Coherent forecasting age distribution of death counts for *multiple* populations

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

(Joint work with Steven Haberman, Cass Business School)

July 1, 2020
Zoom presentation

About

1. In human mortality, three functions are studied[1]:

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

About

1. In human mortality, three functions are studied[1]:
   - *hazard* (e.g., central mortality rate)

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

About

1. In human mortality, three functions are studied[1]:
   - *hazard* (e.g., central mortality rate)
   - *survival* (e.g., life expectancy)

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

About

1. In human mortality, three functions are studied[1]:
   - *hazard* (e.g., central mortality rate)
   - *survival* (e.g., life expectancy)
   - *life table deaths* (e.g., life table death count)

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

About

1. In human mortality, three functions are studied[1]:
   - *hazard* (e.g., central mortality rate)
   - *survival* (e.g., life expectancy)
   - *life table deaths* (e.g., life table death count)
2. Forecasting *central mortality rates* leads to more pessimistic forecasts than using *survival probabilities & life table deaths*

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

About

1. In human mortality, three functions are studied[1]:
   - *hazard* (e.g., central mortality rate)
   - *survival* (e.g., life expectancy)
   - *life table deaths* (e.g., life table death count)
2. Forecasting *central mortality rates* leads to more pessimistic forecasts than using *survival probabilities* & *life table deaths*
3. Modeling life table deaths, resemblance with probability density function[2]

[1] M-P. Bergeron-Boucher, S. Kjaegaard, J. E. Oeppen, J. W. Vaupel (2019) The impact of the choice of life table statistics when forecasting mortality, **41**, article 43

[2] 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

Aims

1. When modeling life table death counts for multiple populations,

Aims

1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy

Aims

1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy
   - how to forecast multiple death counts so that forecasts are non-negative & sum to radix

Aims

1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy
   - how to forecast multiple death counts so that forecasts are non-negative & sum to radix
2. Consider a *compositional data analysis* (CoDa) to forecast future life table death counts

Aims

1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy
   - how to forecast multiple death counts so that forecasts are non-negative & sum to radix
2. Consider a *compositional data analysis* (CoDa) to forecast future life table death counts
3. Improved forecast accuracy of life table death counts is useful

Aims

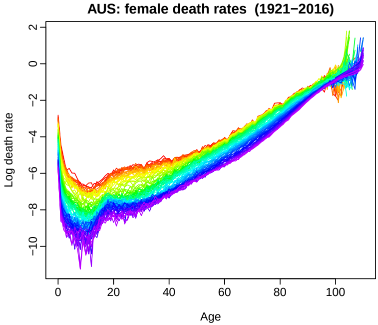
1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy
   - how to forecast multiple death counts so that forecasts are non-negative & sum to radix
2. Consider a *compositional data analysis* (CoDa) to forecast future life table death counts
3. Improved forecast accuracy of life table death counts is useful
   - to demographers for estimating survival probabilities & life expectancy

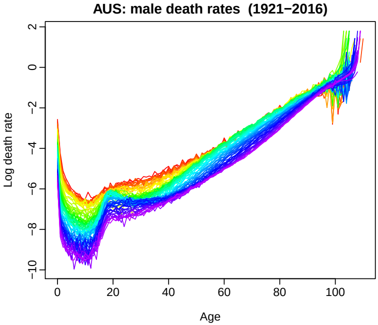
Aims

1. When modeling life table death counts for multiple populations,
   - how to incorporate correlation to improve point & interval forecast accuracy
   - how to forecast multiple death counts so that forecasts are non-negative & sum to radix
2. Consider a *compositional data analysis* (CoDa) to forecast future life table death counts
3. Improved forecast accuracy of life table death counts is useful
   - to demographers for estimating survival probabilities & life expectancy
   - to actuaries for determining fixed-term annuity prices for various ages & maturities

Australian central mortality rate

1. Many approaches for forecasting age-specific central mortality rate[3]

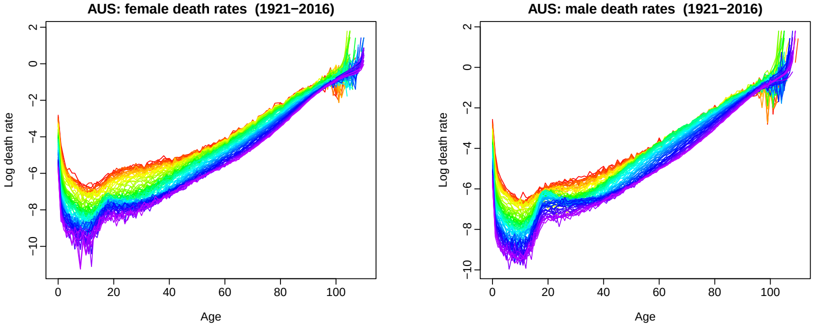




[3] Booth, H. (2006). 'Demographic forecasting: 1980 to 2005 in review', *International Journal of Forecasting*, **22**(3), 547-581

Australian central mortality rate

1. Many approaches for forecasting age-specific central mortality rate[3]

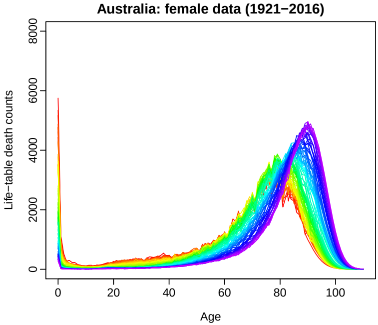


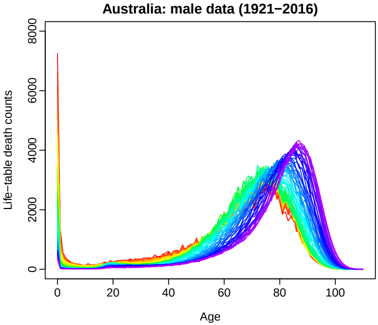
2. Constrained value:
   mortality rate $\in [0, 1] \rightarrow$ log mortality rate $\in (-\infty, 0]$

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

Australian life table death count

1. Instead of central mortality rates, model life table death count[4]

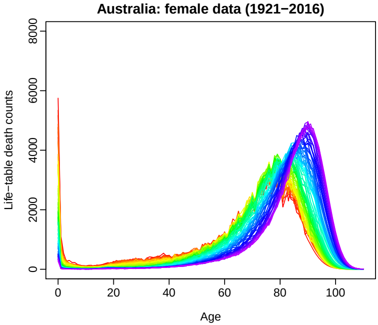


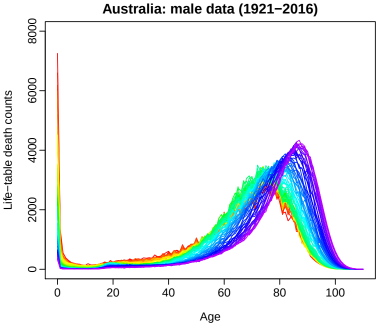


[4] Canudas-Romo, V. (2010). 'Three measures of longevity: Time trends and record values', *Demography*, **47**(2), 299-312

Australian life table death count

1. Instead of central mortality rates, model life table death count[4]





2. Constrained value: non-negativity & summability

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

Advantages of life table death count

1. Visualize & forecast a redistribution of life table deaths (e.g., deaths at younger ages are shifted towards older ages)

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

[6] 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

Advantages of life table death count

1. Visualize & forecast a redistribution of life table deaths (e.g., deaths at younger ages are shifted towards older ages)
2. Life table death counts yield more information on 'central longevity indicators'

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

[6] 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

Advantages of life table death count

1. Visualize & forecast a redistribution of life table deaths (e.g., deaths at younger ages are shifted towards older ages)
2. Life table death counts yield more information on 'central longevity indicators'
   - mean, median, mode age at death[5]

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

[6] 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

Advantages of life table death count

1. Visualize & forecast a redistribution of life table deaths (e.g., deaths at younger ages are shifted towards older ages)
2. Life table death counts yield more information on 'central longevity indicators'
   - mean, median, mode age at death[5]
   - lifespan variability (e.g., standard deviation or inter-quartile)[6]

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

[6]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

Cohort vs period life table death count

1. Cohort life table[7] describes life history of a specific group of individuals, but depends on forecast mortality for those cohorts born more recently (filling gap)

[7] S. Rizzi, S. Kjaergaard, M-P. Bergeron Boucher, C. G. Camarda, R. Lindahl-Jacobsen, J. W. Vaupel (2020) Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model, *International Journal of Forecasting*, in press

