



# City Research Online

## City St George's, University of London

**Citation:** Shang, H.L., Haberman, S. & Xu, R. (2022). Multi-population modelling and forecasting life-table death counts. *Insurance: Mathematics and Economics*, 106, pp. 239-253. doi: 10.1016/j.insmatheco.2022.07.002

This is the accepted version of the paper.

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

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/28373/>

**Link to published version:** <https://doi.org/10.1016/j.insmatheco.2022.07.002>

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

# Multi-population modelling and forecasting life-table death counts

Han Lin Shang  \*

Department of Actuarial Studies and Business Analytics  
Macquarie University

Steven Haberman 

Bayes Business School  
City, University of London

Ruofan Xu 

Department of Econometrics and Business Statistics  
Monash University

## Abstract

When modelling the age distribution of death counts for multiple populations, we ought to consider three features: (1) how to incorporate any possible correlation among multiple populations to improve point and interval forecast accuracy through multi-population joint modelling, (2) how to forecast age distribution of death counts so that the forecasts are non-negative and have a constrained integral. (3) how to construct a prediction interval that is well-calibrated in terms of coverage. Within the framework of compositional data analysis, we apply a log-ratio transform to transform a constrained space into an unconstrained space. We apply multivariate and multilevel functional time series methods to forecast period life-table death counts in the unconstrained space. Through the inverse log-ratio transformation, the forecast period life-table death counts are obtained. Using the age-specific period life-table death counts in England & Wales and Sweden obtained from [Human Mortality Database \(2022\)](#), we investigate one-step-ahead to 30-step-ahead point and interval forecast accuracies of the proposed models and make our recommendations.

Keywords: age distribution of death counts; compositional data analysis; functional principal component analysis; log-ratio transformation; multivariate and multilevel functional principal component regression

JEL code: C14, C32, J11

---

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

# 23 1 Introduction

24 Actuaries and demographers have been interested in developing methods for mortality fore-  
25 casting for many years. In particular, actuaries have produced mortality forecasts since the  
26 beginning of the 20<sup>th</sup> century in response to the adverse financial effects of mortality improve-  
27 ments overtime on life annuities and pensions (Pollard 1987). In the literature on human  
28 mortality, three functions are generally studied: hazard function, survival function and proba-  
29 bility density function. These three functions are complementary; knowing any one of them  
30 allows us to uniquely derive the other two (for detail on the life table and its indicators, see  
31 Preston et al. (2001), Chapter 3, or Dickson et al. (2009), Chapters 2–3).

32 Several authors have proposed new approaches for forecasting the age-specific hazard  
33 function (i.e., central mortality rates) using statistical models (see, e.g., Booth 2006, Booth and  
34 Tickle 2008, Cairns et al. 2008, Shang et al. 2011, for reviews). Instead of modelling central  
35 mortality rates, we consider modelling the life-table death distribution (see, e.g., Basellini et al.  
36 2020). Observed over time, we could model and forecast a redistribution of the density of  
37 life-table deaths, where deaths at younger ages are shifted towards older ages. In addition  
38 to providing a very informative description of the mortality experience of a population, the  
39 life-table death counts yield readily available information on ‘central longevity indicators’ (i.e.,  
40 mean, median and mode age at death, see Cheung et al. 2005, Canudas-Romo 2010), as well as  
41 lifespan variability (for example, standard deviation or inter-quartile range, see Robine 2001,  
42 Vaupel et al. 2011, Horiuchi et al. 2013, van Raalte and Caswell 2013, van Raalte et al. 2014,  
43 Aburto and van Raalte 2018).

44 Within the analysis of life-table death counts, we may consider both cohort and period life  
45 tables. The period life table represents the mortality conditions in a period (see also Oeppen  
46 2008, Bergeron-Boucher et al. 2017, 2018). The cohort life table depicts the life history of a specific  
47 group of individuals. Still, it is dependent on projected mortality rates for those cohorts born  
48 more recently than around 100 years ago. Cohort mortality developments are observed, and  
49 they may differ from those of the synthetic cohorts assumed in period life tables (Goldstein and  
50 Wachter 2006). Unlike the period data, the cohort data are partially complete, which presents a  
51 challenge.

52 Many methods proposed for completing cohort-based life tables depend on parametric  
53 models, which are fitted to the incomplete cohort data and then extrapolated beyond the  
54 truncation age. As noted by Booth and Tickle (2008), *P*-spline regression methods seem to  
55 provide a promising approach, and, in this context, one could apply the method of Rizzi et al.  
56 (2021) who have proposed a penalised composite link model to complete the mortality profile.

57 With the completed mortality profile, statistical methods used for modelling and forecasting  
58 period life-table death counts are applicable. Whether we are considering a period or cohort life  
59 table, the age-specific death counts can be naturally viewed as a probability density function,  
60 an example of compositional data. Without deviating from our methodological contributions,  
61 we focus on period life table results and present cohort life table results in a supplement.

62 In a novel approach that differs from early work of [Bergeron-Boucher et al. \(2018\)](#), [Kokoszka](#)  
63 [et al. \(2019\)](#), [Shang and Haberman \(2020\)](#) and [Zhang et al. \(2022\)](#), we jointly model and forecast  
64 the age distribution of life-table death counts for multiple populations to capture correlations  
65 among multiple series and improve forecast accuracy. Multiple-population modelling and  
66 forecasting have attracted increasing attention in actuarial science (see, e.g., [Jarner and Kryger](#)  
67 [2011](#), [Cairns et al. 2011](#), [Dowd et al. 2011](#), [Russolillo et al. 2011](#), [Hatzopoulos and Haberman](#)  
68 [2013](#), [Villegas et al. 2017](#)) and demography (see, e.g., [Li and Lee 2005](#), [Hyndman et al. 2013](#)). Our  
69 extension links multiple-population forecasting methods with the compositional data analysis  
70 (CoDa) framework (see also, [Shang and Kearney 2022](#), [Bergeron-Boucher et al. 2018](#)). In this  
71 paper, multiple populations refer to the female and male life-table death counts.

72 We also consider a nonparametric bootstrap method for constructing prediction intervals of  
73 life-table death counts. The procedure exploits a univariate autoregressive representation of  
74 the time series of principal component scores appearing in the functional principal component  
75 analysis of the functional process. Via the CoDa transformation, the bootstrap method generates  
76 functional replicates that mimic the temporal dependence of the (original) unconstrained  
77 functional time series.

