},\cdots,\phi_{L,x}\}$ denotes first $L$ sets of principal components

# CoDa in action

4. Principal component analysis: applied to the matrix $\boldsymbol{H}_x = \{h_{t,1}, \dots, h_{t,K}\}$ to obtain the estimated principal components and their associated scores,

$$h_{t,x} = \sum_{\ell=1}^{\min(n,K)} \beta_{t,\ell}\phi_{\ell,x} \approx \sum_{\ell=1}^{L} \beta_{t,\ell}\phi_{\ell,x}$$

- $\{\phi_{1,x}, \cdots, \phi_{L,x}\}$ denotes first $L$ sets of principal components
- $\{\beta_{t,1}, \dots, \beta_{t,L}\}$ denotes first $L$ sets of principal component scores for time $t$

# CoDa in action

4. Principal component analysis: applied to the matrix $\boldsymbol{H}_x = \{h_{t,1},\dots,h_{t,K}\}$ to obtain the estimated principal components and their associated scores,

$$h_{t,x} = \sum_{\ell=1}^{\min(n,K)} \beta_{t,\ell}\phi_{\ell,x} \approx \sum_{\ell=1}^{L} \beta_{t,\ell}\phi_{\ell,x}$$

- $\{\phi_{1,x},\cdots,\phi_{L,x}\}$ denotes first $L$ sets of principal components
- $\{\beta_{t,1},\dots,\beta_{t,L}\}$ denotes first $L$ sets of principal component scores for time $t$
- $L$ denotes number of retained components

# CoDa in action

5. Forecast of principal component scores: Via an exponential smoothing method, obtain $h$-step-ahead forecast of $\ell^{\text{th}}$ principal component score $\widehat{\beta}_{n+h|n,\ell}$

# CoDa in action

5. Forecast of principal component scores: Via an exponential smoothing method, obtain $h$-step-ahead forecast of $\ell^{\text{th}}$ principal component score $\widehat{\beta}_{n+h|n,\ell}$
6. Conditioning on estimated principal components and observations, forecast of $h_{n+h|n,x}$ is obtained by

$$\widehat{h}_{n+h|n,x} = \sum_{\ell=1}^{L} \widehat{\beta}_{n+h|n,\ell}\widehat{\phi}_{\ell,x}$$

# CoDa in action

5. Forecast of principal component scores: Via an exponential smoothing method, obtain $h$-step-ahead forecast of $\ell^{\text{th}}$ principal component score $\widehat{\beta}_{n+h|n,\ell}$
6. Conditioning on estimated principal components and observations, forecast of $h_{n+h|n,x}$ is obtained by

$$\widehat{h}_{n+h|n,x} = \sum_{\ell=1}^{L} \widehat{\beta}_{n+h|n,\ell} \widehat{\phi}_{\ell,x}$$

7. Transform back to compositional data: take inverse centered log-ratio transformation

$$\widehat{f}_{n+h|n,x} = C\left[\exp^{\widehat{h}_{n+h|n,x}}\right]$$

# CoDa in action

7. $C[\cdot]$ is closure operator, performing standardization

$$\widehat{f}_{n+h|n,x} = \frac{\exp^{\widehat{h}_{n+h|n,x}}}{\exp^{\widehat{h}_{n+h|n,1}} + \cdots + \exp^{\widehat{h}_{n+h|n,K}}}$$

# CoDa in action

7 $C[\cdot]$ is closure operator, performing standardization

$$\widehat{f}_{n+h|n,x} = \frac{\exp^{\widehat{h}_{n+h|n,x}}}{\exp^{\widehat{h}_{n+h|n,1}} + \dots + \exp^{\widehat{h}_{n+h|n,K}}}$$