[8] 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

Cohort vs period life table death count

1. Cohort life table[7] describes life history of a specific group of individuals, but depends on forecast mortality for those cohorts born more recently (filling gap)
2. Choose to study period life table[8], representing mortality conditions in a period

[7] S. Rizzi, S. Kjaergaard, M-P. Bergeron Boucher, C. G. Camarda, R. Lindahl-Jacobsen, J. W. Vaupel (2020) Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model, *International Journal of Forecasting*, in press

[8] 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

CoDa literature

1. Probability density function is an example of compositional data

[9] 'Robust principal component analysis for power transformed compositional data', *Journal of the American Statistical Association: Theory and Methods*, **110**(509), 136-148

[10] 'A directional mixed effects model for compositional expenditure data', *Journal of the American Statistical Association: Applications and Case Studies*, **112**(517), 24-36

[11] Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, in 'European Population Conference', Barcelona, Spain

CoDa literature

1. Probability density function is an example of compositional data
2. Scealy et al. (2017)[9] apply CoDa to study concentration of chemical elements in rock samples

[9] 'Robust principal component analysis for power transformed compositional data', *Journal of the American Statistical Association: Theory and Methods*, **110**(509), 136-148

[10] 'A directional mixed effects model for compositional expenditure data', *Journal of the American Statistical Association: Applications and Case Studies*, **112**(517), 24-36

[11] Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, in 'European Population Conference', Barcelona, Spain

CoDa literature

1. Probability density function is an example of compositional data
2. Scealy et al. (2017)[9] apply CoDa to study concentration of chemical elements in rock samples
3. Scealy & Welsh (2017)[10] apply CoDa to analyse total weekly expenditure on food & housing costs

---

[9] 'Robust principal component analysis for power transformed compositional data', *Journal of the American Statistical Association: Theory and Methods*, **110**(509), 136-148

[10] 'A directional mixed effects model for compositional expenditure data', *Journal of the American Statistical Association: Applications and Case Studies*, **112**(517), 24-36

[11] Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, in 'European Population Conference', Barcelona, Spain

CoDa literature

1. Probability density function is an example of compositional data
2. Scealy et al. (2017)[9] apply CoDa to study concentration of chemical elements in rock samples
3. Scealy & Welsh (2017)[10] apply CoDa to analyse total weekly expenditure on food & housing costs
4. In demography, Oeppen (2008)[11] & Bergeron-Boucher et al. (2017) apply PCA to forecast life table death counts

---

[9] 'Robust principal component analysis for power transformed compositional data', *Journal of the American Statistical Association: Theory and Methods*, **110**(509), 136-148

[10] 'A directional mixed effects model for compositional expenditure data', *Journal of the American Statistical Association: Applications and Case Studies*, **112**(517), 24-36

[11] Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, in 'European Population Conference', Barcelona, Spain

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

- S denotes a simplex

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

- S denotes a simplex
- c is a fixed constant

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

- S denotes a simplex
- c is a fixed constant
- $^\top$ denotes vector transpose

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

- S denotes a simplex
- c is a fixed constant
- $^\top$ denotes vector transpose
- simplex space is $I-1$ dimensional subset of real-valued space R^I

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

   - S denotes a simplex
   - c is a fixed constant
   - $^\top$ denotes vector transpose
   - simplex space is $I-1$ dimensional subset of real-valued space R^I

2. Compositional data set typically equal to

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

   - S denotes a simplex
   - c is a fixed constant
   - $^\top$ denotes vector transpose
   - simplex space is $I-1$ dimensional subset of real-valued space R^I
2. Compositional data set typically equal to
   - 1 (portions)

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

   - S denotes a simplex
   - c is a fixed constant
   - $^\top$ denotes vector transpose
   - simplex space is $I-1$ dimensional subset of real-valued space R^I

2. Compositional data set typically equal to
   - 1 (portions)
   - 100 (percentage)

CoDa

1. Between two constraints, space of compositional data is a simplex

$$S^I = \left\{ [f_t(u_1), \dots, f_t(u_I)]^\top, \quad 0 \leq f_t(u_i) \leq c, \quad \sum_{i=1}^{I} f_t(u_i) = c \right\},$$

   - S denotes a simplex
   - c is a fixed constant
   - $^\top$ denotes vector transpose
   - simplex space is $I-1$ dimensional subset of real-valued space R^I

2. Compositional data set typically equal to
   - 1 (portions)
   - 100 (percentage)
   - 10^6 for parts per million (ppm)

CoDa

1. Standard approach involves using a transformation (e.g., centred log-ratio) of raw data to remove constraints

CoDa

1. Standard approach involves using a transformation (e.g., centred log-ratio) of raw data to remove constraints
2. Apply standard statistical techniques to transformed data in an unconstrained space

Contributions

1. Links multiple-population forecasting with CoDa

Contributions

1. Links multiple-population forecasting with CoDa
2. Forecasts of life table death counts $\longrightarrow$ survival probabilities $\longrightarrow$ fixed-term annuity prices

Contributions

1. Links multiple-population forecasting with CoDa
2. Forecasts of life table death counts $\longrightarrow$ survival probabilities $\longrightarrow$ fixed-term annuity prices
3. Propose a non-parametric bootstrap for constructing prediction intervals for future life table death counts

Australian age- and sex-specific life table death counts

1. Consider Australian age- and sex-specific life table death counts from 1921 to 2016, obtained from HMD

[12] a population experiencing 100,000 births annually

Australian age- and sex-specific life table death counts

1. Consider Australian age- and sex-specific life table death counts from 1921 to 2016, obtained from HMD
2. Life table radix[12] is fixed at 100,000 at age 0 for each year

[12] a population experiencing 100,000 births annually

Australian age- and sex-specific life table death counts

1. Consider Australian age- and sex-specific life table death counts from 1921 to 2016, obtained from HMD
2. Life table radix[12] is fixed at 100,000 at age 0 for each year
3. For life table death counts, there are 111 ages, age 0, 1, . . . , 109, 110+

[12] a population experiencing 100,000 births annually

Australian age- and sex-specific life table death counts

1. Consider Australian age- and sex-specific life table death counts from 1921 to 2016, obtained from HMD
2. Life table radix[12] is fixed at 100,000 at age 0 for each year
3. For life table death counts, there are 111 ages, age 0, 1, . . . , 109, 110+
4. Due to rounding, there are zero counts for age 110+ at some years

[12] a population experiencing 100,000 births annually

Australian age- and sex-specific life table death counts

1. Consider Australian age- and sex-specific life table death counts from 1921 to 2016, obtained from HMD
2. Life table radix[12] is fixed at 100,000 at age 0 for each year
3. For life table death counts, there are 111 ages, age 0, 1, . . . , 109, 110+
4. Due to rounding, there are zero counts for age 110+ at some years
5. Prefer to use probability of dying & radix to recalculate estimated death counts (up to 6 decimal places)

---

[12] a population experiencing 100,000 births annually

Compositional data analysis

1. Compute *geometric mean* function

$$\alpha_n^j(u) = \exp\left\{\frac{1}{n}\sum_{t=1}^{n}\ln[f_t^j(u)]\right\}, \qquad j = 1, \dots, J,$$

where $J = 2$ representing female & male, $\ln(\cdot)$:= natural log

Compositional data analysis

1. Compute *geometric mean* function

$$\alpha_n^j(u) = \exp\left\{\frac{1}{n}\sum_{t=1}^{n}\ln[f_t^j(u)]\right\}, \qquad j = 1, \dots, J,$$

   where $J = 2$ representing female & male, $\ln(\cdot) :=$ natural log

2. Treat age as a continuum $u \in [0, 110]$ although age is observed at discrete points

Compositional data analysis

1. Compute *geometric mean* function

$$\alpha_n^j(u) = \exp\left\{\frac{1}{n}\sum_{t=1}^{n}\ln[f_t^j(u)]\right\}, \qquad j = 1, \dots, J,$$

where $J = 2$ representing female & male, $\ln(\cdot) :=$ natural log

2. Treat age as a continuum $u \in [0, 110]$ although age is observed at discrete points

3. Set

$$s_t^j(u) = \frac{f_t^j(u)/\alpha_n^j(u)}{\int_{u=0}^{110} f_t^j(u)/\alpha_n^j(u)du}$$

Compositional data analysis

1. Compute *geometric mean* function

$$\alpha_n^j(u) = \exp\left\{\frac{1}{n}\sum_{t=1}^{n}\ln[f_t^j(u)]\right\}, \qquad j = 1, \dots, J,$$

   where $J = 2$ representing female & male, $\ln(\cdot) :=$ natural log
2. Treat age as a continuum $u \in [0, 110]$ although age is observed at discrete points
3. Set

$$s_t^j(u) = \frac{f_t^j(u)/\alpha_n^j(u)}{\int_{u=0}^{110} f_t^j(u)/\alpha_n^j(u)du}$$