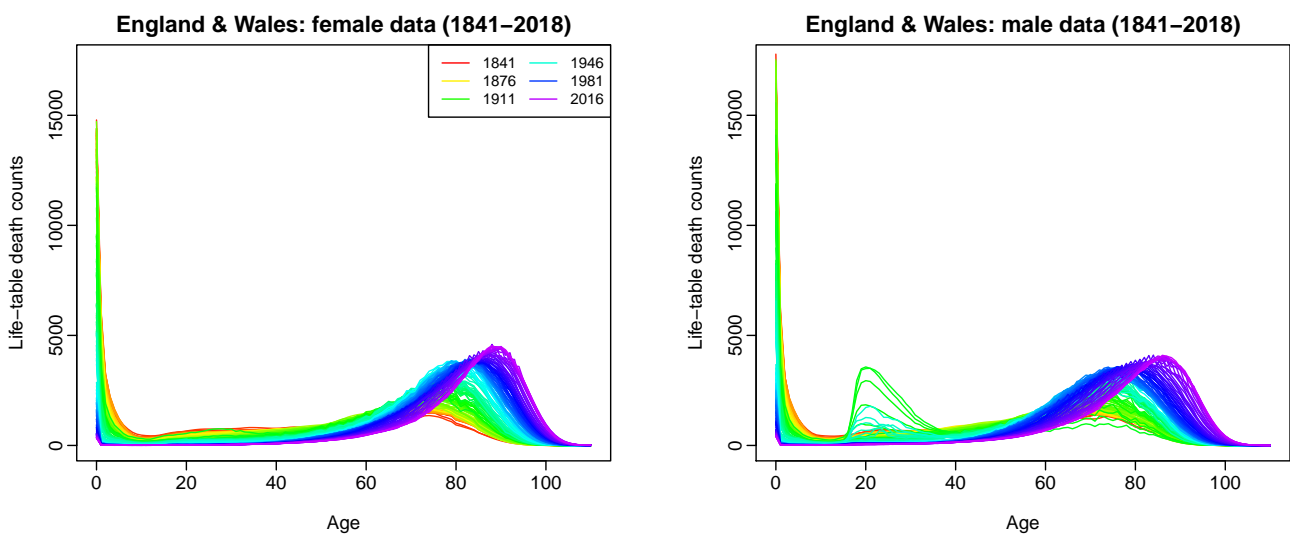
78 To demonstrate our proposed CoDa methods, we conduct a comprehensive analysis of the  
79 age- and sex-specific period life-table death counts from 1841 to 2018 in England & Wales. We  
80 evaluate and compare the one- to 30-step-ahead forecast accuracy between the CoDa methods  
81 for independently and jointly modelling multiple populations. To evaluate point forecast  
82 accuracy, we use the mean absolute percentage error (MAPE), Kullback-Leibler divergence, and  
83 two variants of the Jensen-Shannon divergence. The latter three measures are commonly used  
84 to evaluate density estimation accuracy. To assess the interval forecast accuracy, we consider the  
85 coverage probability difference between the empirical and nominal coverage probabilities and  
86 mean interval score of [Gneiting and Raftery \(2007\)](#) and [Gneiting and Katzfuss \(2014\)](#); refer to  
87 Section 6.1 for details. To provide reliable recommendations, we additionally evaluate the point  
88 and interval forecast accuracy on period life-table death counts in Sweden in Section 7. We also  
89 model cohort life-table death counts in both England & Wales and Sweden in the supplementary  
90 material.

91 The remainder of this paper is organised as follows: Section 2 describes the age- and sex-

92 specific period life-table death counts from 1841 to 2018 in England & Wales. Sections 3 and 4  
 93 couple the CoDa with multivariate and multilevel functional time series forecasting methods of  
 94 [Shang and Kearney \(2022\)](#) for producing the point and interval forecasts of the age distribution  
 95 of life-table death counts. Section 5 studies the goodness-of-fit of a CoDa extension. Using the  
 96 point and interval forecast error criteria in Section 6, we evaluate and compare the one- to 30-  
 97 step-ahead point and interval forecast accuracy among the methods considered. Section 7 also  
 98 presents the forecasting results on the period life-table death counts in Sweden. Conclusions  
 99 are summarised in Section 8, along with some reflections on how the methods presented here  
 100 can be further extended.

## 101 2 Age-distribution of death counts

102 We consider age- and sex-specific life-table death counts from 1841 to 2018 in England & Wales,  
 103 obtained from the [Human Mortality Database \(2022\)](#). We study life-table death counts, where  
 104 the life table radix (i.e., a population experiencing 100,000 births annually) is fixed at 100,000  
 105 at age 0 for each year. For the life-table death counts, there are 111 ages, and these are age  
 106  $0, 1, \dots, 109, 110+$ . Due to rounding, there are zero counts for age 110+ at some years. To rectify  
 107 this problem, we prefer to use the probability of dying and the life table radix to recalculate  
 108 our estimated death counts (up to 6 decimal places). In doing so, we obtain more precise death  
 109 counts than the ones reported in the [Human Mortality Database \(2022\)](#). To some extent, the  
 110 probability of dying relies on smooth rates (see the [Human Mortality Database 2022](#), protocol  
 111 for detail).

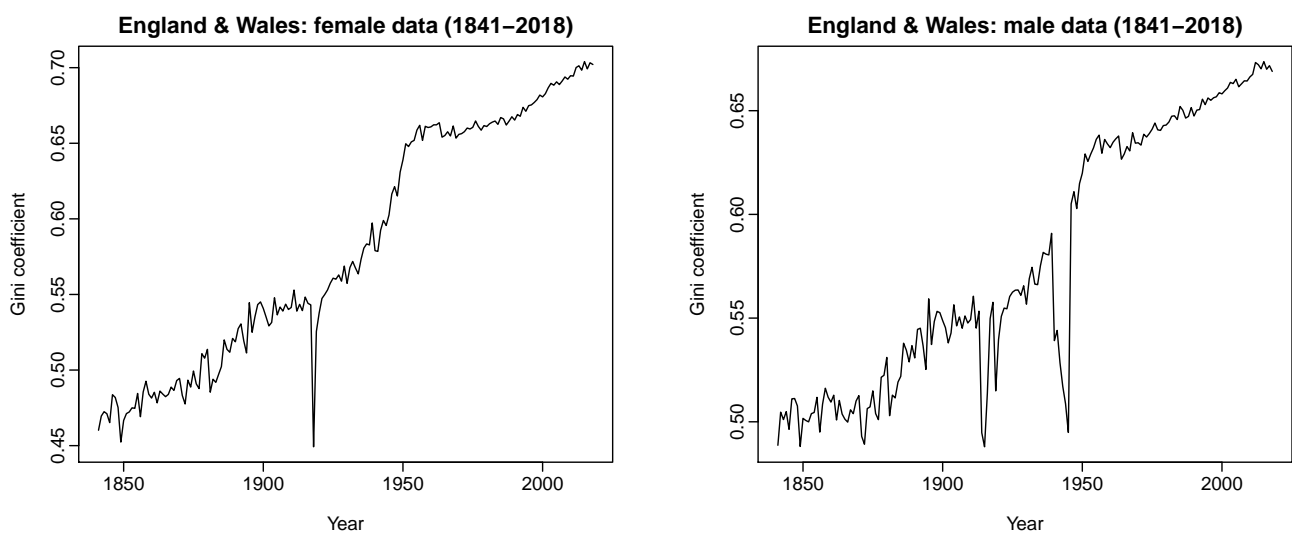


**Figure 1:** Rainbow plots of age-specific period life-table death count from 1841 to 2018 in a single-year group in England & Wales. The oldest years are shown in red, with the most recent years in violet. Curves are ordered chronologically according to the colours of the rainbow.