8 Add back the geometric means, to obtain forecasts of life-table death matrix $\widehat{d}_{n+h|n,x}$:

$$\begin{aligned}\widehat{d}_{n+h|n,x} &= C\left[\widehat{f}_{n+h|n,x} \times \alpha_x\right] \\ &= \left[\frac{\widehat{f}_{n+h|n,1} \times \alpha_1}{\sum_{x=1}^{K} \widehat{f}_{n+h|n,x} \times \alpha_x}, \dots, \frac{\widehat{f}_{n+h|n,K} \times \alpha_K}{\sum_{x=1}^{K} \widehat{f}_{n+h|n,x} \times \alpha_x}\right]\end{aligned}$$

where $\alpha_x$ denotes age-specific geometric mean of $d_{t,x}$

# Selecting the number of components

To determine number of components $L$, determine the value of $L$ as the minimum number of components that reaches a certain level of proportion of total variance explained by $L$ leading components

$$L = \operatorname*{arg\,min}_{L: L \geq 1} \left\{ \sum_{\ell=1}^{L} \widehat{\lambda}_\ell \Big/ \sum_{\ell=1}^{\min\{n,K\}} \widehat{\lambda}_\ell \mathbb{1}_{\{\widehat{\lambda}_\ell > 0\}} \right\}$$

where $\delta = 95\%$, $\mathbb{1}\{\cdot\}$ denotes binary indicator function excluding possible zero eigenvalues. The chosen $L = 1$.

# Bootstrapped forecasts

1. Bootstrapped functional time series can be obtained

$$\widehat{h}^b_{t,x} = \sum_{\ell=1}^{L} \widehat{\beta}^b_{t,\ell}\widehat{\phi}_{\ell,x}, \qquad t = 1,\dots,n,$$

where $\widehat{\beta}^b_{t,\ell}$: bootstrapped $\ell^{\text{th}}$ principal component scores, for $b = 1,\dots,B$ and $B$ is the number of bootstrap replications

# Bootstrapped forecasts

1. Bootstrapped functional time series can be obtained

$$\widehat{h}^b_{t,x} = \sum_{\ell=1}^{L} \widehat{\beta}^b_{t,\ell}\widehat{\phi}_{\ell,x}, \qquad t = 1,\dots,n,$$

   where $\widehat{\beta}^b_{t,\ell}$: bootstrapped $\ell^{\text{th}}$ principal component scores, for $b = 1,\dots,B$ and $B$ is the number of bootstrap replications

2. For each bootstrap replication, we obtain the forecast of $h_{n+h,x}$ as

$$\widehat{h}^b_{n+h,x} = \sum_{\ell=1}^{L} \widehat{\beta}^b_{n+h,\ell}\widehat{\phi}_{\ell,x},$$

   $\widehat{\beta}^b_{n+h,\ell}$: forecast of the bootstrapped principal component scores

# Bootstrapped forecasts

1. Bootstrapped functional time series can be obtained

$$\widehat{h}^b_{t,x} = \sum_{\ell=1}^{L} \widehat{\beta}^b_{t,\ell}\widehat{\phi}_{\ell,x}, \qquad t = 1, \dots, n,$$

where $\widehat{\beta}^b_{t,\ell}$: bootstrapped $\ell^{\text{th}}$ principal component scores, for $b = 1, \dots, B$ and $B$ is the number of bootstrap replications

2. For each bootstrap replication, we obtain the forecast of $h_{n+h,x}$ as

$$\widehat{h}^b_{n+h,x} = \sum_{\ell=1}^{L} \widehat{\beta}^b_{n+h,\ell}\widehat{\phi}_{\ell,x},$$

$\widehat{\beta}^b_{n+h,\ell}$: forecast of the bootstrapped principal component scores

3. By randomly sampling with replacement the observations corresponding to the year index of the in-sample fitted errors, we obtain a set of bootstrapped model residuals

# Forecast reconciliation of death count

Japanese data follow a three-level hierarchy, coupled with sex grouping variable (S. & Hyndman, 2017, JCGS; S. & Haberman, IME)

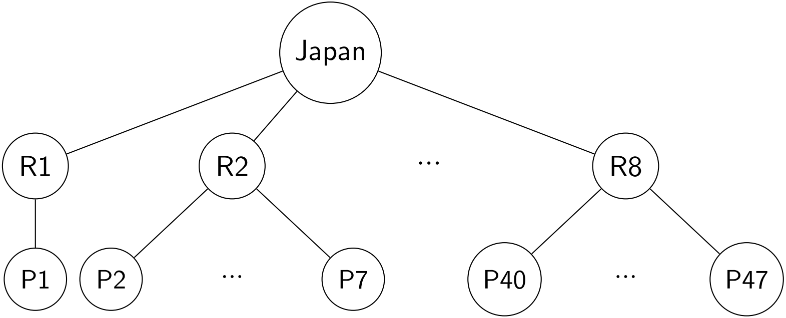



**Figure:** *Japanese geographical hierarchy tree diagram*

Refer to a disaggregated series using notation $X \times S$; $X$ is geographical area and $S$ is sex