4. Geometric mean standardises ranges so that no range dominates weighting

Compositional data analysis

1. Apply centred log-ratio transformation

$$\beta_t^j(u) = \ln\left(\frac{s_t^j(u)}{g_t^j}\right),$$

where g_t^j is geometric mean

$$g_t^j = \exp\left\{\int_{u=0}^{110} \ln\left[s_t^j(u)\right] du\right\}$$

Compositional data analysis

1. Apply centred log-ratio transformation

$$\beta_t^j(u) = \ln\left(\frac{s_t^j(u)}{g_t^j}\right),$$

where g_t^j is geometric mean

$$g_t^j = \exp\left\{\int_{u=0}^{110} \ln\left[s_t^j(u)\right] du\right\}$$

2. Log-ratio transformation removes constraints on $f_t^j(u)$

Compositional data analysis

1. Apply centred log-ratio transformation

$$\beta_t^j(u) = \ln\left(\frac{s_t^j(u)}{g_t^j}\right),$$

where g_t^j is geometric mean

$$g_t^j = \exp\left\{\int_{u=0}^{110} \ln\left[s_t^j(u)\right] du\right\}$$

2. Log-ratio transformation removes constraints on $f_t^j(u)$
   - centering $\rightarrow$ *portion* (ensure summability)

Compositional data analysis

1. Apply centred log-ratio transformation

$$\beta_t^j(u) = \ln\left(\frac{s_t^j(u)}{g_t^j}\right),$$

where g_t^j is geometric mean

$$g_t^j = \exp\left\{\int_{u=0}^{110} \ln\left[s_t^j(u)\right] du\right\}$$

2. Log-ratio transformation removes constraints on $f_t^j(u)$
   - centering $\rightarrow$ *portion* (ensure summability)
   - logarithm removes non-negativity constraint

Univariate functional principal component decomposition

Applying *functional* principal component analysis to $\boldsymbol{\beta}^j(u) = \{\beta_1^j(u), \dots, \beta_n^j(u)\}$,

$$\begin{aligned} \beta_t^j(u) &= \sum_{\ell=1}^{n} \widehat{\gamma}_{t,\ell}^j \widehat{\phi}_{\ell}^j(u) \qquad \text{(dimension reduction)} \\ &= \sum_{\ell=1}^{L_j} \widehat{\gamma}_{t,\ell}^j \widehat{\phi}_{\ell}^j(u) + \widehat{\omega}_t^j(u), \end{aligned}$$

- $[\widehat{\phi}_1^j(u), \dots, \widehat{\phi}_{L_j}^j(u)]$ are first L_j estimated FPCs
- $(\widehat{\gamma}_{t,1}^j, \dots, \widehat{\gamma}_{t,L_j}^j)$ are PC scores at year t
- L_j is number of retained components for j^{th} population
- $\widehat{\omega}_t^j(u)$:= model residual function for j^{th} population at age u & year t

Selection of retained number of components

1. Determine L_j by eigenvalue ratio (ER) and growth ratio (GR) estimators[13]

[13] Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica*, **81**(3), 1203-1227

Selection of retained number of components

1. Determine L_j by eigenvalue ratio (ER) and growth ratio (GR) estimators[13]
2. ER estimator is obtained by maximising ratio of two adjacent eigenvalues of Var$(\boldsymbol{\beta}^j)$ arranged in descending order

$$\text{ER}(\ell) = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_{\ell+1}}, \qquad \ell = 1, 2, \dots, \text{Lmax}$$

where $\boldsymbol{\beta}^j$ is a discretized matrix

[13] Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica*, **81**(3), 1203-1227

Selection of retained number of components

1. Determine L_j by eigenvalue ratio (ER) and growth ratio (GR) estimators[13]
2. ER estimator is obtained by maximising ratio of two adjacent eigenvalues of Var$(\boldsymbol{\beta}^j)$ arranged in descending order

$$\text{ER}(\ell) = \frac{\tilde{\mu}_\ell}{\tilde{\mu}_{\ell+1}}, \qquad \ell = 1, 2, \dots, \text{Lmax}$$

   where $\boldsymbol{\beta}^j$ is a discretized matrix

3. GR estimator is

$$\tilde{\mu}_\ell^* = \frac{\tilde{\mu}_\ell}{\sum_{s=\ell+1}^{\min(n,I)} \tilde{\mu}_s}$$

$$\text{GR}(\ell) = \frac{\ln(1+\tilde{\mu}_\ell^*)}{\ln(1+\tilde{\mu}_{\ell+1}^*)}$$

[13] Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica*, **81**(3), 1203-1227

Selection of number of components

1. Optimal number L_j is

$$L_j = \max\left\{\max_{1\leq\ell\leq \text{Lmax}} \text{ER}(\ell), \max_{1\leq\ell\leq \text{Lmax}} \text{GR}(\ell)\right\}$$

Selection of number of components

1. Optimal number L_j is

$$L_j = \max\left\{\max_{1\leq\ell\leq\mathsf{Lmax}} \mathsf{ER}(\ell), \max_{1\leq\ell\leq\mathsf{Lmax}} \mathsf{GR}(\ell)\right\}$$

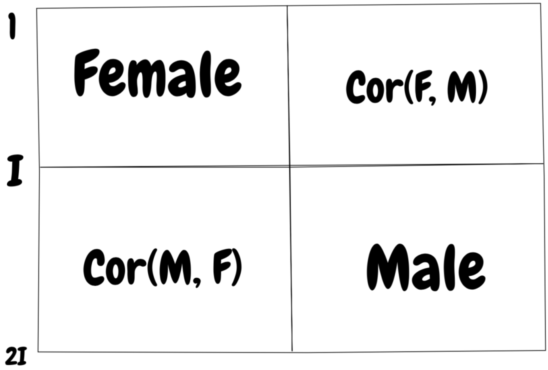
2. $\mathsf{Lmax} = \#\left\{\ell \geq 1 | \tilde{\mu}_\ell \geq \sum_{\ell=1}^{\min(n,I)} \tilde{\mu}_\ell / \min(n, I)\right\}$

Multivariate FPC decomposition

1. Stack $\boldsymbol{\beta}_t(u) = [\beta_t^1(u), \beta_t^2(u)]$ into a long vector of functional time series

Multivariate FPC decomposition

1. Stack $\boldsymbol{\beta}_t(u) = \left[\beta_t^1(u), \beta_t^2(u)\right]$ into a long vector of functional time series
2. Computationally, as $\boldsymbol{\beta}^j(u)$ is a matrix of $n \times I$, $\boldsymbol{\beta}(u)$ is a matrix of $n \times (2I)$



Multivariate FPC decomposition

Apply FPCA to $\boldsymbol{\beta}(u)$,

$$\begin{aligned}\beta_t^j(u) &= \sum_{\ell=1}^{n} \widehat{\gamma}_{t,\ell}^j \widehat{\phi}_\ell^j(u) \qquad \text{(dimension reduction)} \\ &\approx \sum_{\ell=1}^{L} \widehat{\gamma}_{t,\ell}^j \widehat{\phi}_\ell^j(u),\end{aligned}$$

- $\left[\widehat{\phi}_1^j(u), \dots, \widehat{\phi}_L^j(u)\right]$:= first L estimated FPCs extracted from variance of *stacked* functional time series
- $(\widehat{\gamma}_{\ell,1}^j, \dots, \widehat{\gamma}_{\ell,L}^j)$ are PC scores for j^{th} population
- Determine a common number of components, L, using ER tests

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]
2. Seek to find *common pattern* among populations & population-specific pattern

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]
2. Seek to find *common pattern* among populations & population-specific pattern
3. Multilevel FPC regression is

$$\beta_t^j(u) = U_t(u) + R_t^j(u) + e_t^j(u),$$

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]
2. Seek to find *common pattern* among populations & population-specific pattern
3. Multilevel FPC regression is

$$\beta_t^j(u) = U_t(u) + R_t^j(u) + e_t^j(u),$$

- $U_t(u)$:= a common trend

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]
2. Seek to find *common pattern* among populations & population-specific pattern
3. Multilevel FPC regression is

$$\beta_t^j(u) = U_t(u) + R_t^j(u) + e_t^j(u),$$

- $U_t(u)$:= a common trend
- $R^j(u)$:= a population-specific trend

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Model $\beta_t^j(u)$ via multilevel FPCA[14]
2. Seek to find *common pattern* among populations & population-specific pattern
3. Multilevel FPC regression is

$$\beta_t^j(u) = U_t(u) + R_t^j(u) + e_t^j(u),$$

- $U_t(u)$:= a common trend
- $R^j(u)$:= a population-specific trend
- $e_t^j(u)$:= population-specific error term