112 To understand the principal features of the data, Figure 1 presents rainbow plots of the  
113 female and male age-specific period life-table death counts in England & Wales from 1841 to  
114 2018 in a single-year group.

115 In the male population, there are a few bumps in the death counts between age 20 and 40  
116 for cohorts between 1900 and 1920 due to the First and Second World Wars. Apart from that,  
117 both sub-figures demonstrate a decreasing trend in infant death counts and a typical negatively  
118 skewed distribution for the life-table death counts, where the peaks shift to higher ages for both  
119 females and males. This shift is a primary driver of the longevity risk, which is a major issue for  
120 insurers and pension funds, especially in the selling and risk management of annuity products  
121 (see [Denuit et al. 2007](#), for a discussion).

122 The spread of the distribution indicates lifespan variability. A decrease in variability over  
123 time can be observed directly. It can be measured, for example, with the interquartile range  
124 of life-table ages at death or the Gini coefficient (for comprehensive reviews, see [Wilmoth  
125 and Horiuchi 1999](#), [Shkolnikov et al. 2003](#), [van Raalte and Caswell 2013](#), [Debón et al. 2017](#)).  
126 Figure 2 presents an example where the age-at-death distribution provides important insights  
127 on longevity and lifespan variability that cannot be grasped directly from an examination of  
128 either the central mortality rate or the survival function.



**Figure 2:** *Gini coefficients for period female and male life-table death count from 1841 to 2018.*

129 From Figure 2, the effects of the first and second World Wars are apparent for the male data  
130 based on Gini's coefficient. For the female data, there is a sudden drop around 1918 which  
131 possibly relates to the Spanish flu.

### 3 Constrained functional time-series forecasting methods

Density functions are non-negative functions that integrate into one. They naturally share some features with compositional data (see, e.g., [Aitchison 1986](#), [Pawlowsky-Glahn et al. 2015](#)). Compositional data arise in many scientific fields, such as geology (geochemical elements), economics (income/expenditure distribution), medicine (body composition), the food industry (food composition), chemistry (chemical composition), agriculture (nutrient balance bionomics), environmental science (soil contamination), ecology (abundance of different species) and demography (life-table death counts). In statistics, [Scealy et al. \(2017\)](#) use CoDa to study the concentration of chemical elements in sediment or rock samples. [Scealy and Welsh \(2017\)](#) apply CoDa to analyse total weekly expenditure on food and housing costs for households in a chosen set of domains. [Delicado \(2011\)](#), [Kokoszka et al. \(2019\)](#) and [Shang and Haberman \(2020\)](#) use CoDa to analyse density functions and implement dimension-reduction techniques on the constrained compositional data space. In demography, [Oeppen \(2008\)](#) and [Bergeron-Boucher et al. \(2017\)](#) put forward a principal component approach to forecast life-table death counts within a CoDa framework by considering age-specific life-table death counts as compositional data.

For a given period  $t$ , compositional data are defined as a random vector of  $I$  non-negative components,  $[\mathcal{X}_t(u_1), \dots, \mathcal{X}_t(u_I)]$ , whose sum is a specified constant, set typically equal to 1 (portion), 100 (percentage) or  $10^6$  for parts per million (ppm) in geochemical trace element compositions ([Aitchison 1986](#), p.1). Between the positivity and summability constraints, the sample space of compositional data is thereby a simplex

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

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

In the CoDa framework, the standard approach involves breaking the summability constraint using a transformation of the raw data to remove the constraint before applying conventional statistical techniques to the transformed data in an unconstrained space. Among many possible transformations, the centred log-ratio transformation is commonly used ([Aitchison and Shen 1980](#), [Aitchison 1982](#), [1986](#)). The algorithm for implementing the CoDa method consists of the following steps:

1) Compute the geometric mean function

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

where  $\ln(\cdot)$  denotes natural logarithm, and  $j$  denotes a population index in our multiple population setting; here, we consider  $J = 2$  representing the female and male data. We treat age as a continuum  $u \in [0, 110]$  although age is observed at discrete points, and set

$$s_t^j(u) = \frac{\mathcal{X}_t^j(u) / \alpha_n^j(u)}{\int_{v=0}^{110} \mathcal{X}_t^j(v) / \alpha_n^j(v) dv}.$$

The geometric mean standardises the ranges so that no range dominates the weighting.

2) Apply the centred log-ratio transformation given by

$$\beta_t^j(u) = \ln \left( \frac{s_t^j(u)}{g_t^j} \right), \quad (2)$$

where  $g_t^j$  is the geometric mean given by

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

The log-ratio transformation in (2) removes the constraints on  $\mathcal{X}_t^j(u)$ . For a given population  $j$ ,  $\beta_t^j(u)$  can be viewed as an unconstrained functional time series, and we present three methods below for modelling and forecasting multi-population functional time series.

3a) Univariate functional principal component (FPC) decomposition (see, e.g., [Kokoszka et al.](#)

[2019](#), [Shang and Haberman 2020](#)): Apply functional principal component analysis (FPCA)

to the transformed data  $\beta^j(u) = \{\beta_1^j(u), \dots, \beta_n^j(u)\}$ , i.e., compute the Karhunen-Loève

expansion of a functional realisation

$$\beta_t^j(u) = \sum_{\ell=1}^n \hat{\gamma}_{t,\ell}^j \hat{\phi}_\ell^j(u) = \sum_{\ell=1}^{L_j} \hat{\gamma}_{t,\ell}^j \hat{\phi}_\ell^j(u) + \hat{\omega}_t^j(u), \quad (3)$$

where  $\hat{\omega}_t^j(u)$  denotes model residual function for the  $j^{\text{th}}$  population at age  $u$  and year  $t$ ,

$[\hat{\phi}_1^j(u), \dots, \hat{\phi}_{L_j}^j(u)]$  are the first  $L_j$  estimated FPCs, and  $(\hat{\gamma}_{t,1}^j, \dots, \hat{\gamma}_{t,L_j}^j)$  are their scores at

year  $t$ .

We determine  $L_j$  by three eigenvalue ratio criteria (see [Ahn and Horenstein 2013](#), [Li et al.](#)

2020). The first one is known as the eigenvalue ratio (ER) estimator, which is simply obtained by maximising the ratio of two adjacent eigenvalues of  $\beta^j(\beta^j)^\top / (nI)$  arranged in descending order, where  $\beta^j = [\beta^j(u_1), \dots, \beta^j(u_I)]$  and  $I$  denotes the total number of discrete grid points:

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

The second criterion function, known as the growth ratio (GR), is given by

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

where

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

171 The optimal number  $L_j$  is selected as

$$L_j = \max \left\{ \underset{1 \leq \ell \leq Lmax}{\operatorname{argmax}} \text{ER}(\ell), \underset{1 \leq \ell \leq Lmax}{\operatorname{argmax}} \text{GR}(\ell) \right\}. \quad (4)$$

172 In general, the  $Lmax$  is a pre-specified positive integer, which can be set as  $Lmax =$   
 173  $\#\{\ell | \tilde{\mu}_{nI,\ell} \geq \sum_{\ell=1}^n \tilde{\mu}_{nI,\ell} / n, \ell \geq 1\}$ . In other words,  $Lmax$  counts the number of eigenvalues,  
 174 which is greater than the sum of all eigenvalues divides by sample size.