$$
\underbrace{\begin{bmatrix}
\text{d}_{\text{Japan*T},t} \\
\text{d}_{\text{Japan*F},t} \\
\text{d}_{\text{Japan*M},t} \\
\text{d}_{\text{R1*T},t} \\
\vdots \\
\text{d}_{\text{R8*T},t} \\
\text{d}_{\text{R1*F},t} \\
\vdots \\
\text{d}_{\text{R8*F},t} \\
\text{d}_{\text{R1*M},t} \\
\vdots \\
\text{d}_{\text{R8*M},t} \\
\text{d}_{\text{P1*T},t} \\
\vdots \\
\text{d}_{\text{P47*T},t} \\
\text{d}_{\text{P1*F},t} \\
\text{d}_{\text{P1*M},t} \\
\vdots \\
\text{d}_{\text{P47*F},t} \\
\text{d}_{\text{P47*M},t}
\end{bmatrix}}_{\boldsymbol{d}_t}
=
\underbrace{\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix}}_{\boldsymbol{S}:=\text{summing matrix}}
\underbrace{\begin{bmatrix}
\text{d}_{\text{P1*F},t} \\
\text{d}_{\text{P1*M},t} \\
\text{d}_{\text{P2*F},t} \\
\text{d}_{\text{P2*M},t} \\
\vdots \\
\text{d}_{\text{P47*F},t} \\
\text{d}_{\text{P47*M},t}
\end{bmatrix}}_{\boldsymbol{b}_t}
$$

# Bottom-up method

1. Generates independent forecasts for each series at most disaggregated level, aggregate these to produce required forecasts

# Bottom-up method

1. Generates independent forecasts for each series at most disaggregated level, aggregate these to produce required forecasts
2. Using summing matrix, obtain reconciled forecasts

$$\overline{\boldsymbol{D}}_{n+h|n} = \boldsymbol{S} \times \widehat{\boldsymbol{b}}_{n+h|n}$$

where $\overline{\boldsymbol{D}}_{n+h|n}$ denotes reconciled forecasts

# Bottom-up method

1. Generates independent forecasts for each series at most disaggregated level, aggregate these to produce required forecasts
2. Using summing matrix, obtain reconciled forecasts

$$\overline{\boldsymbol{D}}_{n+h|n} = \boldsymbol{S} \times \widehat{\boldsymbol{b}}_{n+h|n}$$

   where $\overline{\boldsymbol{D}}_{n+h|n}$ denotes reconciled forecasts
3. Performs well when there is a strong *signal-to-noise* ratio

# Optimal-combination method

1. Optimal-combination method combines independent forecasts through linear regression, generated revised forecasts are as close as possible to independent forecasts but consistent with respect to the group structure

# Optimal-combination method

1. Optimal-combination method combines independent forecasts through linear regression, generated revised forecasts are as close as possible to independent forecasts but consistent with respect to the group structure
2. Using independent forecasts are responses, linear regression

$$\boldsymbol{D}_{n+h} = \boldsymbol{S}\boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h},$$

# Optimal-combination method

1. Optimal-combination method combines independent forecasts through linear regression, generated revised forecasts are as close as possible to independent forecasts but consistent with respect to the group structure
2. Using independent forecasts are responses, linear regression

$$\boldsymbol{D}_{n+h} = \boldsymbol{S}\boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h},$$

- $\boldsymbol{D}_{n+h}$ is a matrix of $h$-step-ahead values for all series;

# Optimal-combination method

1. Optimal-combination method combines independent forecasts through linear regression, generated revised forecasts are as close as possible to independent forecasts but consistent with respect to the group structure
2. Using independent forecasts are responses, linear regression

$$\boldsymbol{D}_{n+h} = \boldsymbol{S}\boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h},$$