[14] Shang, H. L. (2016) Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method, *Annals of Applied Statistics*, **10**(3), 1639-1672

Multilevel FPC decomposition

1. Common trend $U_t(u)$ can be a simple average of $\beta_t^1(u)$ & $\beta_t^2(u)$

Multilevel FPC decomposition

1. Common trend $U_t(u)$ can be a simple average of $\beta_t^1(u)$ & $\beta_t^2(u)$
2. Common trend & population-specific trend can be modelled via a two-stage FPCA

$$U_t(u) = \sum_{k=1}^{K} \widehat{\gamma}_{t,k}\widehat{\phi}_k(u)$$

$$R_t^j(u) = \sum_{l=1}^{L} \widehat{\eta}_{t,l}^j\widehat{\psi}_l^j(u),$$

Multilevel FPC decomposition

1. Common trend $U_t(u)$ can be a simple average of $\beta_t^1(u)$ & $\beta_t^2(u)$
2. Common trend & population-specific trend can be modelled via a two-stage FPCA

$$U_t(u) = \sum_{k=1}^{K} \widehat{\gamma}_{t,k}\widehat{\phi}_k(u)$$

$$R_t^j(u) = \sum_{l=1}^{L} \widehat{\eta}_{t,l}^j\widehat{\psi}_l^j(u),$$

3. K & L are determined by ER tests, respectively

Within-cluster variability

1. Let $\lambda_k := k^{\text{th}}$ eigenvalue of common trend

Within-cluster variability

1. Let $\lambda_k :=$ k^{th} eigenvalue of common trend
2. Let $\lambda_l^j :=$ l^{th} eigenvalue of population-specific trend

Within-cluster variability

1. Let $\lambda_k := k^{\text{th}}$ eigenvalue of common trend
2. Let $\lambda_l^j := l^{\text{th}}$ eigenvalue of population-specific trend
3. A possible measure of within-cluster variability is

$$\frac{\sum_{k=1}^{\infty}\lambda_k}{\sum_{k=1}^{\infty}\lambda_k+\sum_{l=1}^{\infty}\lambda_l^j} = \frac{\int_{\mathcal{I}}\text{Var}[\boldsymbol{R}(u)]du}{\int_{\mathcal{I}}\text{Var}[\boldsymbol{R}(u)]du+\int_{\mathcal{I}}\text{Var}[\boldsymbol{U}^j(u)]du},$$

Within-cluster variability

1. Let $\lambda_k := k^{\text{th}}$ eigenvalue of common trend
2. Let $\lambda_l^j := l^{\text{th}}$ eigenvalue of population-specific trend
3. A possible measure of within-cluster variability is

$$\frac{\sum_{k=1}^{\infty}\lambda_k}{\sum_{k=1}^{\infty}\lambda_k+\sum_{l=1}^{\infty}\lambda_l^j} = \frac{\int_{\mathcal{I}}\text{Var}[\boldsymbol{R}(u)]du}{\int_{\mathcal{I}}\text{Var}[\boldsymbol{R}(u)]du+\int_{\mathcal{I}}\text{Var}[\boldsymbol{U}^j(u)]du},$$

4. When common factor can explain main mode of total variability, value of within-cluster variability is close to 1

Forecast FPC scores

1. PC scores can be obtained via a univariate or multivariate time series forecasting method

Forecast FPC scores

1. PC scores can be obtained via a univariate or multivariate time series forecasting method
2. For modelling presence of *nonstationarity*, consider four univariate time series forecasting methods

Forecast FPC scores

1. PC scores can be obtained via a univariate or multivariate time series forecasting method
2. For modelling presence of *nonstationarity*, consider four univariate time series forecasting methods
3. Use a univariate time series forecasting method, obtain h-steps-ahead forecast $\widehat{\gamma}^{j}_{n+h|n,\ell}$ of ℓ^{th} PC score

Forecast FPC scores

1. PC scores can be obtained via a univariate or multivariate time series forecasting method
2. For modelling presence of *nonstationarity*, consider four univariate time series forecasting methods
3. Use a univariate time series forecasting method, obtain h-steps-ahead forecast $\widehat{\gamma}^j_{n+h|n,\ell}$ of ℓ^{th} PC score
4. Conditioning on estimated PCs & observed data, forecast of $\beta^j_{n+h}(u)$ is

$$\widehat{\beta}^j_{n+h|n}(u) = \sum_{\ell=1}^{L_j} \widehat{\gamma}_{n+h|n,\ell}\widehat{\phi}^j_{\ell}(u)$$

CoDa back-transformation

1. Transform back to compositional data, taking inverse centered log-ratio transformation

$$\widehat{s}^{j}_{n+h|n}(u) = \frac{\exp\left[\widehat{\beta}^{j}_{n+h|n}(u)\right]}{\int_{u=0}^{110} \exp\left[\widehat{\beta}^{j}_{n+h|n}(u)\right] du}$$

CoDa back-transformation

1. Transform back to compositional data, taking inverse centered log-ratio transformation

$$\widehat{s}^{j}_{n+h|n}(u) = \frac{\exp\left[\widehat{\beta}^{j}_{n+h|n}(u)\right]}{\int_{u=0}^{110}\exp\left[\widehat{\beta}^{j}_{n+h|n}(u)\right]du}$$

2. Add back geometric means, to obtain forecasts of density function

$$\widehat{f}_{n+h|n}(u) = \frac{\widehat{s}^{j}_{n+h|n}(u)\alpha^{j}_{n}(u)}{\int_{u=0}^{110}\widehat{s}^{j}_{n+h|n}(u)\alpha^{j}_{n}(u)du}$$

where $\alpha^{j}_{n}(u)$:= geometric mean function

CoDa back-transformation

1. Transform back to compositional data, taking inverse centered log-ratio transformation

$$\widehat{s}^j_{n+h|n}(u) = \frac{\exp\left[\widehat{\beta}^j_{n+h|n}(u)\right]}{\int_{u=0}^{110}\exp\left[\widehat{\beta}^j_{n+h|n}(u)\right]du}$$

2. Add back geometric means, to obtain forecasts of density function

$$\widehat{f}_{n+h|n}(u) = \frac{\widehat{s}^j_{n+h|n}(u)\alpha^j_n(u)}{\int_{u=0}^{110}\widehat{s}^j_{n+h|n}(u)\alpha^j_n(u)du}$$

where $\alpha^j_n(u)$:= geometric mean function

3. Multiply by radix (10^5)

Nonparametric bootstrap

Bootstrap takes into account two sources of uncertainty[15]

- model residuals in PC decomposition of unconstrained data
- forecast errors between estimated scores from holdout sample & forecast PC scores

[15] R. J. Hyndman and H. L. Shang (2009) Forecasting functional time series (with discussion), *Journal of the Korean Statistical Society*, **38**(3), 199-221

Bootstrap forecast errors of PC scores

1. Using a univariate time series forecasting method, obtain h-step-ahead forecasts for PC scores $\{\widehat{\gamma}^j_{1,\ell}, \dots, \widehat{\gamma}^j_{n,\ell}\}$

Bootstrap forecast errors of PC scores

1. Using a univariate time series forecasting method, obtain h-step-ahead forecasts for PC scores $\{\widehat{\gamma}^j_{1,\ell},\dots,\widehat{\gamma}^j_{n,\ell}\}$
2. Let h-step-ahead forecast errors be

$$\zeta^j_{t,h,\ell} = \widehat{\gamma}^j_{t,\ell} - \widehat{\gamma}^j_{t|t-h,\ell}, \qquad t = h+1,\dots,n$$

Bootstrap forecast errors of PC scores

1. Using a univariate time series forecasting method, obtain h-step-ahead forecasts for PC scores $\{\widehat{\gamma}_{1,\ell}^{j}, \dots, \widehat{\gamma}_{n,\ell}^{j}\}$
2. Let h-step-ahead forecast errors be

$$\zeta_{t,h,\ell}^{j} = \widehat{\gamma}_{t,\ell}^{j} - \widehat{\gamma}_{t|t-h,\ell}^{j}, \qquad t = h+1, \dots, n$$