175 The third criterion estimates the number of components as the integer minimising ratios  
 176 of two adjacent eigenvalues given by

$$L_j = \underset{1 \leq \ell \leq Lmax}{\operatorname{argmin}} \left\{ \frac{\tilde{\mu}_{\ell+1}}{\tilde{\mu}_\ell} \times \mathbb{1} \left( \frac{\tilde{\mu}_\ell}{\tilde{\mu}_1} \geq \theta \right) + \mathbb{1} \left( \frac{\tilde{\mu}_\ell}{\tilde{\mu}_1} < \theta \right) \right\}, \quad (5)$$

177 where  $\theta$  is a pre-specified small positive number, which can be set as  $\theta = 1 / \ln[\max(\tilde{\mu}_1, n)]$   
 178 and  $\mathbb{1}(\cdot)$  is the binary indicator function.

179 By taking a conservative view, the actual number of retained FPCs is more likely to be  
 180 within the maximum of the estimated numbers of retained components in (4) and (5).

3b) Multivariate FPC decomposition: We stack the unconstrained data  $[\beta_t^1(u), \beta_t^2(u)]$  into a long vector of functional time series  $\beta_t(u) = [\beta_t^1(u), \beta_t^2(u)]$ . Computationally,  $\beta^j(u)$  is a matrix of  $n \times I$ , then  $\beta(u)$  is a matrix of  $n \times (2I)$ . Then, we apply FPCA to the transformed data matrix  $\beta(u) = [\beta_1(u), \dots, \beta_n(u)]$ , i.e., compute the Karhunen-Loève expansion of a

functional realisation

$$\beta_t^j(u) = \sum_{\ell=1}^n \hat{\gamma}_{t,\ell}^j \hat{\phi}_\ell^j(u) \approx \sum_{\ell=1}^L \hat{\gamma}_{t,\ell}^j \hat{\phi}_\ell^j(u),$$

where  $[\hat{\phi}_1^j(u), \dots, \hat{\phi}_L^j(u)]$  are the first  $L$  estimated FPCs extracted from the variance-covariance of the *stacked* functional time series, and  $(\hat{\gamma}_{\ell,1}^j, \dots, \hat{\gamma}_{\ell,L}^j)$  are their corresponding scores for the  $j^{\text{th}}$  population. For the first population, we truncate the first  $I$  elements; for the second population, we truncate the second  $I$  elements. The optimal number of  $L$  is determined by the eigenvalue ratio criteria in (4).

3c) Multilevel FPC decomposition: We can model  $\beta_t^j(u)$  via a multilevel FPCA. We seek to find a common pattern shared among populations and a series-specific pattern for each  $j$  series. Mathematically, the multilevel FPC regression can be expressed as

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

where  $U_t(u)$  denotes a common trend, such as the average of all series;  $R_t^j(u)$  denotes a series-specific trend, and  $e_t^j(u)$  denotes an error term. The common trend  $U_t(u)$  can be a simple average of  $\beta_t^{(1)}(u)$  and  $\beta_t^{(2)}(u)$  (see, e.g., Li and Lee 2005, Shang 2016). The common trend and series-specific trend can be modelled via a two-stage FPC analyses,

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

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

where  $K$  and  $L$  denote the number of retained components,  $[\hat{\phi}_1(u), \dots, \hat{\phi}_K(u)]$  denotes the estimated FPCs for the common trend, while  $[\hat{\psi}_1^j(u), \dots, \hat{\psi}_L^j(u)]$  denotes the estimated FPCs for the population-specific trend,  $(\hat{\gamma}_{t,1}, \dots, \hat{\gamma}_{t,K})$  denotes the estimated principal component score for period  $t$ , while  $(\hat{\gamma}_{t,1}^j, \dots, \hat{\gamma}_{t,L}^j)$  denotes the estimated principal component score for period  $t$ . The optimal numbers of  $K$  and  $L$  are determined by the eigenvalue ratio criteria in (4).

An important parameter is the proportion of variability explained by the aggregate data, which is the variance explained by the within-cluster variability (Di et al. 2009). A possible measure of within-cluster variability is given by

$$\frac{\sum_{k=1}^{\infty} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k + \sum_{l=1}^{\infty} \lambda_l^j} = \frac{\int_{u=0}^{110} \text{Var}[\mathbf{R}(u)] du}{\int_{u=0}^{110} \text{Var}[\mathbf{R}(u)] du + \int_{u=0}^{110} \text{Var}[\mathbf{U}^j(u)] du}, \quad (6)$$

where  $\lambda_k$  represents the  $k^{\text{th}}$  eigenvalue of the common trend, and  $\lambda_l^j$  represents the  $l^{\text{th}}$  eigenvalue of the population-specific trend. When the common factor can explain the primary mode of total variability, the value of within-cluster variability is close to 1.

- 4) Forecast the FPC scores. Using a univariate time series forecasting method, we obtain the  $h$ -step-ahead forecast  $\hat{\gamma}_{n+h|n,\ell}^j$  of the  $\ell^{\text{th}}$  principal component score (see, e.g., Hyndman and Shang 2009, Aue et al. 2015). Conditioning on the estimated principal components and observed data, the forecast of  $\beta_{n+h}^j(u)$  is given by

$$\hat{\beta}_{n+h|n}^j(u) = \sum_{\ell=1}^{L_j} \hat{\gamma}_{n+h|n,\ell} \hat{\phi}_{\ell}^j(u). \quad (7)$$

Similarly, the principal component scores in the multivariate and multilevel FPC decompositions can be obtained by a univariate or multivariate time series forecasting method. We consider four commonly used univariate time series forecasting methods (see Section 6). These methods can model the possible presence of nonstationarity in each set of principal component scores.

- 5) Transform back to the compositional data, i.e., take the inverse centred log-ratio transformation given by

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

where  $\hat{\beta}_{n+h|n}^j(u)$  denotes the forecasts in (7).

- 6) Finally, we add back the geometric means to obtain the forecasts of the density function

$$\hat{\chi}_{n+h|n}^j(u) = \frac{\hat{s}_{n+h|n}^j(u) \alpha_n^j(u)}{\int_{u=0}^{110} \hat{s}_{n+h|n}^j(u) \alpha_n^j(u) du},$$

where  $\alpha_n^j(u)$  is the geometric mean function given in (1).

## 4 Nonparametric bootstrap

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. The forecast uncertainty stems from systematic deviations (e.g., parameter and model uncertainty) and random fluctuations (e.g., due to the model error term). As pointed out by Chatfield (2000, Chapter 7), it is essential in both demographic and actuarial applications to provide interval forecasts as well as point forecasts to (1) assess future uncertainty levels;

215 (2) enable different strategies to be planned for a range of possible outcomes indicated by  
 216 the interval forecasts; (3) compare forecasts from different methods more thoroughly; and  
 217 (4) explore different scenarios based on various assumptions.