- $\boldsymbol{D}_{n+h}$ is a matrix of $h$-step-ahead values for all series;
- $\boldsymbol{\beta}_{n+h} = \text{E}[\boldsymbol{b}_{n+h}|\boldsymbol{D}_1, \dots, \boldsymbol{D}_n]$ is unknown mean of independent forecasts of the bottom-level series;

# Optimal-combination method

1. Optimal-combination method combines independent forecasts through linear regression, generated revised forecasts are as close as possible to independent forecasts but consistent with respect to the group structure
2. Using independent forecasts are responses, linear regression

$$\boldsymbol{D}_{n+h} = \boldsymbol{S}\boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h},$$

- $\boldsymbol{D}_{n+h}$ is a matrix of $h$-step-ahead values for all series;
- $\boldsymbol{\beta}_{n+h} = \mathsf{E}[\boldsymbol{b}_{n+h}|\boldsymbol{D}_1, \dots, \boldsymbol{D}_n]$ is unknown mean of independent forecasts of the bottom-level series;
- $\boldsymbol{\epsilon}_{n+h}$ denotes reconciliation errors

# Estimating regression coefficient

To estimate regression coefficient, Hyndman et al. (2011) and Hyndman et al. (2016) proposed a weighted least-squares solution

$$\widehat{\boldsymbol{\beta}}_{n+h} = \left( \boldsymbol{S}^\top \underbrace{\boldsymbol{W}_h^{-1}}_{\text{pain☹}} \boldsymbol{S} \right)^{-1} \boldsymbol{S}^\top \boldsymbol{W}_h^{-1} \widehat{\boldsymbol{D}}_{n+h}$$

where $\boldsymbol{W}_h$ is a diagonal matrix

# How to estimate $\boldsymbol{W}_h$?

1. Assuming error terms follow same group structure, $\boldsymbol{W}_h = k_h \boldsymbol{I}$ and $\boldsymbol{I}$ is identity matrix. Revised forecasts are

$$\overline{\boldsymbol{D}}_{n+h} = \boldsymbol{S}\widehat{\boldsymbol{\beta}}_{n+h} = \boldsymbol{S}(\boldsymbol{S}^\top \boldsymbol{S})^{-1}\boldsymbol{S}^\top \widehat{\boldsymbol{D}}_{n+h},$$

where $k_h$ is a constant (OLS)

# How to estimate $\boldsymbol{W}_h$?

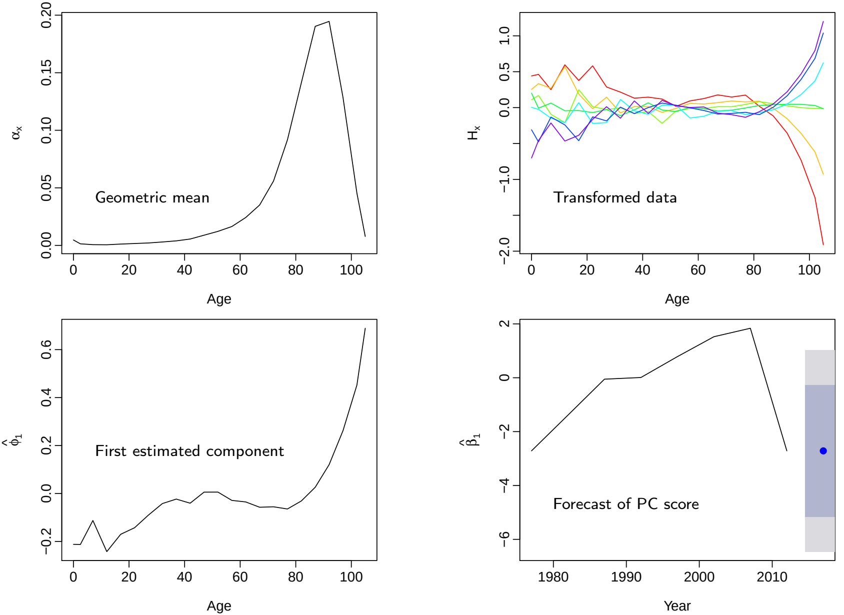
1. Assuming error terms follow same group structure, $\boldsymbol{W}_h = k_h \boldsymbol{I}$ and $\boldsymbol{I}$ is identity matrix. Revised forecasts are

$$\overline{\boldsymbol{D}}_{n+h} = \boldsymbol{S}\widehat{\boldsymbol{\beta}}_{n+h} = \boldsymbol{S}(\boldsymbol{S}^\top \boldsymbol{S})^{-1}\boldsymbol{S}^\top \widehat{\boldsymbol{D}}_{n+h},$$

   where $k_h$ is a constant (OLS)