3. By *sampling with replacement* from $\widehat{\zeta}_{t,h,\ell}^{j}$, obtain $\widehat{\zeta}_{*,h,\ell}^{j,b}$

Bootstrap forecast errors of PC scores

1. Using a univariate time series forecasting method, obtain h-step-ahead forecasts for PC scores $\{\widehat{\gamma}_{1,\ell}^{j},\dots,\widehat{\gamma}_{n,\ell}^{j}\}$
2. Let h-step-ahead forecast errors be

$$\zeta_{t,h,\ell}^{j} = \widehat{\gamma}_{t,\ell}^{j} - \widehat{\gamma}_{t|t-h,\ell}^{j}, \qquad t = h+1,\dots,n$$

3. By *sampling with replacement* from $\widehat{\zeta}_{t,h,\ell}^{j}$, obtain $\widehat{\zeta}_{*,h,\ell}^{j,b}$
4. Obtain bootstrap samples of $\gamma_{n+h,\ell}$:

$$\widehat{\gamma}_{n+h|n,\ell}^{j,b} = \widehat{\gamma}_{n+h|n,\ell} + \widehat{\zeta}_{*,h,\ell}^{j,b}, \qquad b = 1,\dots,10,000$$

Bootstrap model residuals

1. When first L_j PCs capture main mode of variation, model residuals should be random noise

Bootstrap model residuals

1. When first L_j PCs capture main mode of variation, model residuals should be random noise
2. Bootstrap model residuals by *sampling with replacement* from $\{\widehat{\omega}_1^j(u), \dots, \widehat{\omega}_n^j(u)\}$

Bootstrap model residuals

1. When first L_j PCs capture main mode of variation, model residuals should be random noise
2. Bootstrap model residuals by *sampling with replacement* from $\{\widehat{\omega}_1^j(u), \dots, \widehat{\omega}_n^j(u)\}$
3. Adding two components of variability, obtain B variants for $\beta_{n+h}^j(u)$:

$$\widehat{\beta}_{n+h|n}^{j,b}(u) = \sum_{\ell=1}^{L_j} \widehat{\gamma}_{n+h|n,\ell}^{j,b} \widehat{\phi}_\ell^j(u) + \widehat{\omega}_{n+h}^{j,b}(u)$$

Bootstrap model residuals

1. When first L_j PCs capture main mode of variation, model residuals should be random noise
2. Bootstrap model residuals by *sampling with replacement* from $\{\widehat{\omega}_1^j(u),\dots,\widehat{\omega}_n^j(u)\}$
3. Adding two components of variability, obtain B variants for $\beta_{n+h}^j(u)$:

$$\widehat{\beta}_{n+h|n}^{j,b}(u) = \sum_{\ell=1}^{L_j}\widehat{\gamma}_{n+h|n,\ell}^{j,b}\widehat{\phi}_\ell^j(u) + \widehat{\omega}_{n+h}^{j,b}(u)$$

4. Via CoDa back-transformation, obtain bootstrap forecasts of $f_{n+h}^j(u)$ & take quantiles

CoDa with multilevel functional time series method

1. Number of retained components, K & L are determined by ER tests in common & population-specific trends

CoDa with multilevel functional time series method

1. Number of retained components, K & L are determined by ER tests in common & population-specific trends
2. $K = 1$, $L_{\text{female}} = 2$ and $L_{\text{male}} = 3$

CoDa with multilevel functional time series method

1. Number of retained components, K & L are determined by ER tests in common & population-specific trends
2. $K = 1$, $L_{\text{female}} = 2$ and $L_{\text{male}} = 3$
3. By computing within-cluster variability, common trend accounts for

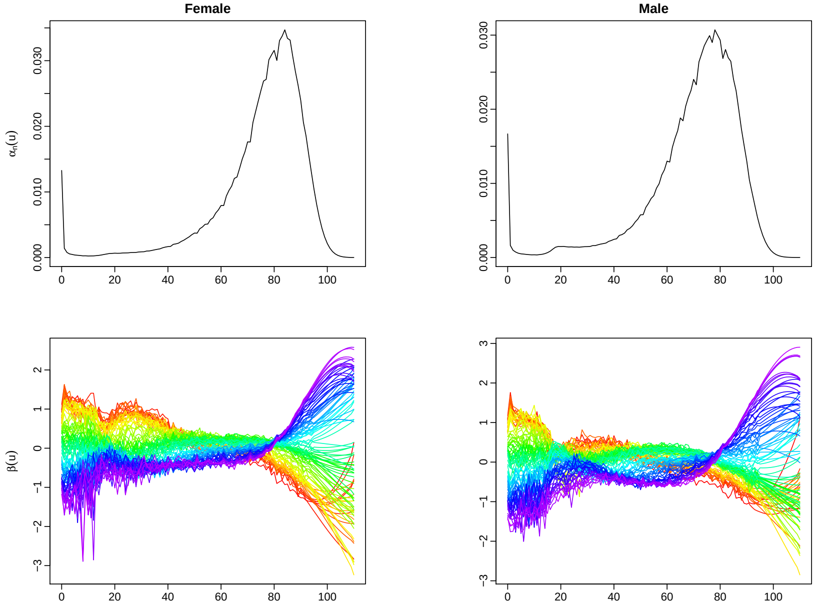
CoDa with multilevel functional time series method

1. Number of retained components, K & L are determined by ER tests in common & population-specific trends
2. $K = 1$, $L_{\text{female}} = 2$ and $L_{\text{male}} = 3$
3. By computing within-cluster variability, common trend accounts for
   - 93.02% of total variation for females

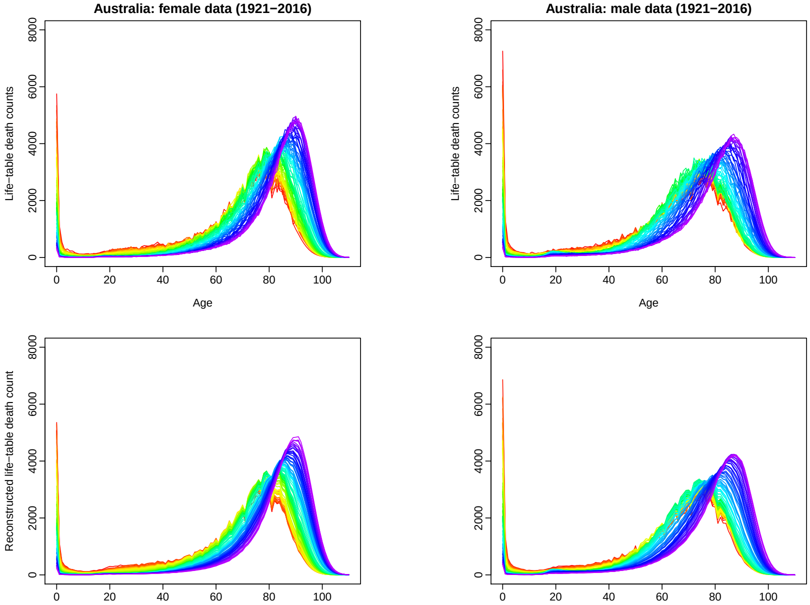
CoDa with multilevel functional time series method

1. Number of retained components, K & L are determined by ER tests in common & population-specific trends
2. $K = 1$, $L_{\text{female}} = 2$ and $L_{\text{male}} = 3$
3. By computing within-cluster variability, common trend accounts for
   - 93.02% of total variation for females
   - 91.45% of total variation for males

Geometric mean & unconstrained data



Actual & reconstructed life table death counts



Forecast life table death counts

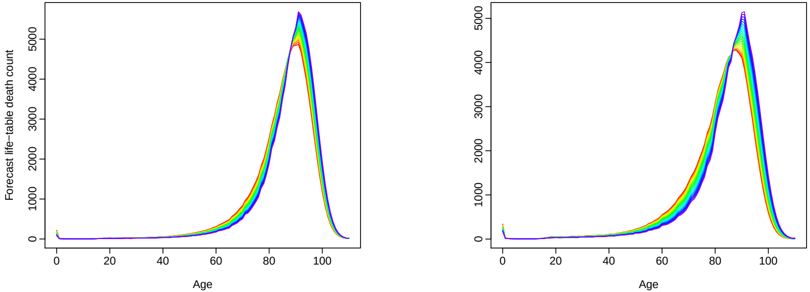


Figure: *CoDa + MLFTS + ETS*

Forecast evaluation setup

1. Using first 76 observations from 1921 to 1996 in Australian females & males, produce 1- to 20-step-ahead forecasts

Forecast evaluation setup

1. Using first 76 observations from 1921 to 1996 in Australian females & males, produce 1- to 20-step-ahead forecasts
2. Via an *expanding window* approach, re-estimate parameters using first 77 observations from 1921 to 1997, produce 1- to 19-step-ahead forecasts