Our aim is to construct a prediction band for the unconstrained functional time series  $\beta_{n+h}$  associated with the predictor  $\widehat{\beta}_{n+h}$ . The prediction band, denoted by  $[\widehat{\beta}_{n+h}(u) - L_{n,h}(u), \widehat{\beta}_{n+h}(u) + U_{n,h}(u)]$ , is given as

$$\lim_{n \rightarrow \infty} \Pr \left( \widehat{\beta}_{n+h}(u) - L_{n,h}(u) \leq \beta_{n+h}(u) \leq \widehat{\beta}_{n+h}(u) + U_{n,h}(u), \quad \forall u \in \mathcal{I} | \beta_{n,k} \right) = 1 - \alpha,$$

where  $\beta_{n,k} = (\beta_n, \beta_{n-1}, \dots, \beta_{n-k+1})$  are the last  $k$  observations. The crux of the problem is on the estimation of the conditional distribution of the prediction error  $\mathcal{E}_{n+h} = \beta_{n+h} - \widehat{\beta}_{n+h}$  given  $\beta_{n,k}$ . The prediction error can be decomposed as

$$\begin{aligned} \mathcal{E}_{n+h} &:= \beta_{n+h} - \widehat{\beta}_{n+h} \\ &= \epsilon_{n+h} + [f(\beta_{n+h-1}, \beta_{n+h-2}, \dots) - g(\beta_{n+h-1}, \beta_{n+h-2}, \dots, \beta_{n+h-k})] + \\ &\quad + \left[ g(\beta_{n+h-1}, \beta_{n+h-2}, \dots, \beta_{n+h-k}) - \widehat{g}(\widehat{\beta}_{n+h-1}, \widehat{\beta}_{n+h-2}, \dots, \widehat{\beta}_{n+h-k}) \right] \\ &= \mathcal{E}_{I,n+h} + \mathcal{E}_{M,n+h} + \mathcal{E}_{E,n+h}. \end{aligned}$$

218  $\mathcal{E}_{I,n+h}$  is the error attributable to the independent and identically distributed innovation,  $\mathcal{E}_{M,n+h}$   
 219 is the model mis-specification error, and  $\mathcal{E}_{E,n+h}$  is the error attributable to estimation of the  
 220 unknown function  $g$  and of random element  $(\beta_{n+h-1}, \dots, \beta_{n+h-k})$  used for the  $h$ -step-ahead  
 221 prediction.

222 The construction of the prediction interval for CoDa has previously been considered by  
 223 Bergeron-Boucher et al. (2017) and Shang and Haberman (2020). However, the existing methods  
 224 do not take into account a model mis-specification error. Our bootstrap method does not  
 225 only take into account the three sources of uncertainty, but it also considers multi-population  
 226 modelling of unconstrained data  $\beta^j(u) = \{\beta_1^j(u), \dots, \beta_n^j(u)\}$ .

Univariate functional time series method (see Kokoszka et al. 2019, Shang and Haberman 2020):

Using a univariate time series forecasting model, we can obtain multi-step-ahead forecasts for the principal component scores,  $\{\widehat{\gamma}_{1,\ell}^j, \dots, \widehat{\gamma}_{n,\ell}^j\}$  for  $\ell = 1, \dots, L$ . Let the  $h$ -step-ahead forecast errors be  $\zeta_{t,h,\ell}^j = \widehat{\gamma}_{t,\ell}^j - \widehat{\gamma}_{t|t-h,\ell}^j$ , for  $t = h+1, \dots, n$ . These can then be sampled with replacement to give a bootstrap sample of  $\gamma_{n+h,\ell}$ :

$$\widehat{\gamma}_{n+h|n,\ell}^{j,b} = \widehat{\beta}_{n+h|n,\ell} + \widehat{\zeta}_{*,h,\ell}^{j,b}, \quad b = 1, \dots, B,$$

227 where  $B = 1,000$  symbolises the total number of bootstrap replications and  $\widehat{\zeta}_{*,h,\ell}^{j,b}$  are sampled  
 228 with replacement from  $\widehat{\zeta}_{t,h,\ell}^j$ .

229 While the first  $L_j$  principal components approximate the data  $\beta^j(u)$  relatively well, the  
 230 model residuals should contribute nothing but random noise. We apply a white noise test of  
 231 [Bagchi et al. \(2018\)](#) to examine whether or not the residuals are white noise. Consequently, we  
 232 can bootstrap the model fit errors in (3) by sampling with replacement from the model residual  
 233 term  $\{\widehat{\omega}_1^j(u), \dots, \widehat{\omega}_n^j(u)\}$  for a given age  $u$ .

Adding both components of variability, we obtain  $B$  variants for  $\beta_{n+h}^j(u)$ :

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

234 With the bootstrapped  $\{\widehat{\beta}_{n+h|n}^{j,1}(u), \dots, \widehat{\beta}_{n+h|n}^{j,B}(u)\}$ , we fit a functional time series model, where  
 235 the retained number of principal components is estimated from the ER and GR tests and is  
 236 allowed to be different values among the bootstrap samples. By conditioning on the estimated  
 237 mean function and FPCs in each bootstrap sample, we then obtain the forecasts.