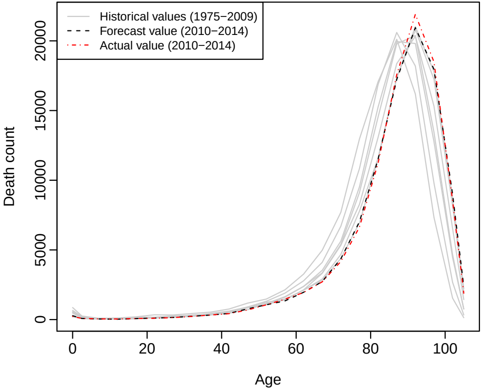
2. Assuming $\boldsymbol{W}_h = k_h \times \boldsymbol{W}_1$, we approximate $\boldsymbol{W}_1$ by its diagonal using in-sample fitted errors. Assigning weights as inverse proportion to variance, so places smallest weights where we have largest residual variance (WLS)

# Model fitting (Okinawa female data)



# Forecast death counts

Based on historical death from 1975 to 2009, produce one-step-ahead point forecasts of age-specific life-table death between 2010 and 2014




Age distribution of death counts continues to be negative skewed with more deaths occurring at older ages

# Expanding window analysis

1. Using the first 6 observations of five-year interval from 1975 to 2004 in Japanese age-specific life-table death counts, produce *one-step-ahead* point forecasts

# Expanding window analysis

1. Using the first 6 observations of five-year interval from 1975 to 2004 in Japanese age-specific life-table death counts, produce *one-step-ahead* point forecasts
2. Re-estimate parameters in the CoDa method using the first 7 observations from 1975 to 2009. Forecasts from estimated models are produced for *one-step-ahead*

# Expanding window analysis

1. Using the first 6 observations of five-year interval from 1975 to 2004 in Japanese age-specific life-table death counts, produce *one-step-ahead* point forecasts
2. Re-estimate parameters in the CoDa method using the first 7 observations from 1975 to 2009. Forecasts from estimated models are produced for *one-step-ahead*
3. With two one-step-ahead forecasts, evaluate out-of-sample forecast accuracy

# Point forecast evaluation

1. MAPE measures how close forecasts are to the actual values of variable being forecast, regardless of direction of forecast errors

# Point forecast evaluation

1. MAPE measures how close forecasts are to the actual values of variable being forecast, regardless of direction of forecast errors
2. For each series $k$, error can be expressed as

$$\text{MAPE}_k = \frac{1}{23 \times 2} \sum_{\xi=1}^{2} \sum_{x=1}^{23} \left| \frac{d_{n+\xi,x}^k - \widehat{d}_{n+\xi,x}}{d_{n+\xi,x}^k} \right| \times 100,$$

where $d_{n+\xi,x}^k$ denotes actual holdout sample for age $x$ and forecasting year $\xi$ in $k^{\text{th}}$ series

# Point forecast evaluation

1. MAPE measures how close forecasts are to the actual values of variable being forecast, regardless of direction of forecast errors
2. For each series $k$, error can be expressed as

$$\text{MAPE}_k = \frac{1}{23 \times 2} \sum_{\xi=1}^{2} \sum_{x=1}^{23} \left| \frac{d_{n+\xi,x}^k - \widehat{d}_{n+\xi,x}}{d_{n+\xi,x}^k} \right| \times 100,$$

   where $d_{n+\xi,x}^k$ denotes actual holdout sample for age $x$ and forecasting year $\xi$ in $k^{\text{th}}$ series
3. By averaging $\text{MAPE}_m$ across number of series within each level of disaggregation, obtain an overall assessment of point forecast accuracy for each level within collection of series

$$\text{MAPE} = \frac{1}{M_i} \sum_{m=1}^{M_i} \text{MAPE}_m$$

   where $M_i$ denotes number of series at $i^{\text{th}}$ level of disaggregation

# Interval forecast evaluation

1. Consider the common case of symmetric $100(1-\gamma)\%$ prediction intervals, with lower and upper bounds that are predictive quantiles at $\gamma/2$ and $1-\gamma/2$, denoted by $\widehat{d}^{l}_{n+\xi,x}$ and $\widehat{d}^{u}_{n+\xi,x}$