Forecast evaluation setup

1. Using first 76 observations from 1921 to 1996 in Australian females & males, produce 1- to 20-step-ahead forecasts
2. Via an *expanding window* approach, re-estimate parameters using first 77 observations from 1921 to 1997, produce 1- to 19-step-ahead forecasts
3. Iterate this process by increasing a year until reaching end of data period in 2016

Forecast evaluation setup

1. Using first 76 observations from 1921 to 1996 in Australian females & males, produce 1- to 20-step-ahead forecasts
2. Via an *expanding window* approach, re-estimate parameters using first 77 observations from 1921 to 1997, produce 1- to 19-step-ahead forecasts
3. Iterate this process by increasing a year until reaching end of data period in 2016
4. Produces 20 one-step-ahead forecasts, 19 two-step-ahead forecasts, . . . , one 20-step-ahead forecast

Forecast evaluation setup

1. Using first 76 observations from 1921 to 1996 in Australian females & males, produce 1- to 20-step-ahead forecasts
2. Via an *expanding window* approach, re-estimate parameters using first 77 observations from 1921 to 1997, produce 1- to 19-step-ahead forecasts
3. Iterate this process by increasing a year until reaching end of data period in 2016
4. Produces 20 one-step-ahead forecasts, 19 two-step-ahead forecasts, . . . , one 20-step-ahead forecast
5. Compare these forecasts with holdout samples between 1997 & 2016

Point forecast error criterion

Consider MAPE, it can be defined as

$$\text{MAPE}(h) = \frac{1}{111 \times (21-h)} \sum_{\xi=h}^{20} \sum_{i=1}^{111} \left| \frac{f_{n+\xi}(u_i) - \widehat{f}_{n+\xi|n}(u_i)}{f_{n+\xi}(u_i)} \right| \times 100,$$

- $f_{n+\xi}(u_i)$:= holdout sample for ${u_i}^{\text{th}}$ age & ξ^{th} forecasting year
- $\widehat{f}_{n+\xi|n}(u_i)$:= point forecasts

Density evaluation metric – Kullback-Leibler divergence

Discrete version of Kullback-Leibler divergence

$$\begin{aligned}\mathsf{KLD}(h) =& D_{\mathsf{KL}}\left[f_{n+\xi}(u_i)||\widehat{f}_{n+\xi|n}(u_i)\right] + D_{\mathsf{KL}}\left[\widehat{f}_{n+\xi|n}(u_i)||f_{n+\xi}(u_i)\right] \\ =& \frac{1}{111\times(21-h)}\sum_{\xi=h}^{20}\sum_{i=1}^{111} f_{n+\xi}(u_i)\cdot[\ln f_{n+\xi}(u_i) - \ln \widehat{f}_{n+\xi|n}(u_i)] + \\ & \frac{1}{111\times(21-h)}\sum_{\xi=h}^{20}\sum_{i=1}^{111} \widehat{f}_{n+\xi|n}(u_i)\cdot[\ln \widehat{f}_{n+\xi|n}(u_i) - \ln f_{n+\xi}(u_i)]\end{aligned}$$

Density evaluation metric – Jensen-Shannon divergence

Jensen-Shannon divergence

$$\text{JSD}(h) = \frac{1}{2}D_{\text{KL}}\left[f_{n+\xi}(u_i)||\delta_{n+\xi}(u_i)\right] + \frac{1}{2}D_{\text{KL}}\left[\widehat{f}_{n+\xi|n}(u_i)||\delta_{n+\xi}(u_i)\right]$$

- $\delta_{n+\xi}(u_i)$:= a common quantity between $f_{n+\xi}(u_i)$ & $\widehat{f}_{n+\xi|n}(u)$
- Consider simple mean $\delta_{n+\xi}(u_i) = \frac{1}{2}\left[f_{n+\xi}(u_i) + \widehat{f}_{n+\xi|n}(u_i)\right]$
- Consider geometric mean $\delta_{n+\xi}(u_i) = \sqrt{f_{n+\xi}(u_i)\widehat{f}_{n+\xi|n}(u_i)}$

Interval forecast error criterion – interval score

A scoring rule for prediction intervals $f_{n+\xi}(u_i)$ is

$$
\begin{aligned}
S_{\nu,\xi}\left[\widehat{f}^{\text{lb}}_{n+\xi}(u_i), \widehat{f}^{\text{ub}}_{n+\xi}(u_i), f_{n+\xi}(u_i)\right] = \left[\widehat{f}^{\text{ub}}_{n+\xi}(u_i) - \widehat{f}^{\text{lb}}_{n+\xi}(u_i)\right] \\
+\frac{2}{\nu}[\widehat{f}^{\text{lb}}_{n+\xi}(u_i) - f_{n+\xi}(u_i)]\mathbb{1}\{f_{n+\xi}(u_i) < \widehat{f}^{\text{lb}}_{n+\xi}(u_i)\} \\
+\frac{2}{\nu}[f_{n+\xi}(u_i) - \widehat{f}^{\text{ub}}_{n+\xi}(u_i)]\mathbb{1}\{f_{n+\xi}(u_i) > \widehat{f}^{\text{ub}}_{n+\xi}(u_i)\}
\end{aligned}
$$

- $\mathbb{1}\{\cdot\}$:= binary indicator function
- ν:= level of significance, customarily $\nu = 0.2$ or 0.05

Interval forecast error criterion – interval score

1. Interval score rewards a *narrow* prediction interval, if and only if true observation *lies within* prediction interval

Interval forecast error criterion – interval score

1. Interval score rewards a *narrow* prediction interval, if and only if true observation *lies within* prediction interval
2. For different ages & years in forecasting period, mean interval score is

$$\overline{S}_{\nu}(h) = \frac{1}{111 \times (21-h)} \sum_{\xi=h}^{20} \sum_{i=1}^{111} S_{\nu,\xi} \left[\widehat{f}^{\,\text{lb}}_{n+\xi}(u_i), \widehat{f}^{\,\text{ub}}_{n+\xi}(u_i), f_{n+\xi}(u_i)\right]$$

Interval forecast error criterion – interval score

1. Interval score rewards a *narrow* prediction interval, if and only if true observation *lies within* prediction interval
2. For different ages & years in forecasting period, mean interval score is

$$\overline{S}_{\nu}(h) = \frac{1}{111 \times (21-h)} \sum_{\xi=h}^{20} \sum_{i=1}^{111} S_{\nu,\xi} \left[\widehat{f}^{\text{lb}}_{n+\xi}(u_i), \widehat{f}^{\text{ub}}_{n+\xi}(u_i), f_{n+\xi}(u_i) \right]$$

3. Averaged over all horizons, obtain overall mean interval score

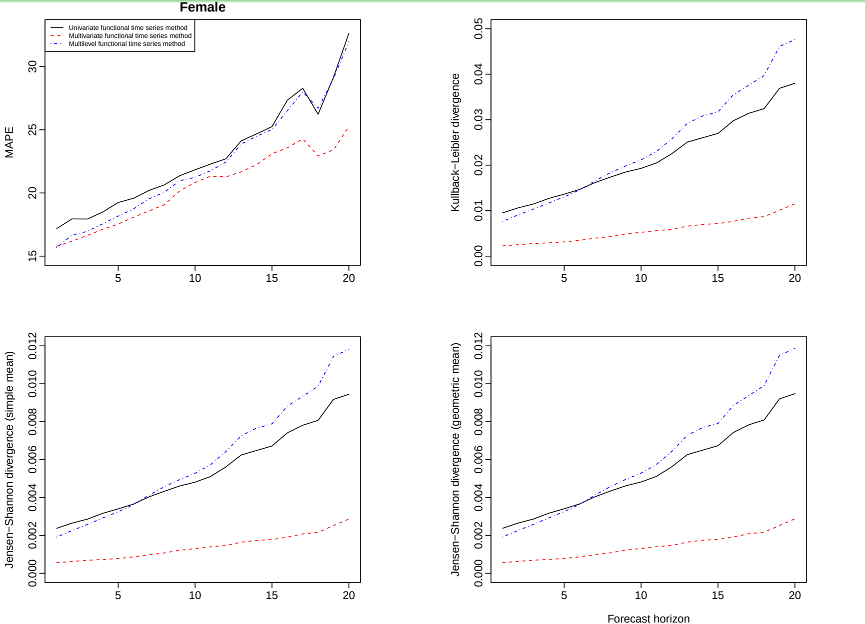
$$\overline{S}_{\nu} = \frac{1}{20} \sum_{h=1}^{20} \overline{S}_{\nu}(h)$$