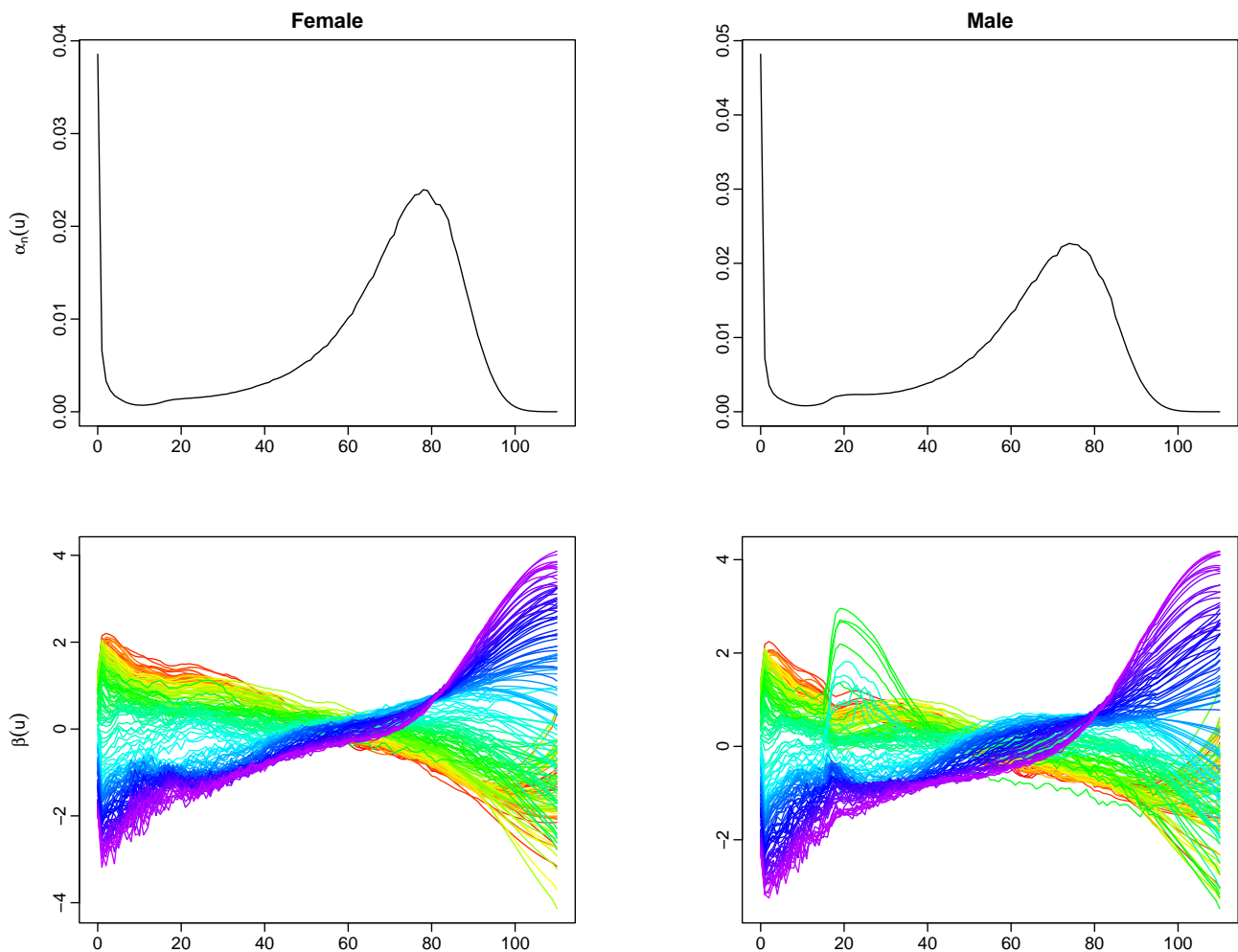
238 Multivariate functional time series method (see [Shang and Kearney 2022](#)): By stacking  $\beta_t(u) =$   
 239  $[\beta_t^1(u), \beta_t^2(u)]$  into a long vector of functional time series, we then apply a FPCA to decompose  
 240  $[\beta_1(u), \dots, \beta_n(u)]$  into  $L$  sets of principal components and their associated scores. We apply  
 241 our bootstrap procedure to construct the bootstrap samples, namely  $\widehat{\beta}_{n+h|n}^{j,b}(u)$ .

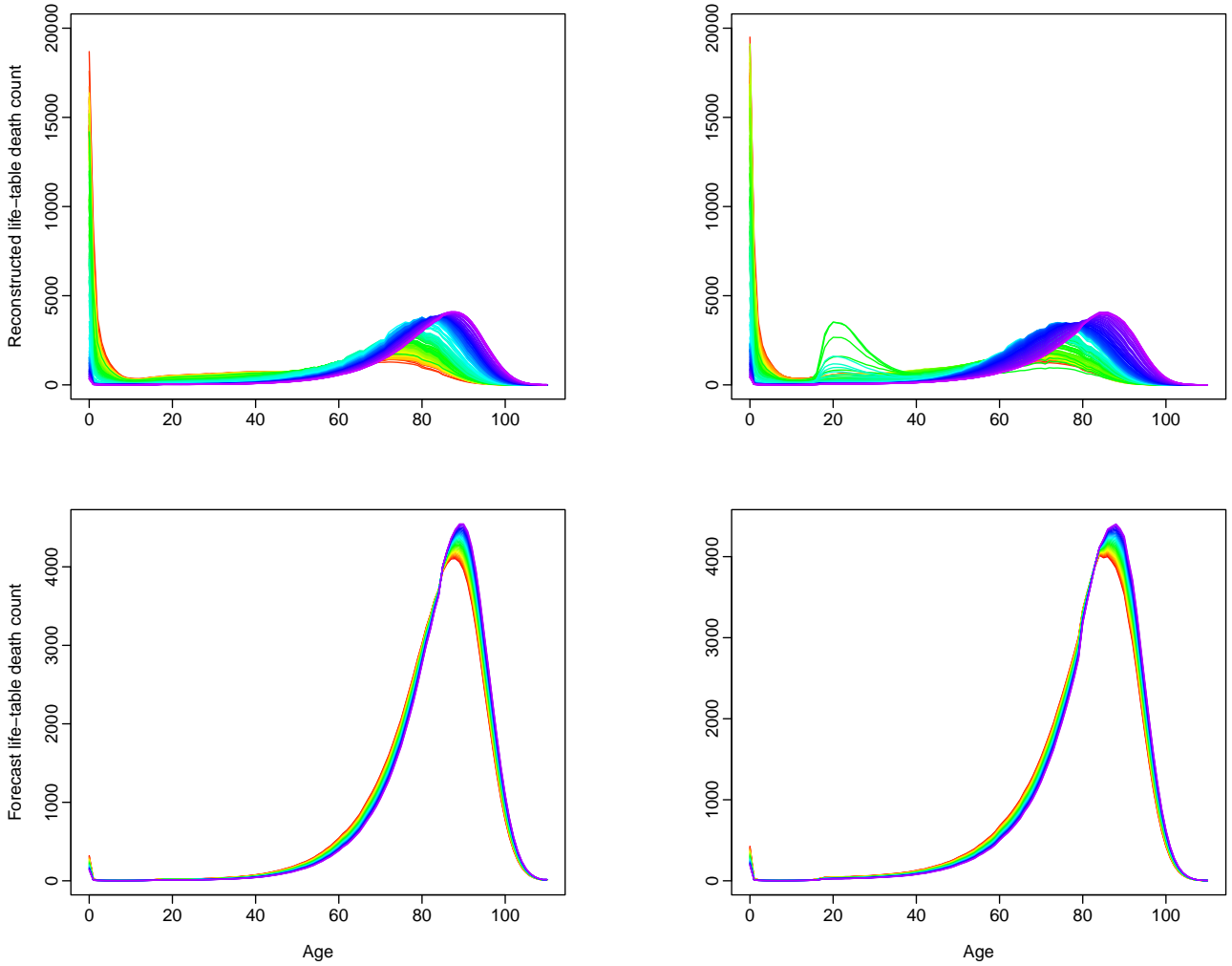
242 Multilevel functional time series method (see [Shang et al. 2016](#), [Shang and Kearney 2022](#)): By  
 243 taking a simple average  $U_t(u) = \bar{\beta}_t(u) = \frac{\beta_t^1(u) + \beta_t^2(u)}{2}$ , we obtain a common series between the  
 244 two populations. First, we apply a FPCA to decompose  $\bar{\beta}(u) = [\bar{\beta}_1(u), \dots, \bar{\beta}_n(u)]$  into  $K$  sets of  
 245 FPCs and their associated scores. From the principal component decomposition of the averaged  
 246 series, we can compute the population-specific residual trend. We again apply a FPCA to  
 247 decompose the population-specific residual trend into  $L$  sets of FPCs and their associated scores.  
 248 Then, we apply the above bootstrap procedure to construct the bootstrap samples, namely  
 249  $\widehat{U}_{n+h|n}^b(u)$  and  $\widehat{R}_{n+h|n}^{j,b}(u)$ . Then,  $\widehat{\beta}_{n+h|n}^{j,b}(u) = \widehat{U}_{n+h|n}^b(u) + \widehat{R}_{n+h|n}^{j,b}(u)$ .