# Interval forecast evaluation

1. Consider the common case of symmetric $100(1-\gamma)\%$ prediction intervals, with lower and upper bounds that are predictive quantiles at $\gamma/2$ and $1-\gamma/2$, denoted by $\widehat{d}^{l}_{n+\xi,x}$ and $\widehat{d}^{u}_{n+\xi,x}$
2. A scoring rule for the interval forecasts at time point $d_{\xi+h,x}$ is

$$S^{k}_{\gamma,\xi}\left[\widehat{d}^{l}_{n+\xi,x}, \widehat{d}^{u}_{n+\xi,x}, d_{n+\xi,x}\right] = \left(\widehat{d}^{u}_{n+\xi,x} - \widehat{d}^{l}_{n+\xi,x}\right) + \frac{2}{\gamma}\left(\widehat{d}^{l}_{n+\xi,x} - d_{n+\xi,x}\right)$$

$$\mathbb{1}\left\{d_{n+\xi,x} < \widehat{d}^{l}_{n+\xi,x}\right\} + \frac{2}{\gamma}\left(d_{n+\xi,x} - \widehat{d}^{u}_{n+\xi,x}\right)\mathbb{1}\left\{d_{n+\xi,x} > \widehat{d}^{u}_{n+\xi,x}\right\}$$

where $\mathbb{1}\{\cdot\}$: binary indicator function, $\gamma$: level of significance

# Interval forecast evaluation

1. Consider the common case of symmetric $100(1-\gamma)\%$ prediction intervals, with lower and upper bounds that are predictive quantiles at $\gamma/2$ and $1-\gamma/2$, denoted by $\widehat{d}^{l}_{n+\xi,x}$ and $\widehat{d}^{u}_{n+\xi,x}$
2. A scoring rule for the interval forecasts at time point $d_{\xi+h,x}$ is

$$S^k_{\gamma,\xi}\left[\widehat{d}^{l}_{n+\xi,x}, \widehat{d}^{u}_{n+\xi,x}, d_{n+\xi,x}\right] = \left(\widehat{d}^{u}_{n+\xi,x} - \widehat{d}^{l}_{n+\xi,x}\right) + \frac{2}{\gamma}\left(\widehat{d}^{l}_{n+\xi,x} - d_{n+\xi,x}\right)$$
$$\mathbb{1}\left\{d_{n+\xi,x} < \widehat{d}^{l}_{n+\xi,x}\right\} + \frac{2}{\gamma}\left(d_{n+\xi,x} - \widehat{d}^{u}_{n+\xi,x}\right)\mathbb{1}\left\{d_{n+\xi,x} > \widehat{d}^{u}_{n+\xi,x}\right\}$$

where $\mathbb{1}\{\cdot\}$: binary indicator function, $\gamma$: level of significance

3. For different ages and years in the forecasting period, mean interval score is

$$\overline{S}^k_{\gamma} = \frac{1}{23\times 2}\sum_{\xi=1}^{2}\sum_{x=1}^{23} S^k_{\gamma,\xi}\left[\widehat{d}^{l}_{n+\xi,x}, \widehat{d}^{u}_{n+\xi,x}; d_{n+\xi,x}\right], \quad \overline{S}_{\gamma}(h) = \frac{1}{M_i}\sum_{k=1}^{M_i}\overline{S}^k_{\gamma}$$

# Point forecast evaluation for forecasting death counts

| | MAPE | | number of smaller errors / number of series at each level | |
|---|---|---|---|---|
| Level | CoDa | RW | CoDa | RW |
| Total | **6.8831** | 8.0765 | 100% | 0% |
| Sex | **7.6630** | 8.2054 | 100% | 0% |
| Region | **8.4605** | 9.3633 | 87.50% | 12.50% |
| Region + Sex | **9.5975** | 10.0833 | 68.75% | 31.25% |
| Prefecture | **10.1161** | 11.5056 | 91.49% | 8.51% |
| Prefecture + Sex | **12.4527** | 13.8352 | 81.91% | 18.09% |