Point forecast accuracy – Australian female data

Forecasting method	Modelling method	Female			
		MAPE	KLD	JSD^s	JSD^g
ARIMA	FTS	21.7888	0.0177	0.0044	0.0044
	MFTS	28.5893	0.0064	0.0016	0.0016
	MLFTS	24.3259	0.0214	0.0053	0.0054
ETS	FTS	21.7890	0.0177	0.0044	0.0044
	MFTS	28.9389	0.0069	0.0017	0.0017
	MLFTS	21.5710	0.0165	0.0041	0.0041
RW	FTS	34.8394	0.0487	0.0121	0.0121
	MFTS	32.8397	0.0227	0.0057	0.0057
	MLFTS	37.2417	0.0478	0.0118	0.0119
RWD	FTS	22.8552	0.0217	0.0054	0.0054
	MFTS	**20.4531**	**0.0057**	**0.0014**	**0.0014**
	MLFTS	22.2698	0.0245	0.0061	0.0061

CoDa with multivariate functional time series & RWD forecasting method

CoDa + RWD forecasting method

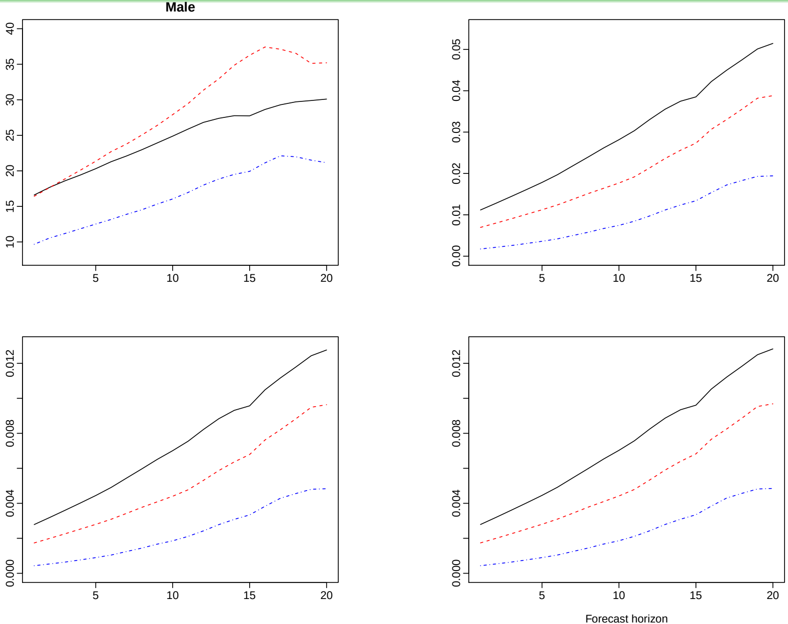


Point forecast accuracy – Australian male data

Forecasting method	Modelling method	Male			
		MAPE	KLD	JSD^s	JSD^g
ARIMA	FTS	30.0997	0.0517	0.0128	0.0129
	MFTS	29.1165	0.0218	0.0054	0.0055
	MLFTS	23.5167	0.0277	0.0069	0.0069
ETS	FTS	24.5508	0.0302	0.0075	0.0075
	MFTS	28.3270	0.0207	0.0051	0.0052
	MLFTS	**16.4944**	**0.0093**	**0.0023**	**0.0023**
RW	FTS	42.6095	0.0787	0.0193	0.0196
	MFTS	44.4279	0.0664	0.0163	0.0166
	MLFTS	35.3865	0.0453	0.0111	0.0113
RWD	FTS	29.9539	0.0565	0.0139	0.0141
	MFTS	28.7224	0.0440	0.0109	0.0110
	MLFTS	22.8291	0.0251	0.0062	0.0063

CoDa with multilevel functional time series & ETS forecasting method

CoDa + ETS forecasting method

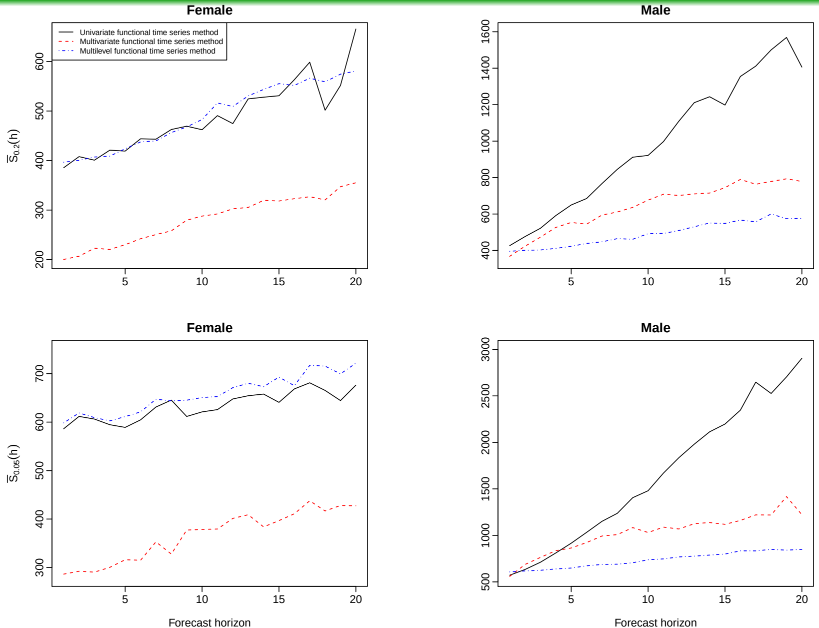


Interval forecast accuracy

Forecasting method	Modeling method	Female $\overline{S}_{0.2}$	Female $\overline{S}_{0.05}$	Male $\overline{S}_{0.2}$	Male $\overline{S}_{0.05}$
ARIMA	FTS	424.5165	662.1788	989.7989	1643.7430
	MFTS	313.0574	433.4390	644.4068	1026.4499
	MLFTS	444.0410	676.9822	**492.2062**	**736.0020**
ETS	FTS	494.4926	875.7863	587.8981	741.4829
	MFTS	354.2009	583.3586	624.9392	908.8275
	MLFTS	516.4761	1015.2346	538.1190	868.0729
RW	FTS	487.1864	633.2024	1248.2391	2499.3450
	MFTS	**280.4400**	**385.3941**	1258.4402	3024.3323
	MLFTS	490.3348	657.3637	611.5622	807.2828
RWD	FTS	474.3986	830.2068	580.0019	838.8614
	MFTS	378.0343	696.1440	620.8607	822.4041
	MLFTS	550.7896	1130.3849	595.7285	1103.8458

CoDa with multivariate functional time series & RW forecasting method
CoDa with multilevel functional time series & ARIMA forecasting method

Interval score



Fixed-term annuities

1. Lifetime annuities, locked in for life, have been perceived to deliver poor value for money[16]

[16] Cannon, E. and Tonks, I. (2008), Annuity Markets, Oxford University Press, Oxford

Fixed-term annuities

1. Lifetime annuities, locked in for life, have been perceived to deliver poor value for money[16]
2. Fixed-term annuities pay a guaranteed level of income for a period of maturity; it provides flexibility in allowing purchaser buy a deferred annuity at a later stage

[16] Cannon, E. and Tonks, I. (2008), Annuity Markets, Oxford University Press, Oxford

Forecast age-specific life table death counts (2017 – 2066)

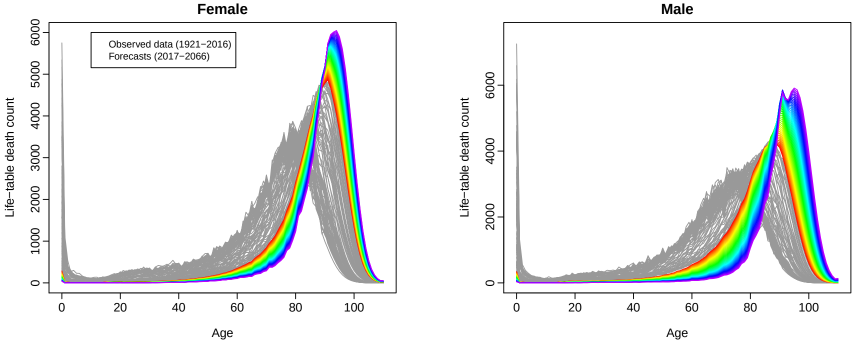


Figure: *For females, use CoDa + MFTS + RWD; For males, use CoDa + MLFTS + ETS*

Survival probabilities

1. Adopt a cohort approach[17] to calculation of survival probabilities
2. τ-year survival probability of a person aged x at $t=0$ is

$$
{}_{\tau}p_x = \prod_{\eta=1}^{\tau} p_{x+\eta-1} = \prod_{\eta=1}^{\tau}(1-q_{x+\eta-1}) = \prod_{\eta=1}^{\tau}\Big(1-\frac{d_{x+\eta-1}}{l_{x+\eta-1}}\Big)
$$