250 With the bootstrapped  $\widehat{\beta}_{n+h|n}^{j,b}(u)$ , we follow steps 5) and 6) in Section 3, in order to obtain the  
 251 bootstrap forecast of  $\mathcal{X}_{n+h}^{j,b}(u)$ . At the  $100(1 - v)\%$  nominal coverage probability, the prediction  
 252 bands are obtained by taking  $v/2$  and  $1 - v/2$  quantiles based on  $\{\widehat{\mathcal{X}}_{n+h|n}^{j,1}(u), \dots, \widehat{\mathcal{X}}_{n+h|n}^{j,B}(u)\}$ .

## 253 5 CoDa model fitting

254 We examine the goodness-of-fit of the CoDa model for multiple populations. In Section 6, the  
255 CoDa with the multivariate and multilevel functional time series methods are our recommenda-  
256 tion for producing the point and interval forecasts for multiple populations. For example, we  
257 consider the CoDa model fitting by the multilevel functional time series method for the female  
258 and male life-table death counts. The number of retained components,  $K$  and  $L$  are determined  
259 by the ER and GR criteria in the common and residual trends, respectively. For the common  
260 trend between the female and male data in England & Wales, the chosen number of components  
261 is 1. For the population-specific trend between the female and male data in England & Wales,  
262 the chosen numbers of components for the residual trend is 1 and 4, respectively. From (6),  
263 we compute the within-cluster variability. The common trend accounts for 96.16% of the total  
264 variation for the female series, and it accounts for 93.63% of the total variation for the male  
265 series.





**Figure 3:** Elements of the CoDa with the multilevel functional time series method for analysing female and male age-specific life-table death counts in England & Wales from 1841 to 2018.

266 From the observed death counts from 1841 to 2018 (i.e., 178 observations), in Figure 3, we  
 267 present the geometric mean of female and male period life-table death counts, given by  $\alpha_x^j$ ,  
 268 the transformed data matrix  $\beta^j(u) = \{\beta_1^j(u), \dots, \beta_n^j(u)\}$ , the reconstructed female and male  
 269 life-table death counts using a multilevel FPCA, and one- to 30-step-ahead forecasts of female  
 270 and male life-table death counts. From the unconstrained data  $\beta^1(u)$  and  $\beta^2(u)$ , we notice that  
 271 the series are nonstationary. Therefore, our bootstrap procedure for constructing prediction  
 272 intervals must be tailored to handle nonstationary functional time series. These forecast life-  
 273 table death counts are obtained by multiplying the estimated principal components with the  
 274 forecast principal component scores using the random walk with drift (RWD) for the years  
 275 between 2019 and 2048. RWD is one of the four univariate forecasting methods employed in  
 276 the example, and which will be discussed in Section 6.4.

## 277 6 Comparisons of point and interval forecast accuracy

### 278 6.1 Expanding window

279 An expanding window analysis of a time series model is commonly used to assess model  
280 and parameter stability over time and prediction accuracy. The expanding window analysis  
281 determines the constancy of a model's parameter by computing parameter estimates and their  
282 resultant forecasts over an expanding window of a fixed size through the sample (for details  
283 [Zivot and Wang 2006](#), pp. 313–314). Using the first 148 observations from 1841 to 1988 in  
284 the female and male age-specific life-table death counts, we produce one- to 30-step-ahead  
285 forecasts. Through an expanding window approach, we re-estimate the parameters in the time  
286 series forecasting models using the first 149 observations from 1841 to 1989. Forecasts from  
287 the estimated models are then produced for one- to 29-step-ahead forecasts. We iterate this  
288 process by increasing the sample size by one year until we reached the data period in 2018.  
289 This process produces 30 one-step-ahead forecasts, 29 two-step-ahead forecasts, . . . , and one  
290 30-step-ahead forecast. We compare these forecasts with the holdout samples between 1989 and  
291 2018 to determine the out-of-sample forecast accuracy.

### 292 6.2 Point forecast error criteria

To evaluate the point forecast accuracy, we consider the MAPE that measures how close the forecasts are to the holdout observations being forecast, regardless of the direction of forecast errors. This error measure can be written as

$$\text{MAPE}(h) = \frac{1}{111 \times (31 - h)} \sum_{\xi=h}^{30} \sum_{i=1}^{111} \left| \frac{\mathcal{X}_{n+\xi}(u_i) - \widehat{\mathcal{X}}_{n+\xi|n}(u_i)}{\mathcal{X}_{n+\xi}(u_i)} \right| \times 100, \quad i = 1, \dots, 111,$$

293 where  $\mathcal{X}_{n+\xi}(u_i)$  denotes the holdout sample for the  $u_i^{\text{th}}$  age and  $\xi^{\text{th}}$  forecasting year, while  
294  $\widehat{\mathcal{X}}_{n+\xi|n}(u_i)$  denotes the point forecasts for the holdout sample.

Since age-specific life-table death counts can be considered a probability density function, we also consider some density evaluation measures. These measures include the discrete version of the Kullback-Leibler divergence ([Kullback and Leibler 1951](#)) and the square root of the Jensen-Shannon divergence ([Shannon 1948](#)). The Kullback-Leibler divergence is intended to measure the loss of information when we choose an approximation. For two probability density functions denoted by  $\mathcal{X}_{n+\xi}(u)$  and  $\widehat{\mathcal{X}}_{n+\xi|n}(u)$ , the discrete version of the Kullback-Leibler divergence is

defined as

$$\begin{aligned} \text{KLD}(h) &= D_{\text{KL}}[\mathcal{X}_{n+\zeta}(u_i) \|\hat{\mathcal{X}}_{n+\zeta|n}(u_i)] + D_{\text{KL}}[\hat{\mathcal{X}}_{n+\zeta|n}(u_i) \|\mathcal{X}_{n+\zeta}(u_i)] \\ &= \frac{1}{111 \times (31-h)} \sum_{\zeta=h}^{30} \sum_{i=1}^{111} \mathcal{X}_{n+\zeta}(u_i) \cdot [\ln \mathcal{X}_{n+\zeta}(u_i) - \ln \hat{\mathcal{X}}_{n+\zeta|n}(u_i)] + \\ &\quad \frac{1}{111 \times (31-h)} \sum_{\zeta=h}^{30} \sum_{i=1}^{111} \hat{\mathcal{X}}_{n+\zeta|n}(u_i) \cdot [\ln \hat{\mathcal{X}}_{n+\zeta|n}(u_i) - \ln \mathcal{X}_{n+\zeta}(u_i)], \end{aligned}$$

295 which is symmetric and non-negative.