# Point forecast evaluation (reconciliation methods)

| Level | BU | OLS | WLS |
|---|---|---|---|
| Total | 7.6064 | 7.3324 | **7.2925** |
| Sex | 7.8143 | 7.5184 | **7.4846** |
| Region | 9.3641 | 9.0333 | **9.0323** |
| Region + Sex | 9.4131 | **9.1015** | 9.1639 |
| Prefecture | 11.1255 | 10.7811 | **10.7494** |
| Prefecture + Sex | 12.4527 | **12.1693** | 12.2157 |
| Overall Mean | 9.6294 | **9.3227** | 9.3231 |

# Interval forecast evaluation

| Level | CoDa | BU | OLS | WLS |
|---|---|---|---|---|
| Total | 1108.76 | 900.91 | **848.53** | 857.71 |
| Sex | 1089.50 | **947.71** | 962.52 | 991.92 |
| Region | 1145.86 | 815.33 | **772.02** | 780.80 |
| Region + Sex | 1123.92 | 771.30 | 724.75 | **719.90** |
| Prefecture | 1201.80 | 900.82 | **791.10** | 792.98 |
| Prefecture + Sex | 1187.09 | 1187.09 | 1110.32 | **1081.02** |
| Overall Mean | 1142.82 | 920.53 | **868.21** | 870.72 |

# Life annuity

1. An annuity is a contract offered by insurers guaranteeing a steady stream of payments for either a fixed term or lifetime of annuitants in exchange for an initial premium fee

# Life annuity

1. An annuity is a contract offered by insurers guaranteeing a steady stream of payments for either a fixed term or lifetime of annuitants in exchange for an initial premium fee
2. Apply forecasts of death counts to calculation of single-premium term immediate annuities

# Life annuity

1. An annuity is a contract offered by insurers guaranteeing a steady stream of payments for either a fixed term or lifetime of annuitants in exchange for an initial premium fee
2. Apply forecasts of death counts to calculation of single-premium term immediate annuities
3. $\tau$ year survival probability of a person aged $x$ currently at $t = 0$ is determined by

$$
\begin{aligned}
{}_\tau p_x &= \prod_{j=1}^{\tau} p_{x+j-1} \\
&= \prod_{j=1}^{\tau} (1 - q_{x+j-1}) = \prod_{j=1}^{\tau} \left(1 - \frac{d_{x+j-1}}{l_{x+j-1}}\right)
\end{aligned}
$$

where $d_{x+j-1}$ denotes number of death counts between ages $x + j - 1$ and $x + j$; $l_{x+j-1}$ denotes number of lives alive at age $x + j - 1$

# Annuity price calculation

Price of an annuity with maturity $T$ year, written for a $x$-year-old with benefit \$1 per year, is given

$$
\begin{aligned}
a_x^T\left(d_{1:T}^x\right) &= \sum_{\tau=1}^{T} B(t=0,\tau) \times \mathsf{E}\left(\mathbb{1}_{T_x>\tau} | d_{1:\tau}^x\right) \\
&= \sum_{\tau=1}^{T} \underbrace{B(t=0,\tau)}_{\text{bond price}} \times \underbrace{{}_{\tau}p_x(d_{1:\tau}^x)}_{\text{survival probability}}
\end{aligned}
$$

1. $B(t=0,\tau)$ is $\tau$-year bond price, where $\tau < T$

# Annuity price calculation

Price of an annuity with maturity $T$ year, written for a $x$-year-old with benefit \$1 per year, is given

$$
\begin{aligned}
a_x^T\left(d_{1:T}^x\right) &= \sum_{\tau=1}^{T} B(t=0,\tau) \times \mathsf{E}\left(\mathbb{1}_{T_x>\tau} | d_{1:\tau}^x\right) \\
&= \sum_{\tau=1}^{T} \underbrace{B(t=0,\tau)}_{\text{bond price}} \times \underbrace{{}_\tau p_x(d_{1:\tau}^x)}_{\text{survival probability}}
\end{aligned}
$$