- $d_{x+\eta-1}$:= number of death counts between ages $x+\eta-1$ & $x+\eta$
- $l_{x+\eta-1}$:= number of lives alive at age $x+\eta-1$

[17] Dickson, D. C. M., Hardy, M. R. and Waters, H. R. (2009), Actuarial Mathematics for Life Contingent Risks, Cambridge University Press, Cambridge

Annuity price

1. Annuity price with maturity T year depends on

Annuity price

1. Annuity price with maturity T year depends on
   - zero-coupon bond price

Annuity price

1. Annuity price with maturity T year depends on
   - zero-coupon bond price
   - future mortality

Annuity price

1. Annuity price with maturity T year depends on
   - zero-coupon bond price
   - future mortality
2. Annuity written for an x-year-old with benefit \$1 dollar per year is

$$a_x^T = \sum_{\tau=1}^{T} B(0,\tau)\,{}_{\tau}p_x,$$

Annuity price

1. Annuity price with maturity T year depends on
   - zero-coupon bond price
   - future mortality
2. Annuity written for an x-year-old with benefit \$1 dollar per year is

$$a_x^T = \sum_{\tau=1}^{T} B(0,\tau)_{\tau}p_x,$$

- $B(0,\tau)$:= τ-year bond price

Annuity price

1. Annuity price with maturity T year depends on
   - zero-coupon bond price
   - future mortality
2. Annuity written for an x-year-old with benefit \$1 dollar per year is

$$a_x^T = \sum_{\tau=1}^{T} B(0,\tau)\,{}_{\tau}p_x,$$

- $B(0,\tau)$:= τ-year bond price
- ${}_{\tau}p_x$:= survival probability

Bond price calculation

1. Zero-coupon bond[18] is

$$B(0, \tau) = \exp^{-w\tau}$$

where w alters for different maturities T

[18] A. J. G. Cairns (2004). Interest Rate Models: An Introduction, Princeton University Press, Princeton, New Jersey

Bond price calculation

1. Zero-coupon bond[18] is

$$B(0, \tau) = \exp^{-w\tau}$$

   where w alters for different maturities T

2. Australian government bond yields for various maturities

	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 30$
Yield (w)	0.480%	1.069%	1.542%	1.794%	1.986%

Table: *Retrieved from https://au.investing.com/rates-bonds/*

[18] A. J. G. Cairns (2004). Interest Rate Models: An Introduction, Princeton University Press, Princeton, New Jersey

Point forecasts of annuity prices

Sex	Age	$T=5$	$T=10$	$T=15$	$T=20$	$T=30$
Female	60	4.8710	9.2088	12.7712	15.6813	19.5944
	65	4.8467	9.1048	12.5139	15.1371	17.9302
	70	4.8023	8.9198	12.0261	14.1290	15.5642
	75	4.7228	8.5484	11.0908	12.3804	12.7786
	80	4.5333	7.7887	9.4198	9.8924	9.8704
	85	4.1748	6.4772	7.1607	7.2023	NA
	90	3.4966	4.7080	4.8372	4.7995	NA
	95	2.6411	3.0746	3.0607	NA	NA
	100	1.7774	1.8806	NA	NA	NA
	105	1.1648	NA	NA	NA	NA
Male	60	4.8351	9.0754	12.4892	15.2103	18.7021
	65	4.7997	8.9371	12.1689	14.5748	17.0265
	70	4.7457	8.7200	11.6208	13.5148	14.7944
	75	4.6534	8.3061	10.6476	11.8059	12.2007
	80	4.4427	7.5268	9.0348	9.5010	9.5014
	85	4.0964	6.3062	7.0032	7.0693	NA
	90	3.4573	4.7212	4.8913	4.8589	NA
	95	2.7505	3.2745	3.272	NA	NA
	100	1.9278	2.0754	NA	NA	NA
	105	1.3020	NA	NA	NA	NA

Interval forecasts of annuity prices

Sex	Age	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 30$
Female	60	(4.899, 4.920)	(9.311, 9.399)	(12.989, 13.204)	(16.048, 16.478)	(20.295, 21.549)
	65	(4.881, 4.916)	(9.231, 9.376)	(12.778, 13.137)	(15.570, 16.285)	(18.628, 20.457)
	70	(4.845, 4.904)	(9.074, 9.321)	(12.338, 12.948)	(14.599, 15.778)	(16.151, 18.419)
	75	(4.775, 4.878)	(8.728, 9.164)	(11.408, 12.446)	(12.783, 14.524)	(13.197, 15.563)
	80	(4.590, 4.787)	(7.948, 8.721)	(9.644, 11.169)	(10.121, 12.178)	(10.093, 12.307)
	85	(4.216, 4.556)	(6.553, 7.633)	(7.228, 8.898)	(7.262, 9.106)	NA
	90	(3.492, 3.960)	(4.669, 5.795)	(4.788, 6.136)	(4.752, 6.103)	NA
	95	(2.571, 3.130)	(2.971, 3.888)	(2.957, 3.910)	NA	NA
	100	(1.612, 2.266)	(1.706, 2.503)	NA	NA	NA
	105	(0.749, 2.046)	NA	NA	NA	NA
Male	60	(4.861, 4.929)	(9.147, 9.433)	(12.582, 13.287)	(15.270, 16.649)	(18.440, 22.283)
	65	(4.815, 4.929)	(8.948, 9.432)	(12.115, 13.282)	(14.366, 16.628)	(16.334, 21.980)
	70	(4.730, 4.927)	(8.627, 9.426)	(11.343, 13.259)	(12.945, 16.535)	(13.754, 21.079)
	75	(4.602, 4.925)	(8.054, 9.406)	(10.055, 13.162)	(10.873, 16.164)	(11.012, 19.065)
	80	(4.305, 4.910)	(7.026, 9.309)	(8.110, 12.759)	(8.327, 15.062)	(8.273, 16.111)
	85	(3.842, 4.853)	(5.573, 8.930)	(5.942, 11.608)	(5.929, 12.652)	NA
	90	(3.068, 4.612)	(3.879, 7.878)	(3.926, 9.186)	(3.896, 9.335)	NA
	95	(2.249, 4.078)	(2.521, 5.908)	(2.503, 6.177)	NA	NA
	100	(1.507, 3.006)	(1.571, 3.641)	NA	NA	NA
	105	(0.927, 2.104)	NA	NA	NA	NA

Future research

1. A robust CoDa[19]

---

[19] Filzmoser, P., Hron, K. and Reimann, C. (2009), 'Principal component analysis for compositional data with outliers', *Environmetrics*, **20**(6), 621-632

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

[21] J. Li (2013) A Poisson common factor model for projecting mortality and life expectancy jointly for females and males, *Population Studies*, **67**(1), 111-126

Future research

1. A robust CoDa[19]
2. Consider other joint modelling techniques, such as product-ratio method[20] and Poisson common factor model[21]

[19] Filzmoser, P., Hron, K. and Reimann, C. (2009), 'Principal component analysis for compositional data with outliers', *Environmetrics*, **20**(6), 621-632

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

[21] J. Li (2013) A Poisson common factor model for projecting mortality and life expectancy jointly for females and males, *Population Studies*, **67**(1), 111-126

Future research

1. A robust CoDa[19]
2. Consider other joint modelling techniques, such as product-ratio method[20] and Poisson common factor model[21]
3. Consider other types of annuity prices, such as a deferred annuity

[19] Filzmoser, P., Hron, K. and Reimann, C. (2009), 'Principal component analysis for compositional data with outliers', *Environmetrics*, **20**(6), 621-632

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

[21] J. Li (2013) A Poisson common factor model for projecting mortality and life expectancy jointly for females and males, *Population Studies*, **67**(1), 111-126

Future research

1. A robust CoDa[19]
2. Consider other joint modelling techniques, such as product-ratio method[20] and Poisson common factor model[21]
3. Consider other types of annuity prices, such as a deferred annuity
4. Consider cohort life table death counts

[19] Filzmoser, P., Hron, K. and Reimann, C. (2009), 'Principal component analysis for compositional data with outliers', *Environmetrics*, **20**(6), 621-632

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

[21] J. Li (2013) A Poisson common factor model for projecting mortality and life expectancy jointly for females and males, *Population Studies*, **67**(1), 111-126

Thank you

RG: `https://www.researchgate.net/profile/Han_Lin_Shang`
Email: hanlin.shang@mq.edu.au