1. $B(t=0,\tau)$ is $\tau$-year bond price, where $\tau < T$
2. $d_{1:\tau}^x$ is first $\tau$ elements of $d_{1:T}^x$

# Annuity price calculation

Price of an annuity with maturity $T$ year, written for a $x$-year-old with benefit \$1 per year, is given

$$a_x^T\left(d_{1:T}^x\right) = \sum_{\tau=1}^{T} B(t=0,\tau) \times \mathsf{E}\left(\mathbb{1}_{T_x>\tau} | d_{1:\tau}^x\right)$$

$$= \sum_{\tau=1}^{T} \underbrace{B(t=0,\tau)}_{\text{bond price}} \times \underbrace{{}_\tau p_x(d_{1:\tau}^x)}_{\text{survival probability}}$$

1. $B(t=0,\tau)$ is $\tau$-year bond price, where $\tau < T$
2. $d_{1:\tau}^x$ is first $\tau$ elements of $d_{1:T}^x$
3. ${}_\tau p_x(d_{1:\tau}^x)$ denotes survival probability given a random $d_{1:\tau}^x$

# Comparison of life annuity premium calculation

1. Compare annuity price estimates for different ages and maturities between methods for a female policyholder living in Japan
2. Assume a constant interest rate at $\eta = 3\%$ and $B(t = 0, \tau) = \exp^{-\eta\tau}$

# Fixed-term annuity price (age = 60) for Japan (F, M, T)

| Series | | $T = 5$ | $T = 10$ | $T = 15$ | $T = 20$ | $T = 25$ | $T = 30$ |
|---|---|---|---|---|---|---|---|
| Female | LB | 4.5255 | 8.3311 | 11.4895 | 14.0474 | 16.0071 | 17.3350 |
| | Mean | 4.5288 | 8.3448 | 11.5274 | 14.1288 | 16.1626 | 17.6018 |
| | UB | 4.5370 | 8.3830 | 11.6356 | 14.3754 | 16.6576 | 18.5063 |
| Male | LB | 4.4540 | 8.0646 | 10.9043 | 13.0075 | 14.4030 | 15.1543 |
| | Mean | 4.4602 | 8.0897 | 10.9659 | 13.1356 | 14.6187 | 15.4637 |
| | UB | 4.4729 | 8.1467 | 11.1276 | 13.4911 | 15.2772 | 16.4944 |
| Total | LB | 4.4912 | 8.2011 | 11.2047 | 13.5497 | 15.2536 | 16.3333 |
| | Mean | 4.4958 | 8.2223 | 11.2618 | 13.6700 | 15.4712 | 16.6753 |
| | UB | 4.5056 | 8.2714 | 11.4018 | 13.9851 | 16.0845 | 17.7222 |

# Thank you

1. A draft paper is available upon request from hanlin.shang@anu.edu.au

# Thank you

1. A draft paper is available upon request from hanlin.shang@anu.edu.au
2. Follow me at Research Gate https://www.researchgate.net/profile/Han_Lin_Shang

References
[1] Aitchison, J. (1982), ‘The statistical analysis of compositional data’, *JRSSB*, **44**(2), 139-177.
[2] Aitchison, J. (1986), The Statistical Analysis of Compositional Data, Chapman & Hall, London.
[3] Boucher, M.-P. B., Canudas-Romo, V. and Vaupel, J. W. (2014), Convergent mortality levels? Coherent mortality forecasts among industrialized countries, in ‘Population Association of America’, Boston, MA.
[4] Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. and Shang, H. L. (2011), ‘Optimal combination forecasts for hierarchical time series’, *CSDA*, **55**, 2579-2589.
[5] Hyndman, R. J., Lee, A. and Wang, E. (2016), ‘Fast computation of reconciled forecasts for hierarchical and grouped time series’, *CSDA*, **97**, 16-32.
[6] Shang, H. L. and Haberman, S. (2017), ‘Grouped multivariate and functional time series forecasting: An application to annuity pricing’, *IME*, **75**, 166-179.
[7] Shang, H. L. and Hyndman, R. J. (2017), ‘Grouped functional time series forecasting: An application to age-specific mortality rates’, *JCGS*, **26**(2), 330-343.