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

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

296 where  $\delta_{n+\zeta}(u_i)$  measures a common quantity between  $\mathcal{X}_{n+\zeta}(u_i)$  and  $\hat{\mathcal{X}}_{n+\zeta|n}(u_i)$ . We consider  
 297 simple mean and geometric mean, given by  $\delta_{n+\zeta}(u_i) = \frac{1}{2}[\mathcal{X}_{n+\zeta}(u_i) + \hat{\mathcal{X}}_{n+\zeta|n}(u_i)]$  or  $\delta_{n+\zeta}(u_i) =$   
 298  $\sqrt{\mathcal{X}_{n+\zeta}(u_i) \hat{\mathcal{X}}_{n+\zeta|n}(u_i)}$ . We denote  $\text{JSD}^s(h)$  for the Jensen-Shannon divergence with the simple  
 299 mean, and  $\text{JSD}^g(h)$  for the Jensen-Shannon divergence with the geometric mean. To make the  
 300 Jensen-Shannon divergence a metric between two probability densities, we take the square root  
 301 of the Jensen-Shannon divergence (see, e.g., [Fuglede and Topsøe 2004](#)).

### 302 6.3 Interval forecast error criteria

To evaluate the interval forecast accuracy, we consider the coverage probability difference (CPD) between the empirical and nominal coverage probabilities and mean interval score of [Gneiting and Raftery \(2007\)](#). For each year in the forecasting period, the  $h$ -step-ahead prediction intervals are calculated at the  $100(1-v)\%$  nominal coverage probability. We consider the common case of the symmetric  $100(1-v)\%$  prediction intervals, with lower and upper bounds that are predictive quantiles at  $v/2$  and  $1-v/2$ , denoted by  $\hat{\mathcal{X}}_{n+\zeta}^{\text{lb}}(u_i)$  and  $\hat{\mathcal{X}}_{n+\zeta}^{\text{ub}}(u_i)$ . The CPD is defined as

$$\text{CPD}_h = \frac{1}{111 \times (31-h)} \sum_{\zeta=h}^{30} \sum_{i=1}^{111} \left[ \mathbb{1}\{\mathcal{X}_{n+\zeta}(u_i) > \hat{\mathcal{X}}_{n+\zeta}^{\text{ub}}(u_i)\} + \mathbb{1}\{\mathcal{X}_{n+\zeta}(u_i) < \hat{\mathcal{X}}_{n+\zeta}^{\text{lb}}(u_i)\} \right].$$

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

$$\overline{\text{CPD}} = \frac{1}{30} \sum_{h=1}^{30} \text{CPD}_h.$$

As defined by [Gneiting and Raftery \(2007\)](#), a scoring rule for the prediction intervals at time

point  $\mathcal{X}_{n+\xi}(u_i)$  is

$$S_{v,\xi} \left[ \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i), \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i), \mathcal{X}_{n+\xi}(u_i) \right] = \left[ \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i) - \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i) \right] \\ + \frac{2}{v} \left[ \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i) - \mathcal{X}_{n+\xi}(u_i) \right] \mathbb{1} \left\{ \mathcal{X}_{n+\xi}(u_i) < \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i) \right\} \\ + \frac{2}{v} \left[ \mathcal{X}_{n+\xi}(u_i) - \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i) \right] \mathbb{1} \left\{ \mathcal{X}_{n+\xi}(u_i) > \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i) \right\},$$

303 where  $\mathbb{1}\{\cdot\}$  represents the binary indicator function and  $v$  denotes the level of significance,  
 304 customarily  $v = 0.2$  or  $0.05$ . The interval score rewards a narrow prediction interval, if and only  
 305 if the true observation lies within the prediction interval. The optimal interval score is achieved  
 306 when  $\mathcal{X}_{n+\xi}(u_i)$  lies between  $\hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i)$  and  $\hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i)$ , and the distance between  $\hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i)$  and  
 307  $\hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i)$  is minimal.

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

$$\bar{S}_v(h) = \frac{1}{111 \times (31 - h)} \sum_{\xi=h}^{30} \sum_{i=1}^{111} S_{v,\xi} \left[ \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i), \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i), \mathcal{X}_{n+\xi}(u_i) \right],$$

where  $S_{v,\xi} \left[ \hat{\mathcal{X}}_{n+\xi}^{\text{lb}}(u_i), \hat{\mathcal{X}}_{n+\xi}^{\text{ub}}(u_i), \mathcal{X}_{n+\xi}(u_i) \right]$  denotes the interval score at the  $i^{\text{th}}$  age and  $\xi^{\text{th}}$  curve  
 in the forecasting period. Averaged over all forecast horizons, we obtain the overall mean  
 interval score

$$\bar{S}_v = \frac{1}{30} \sum_{h=1}^{30} \bar{S}_v(h).$$

## 308 6.4 Forecast results

309 Using the expanding window scheme, we compare the one-step-ahead to 30-step-ahead forecast  
 310 errors between the CoDa with the functional time series methods in Table 1. We consider  
 311 forecasting each set of the estimated principal component scores by four distinct univariate  
 312 time series forecasting methods. These forecasting methods are the autoregressive integrated  
 313 moving average (ARIMA), exponential smoothing (ETS), naïve random walk (RW) and RWD.  
 314 While these methods are widely applied to analyse linear time series, ETS can be tailored to fit  
 315 a nonlinear time series. We aim to identify which one of the four methods produces the most  
 316 accurate forecasts through an empirical comparison.

**Table 1:** A comparison of the point forecast accuracy, as measured by the overall MAPE, KLD and JSD with simple and geometric means, among the CoDa with independent and joint modelling approaches and several benchmark methods using a holdout sample of the period life-table death counts for the female and male data in England & Wales. Further, we consider four univariate time series forecasting methods for our functional time-series forecasting methods within the CoDa. ETS denotes the automatic exponential smoothing method. ARIMA denotes the automatic autoregressive integrated moving average method of Hyndman and Khandakar (2008). FTS denotes the univariate functional time series method. MFTS denotes the multivariate functional time series method. MLFTS denotes the multilevel functional time series method. The smallest errors are highlighted in bold.

Modelling method	Forecasting method	Female				Male			
		MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>	MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>
<u>CoDa</u>									
FTS	ARIMA	29.900	0.041	0.010	0.010	42.737	0.130	0.031	0.033
	ETS	33.405	0.032	0.008	0.008	47.223	0.153	0.037	0.038
	RW	31.422	0.070	0.017	0.017	47.077	0.152	0.036	0.038
	RWD	29.914	0.042	0.010	0.010	43.175	0.127	0.031	0.032
MFTS	ARIMA	18.966	0.021	0.005	0.005	38.318	0.099	0.024	0.025
	ETS	32.214	0.044	0.011	0.011	52.609	0.123	0.030	0.031
	RW	33.953	0.041	0.010	0.010	54.951	0.118	0.029	0.029
	RWD	<b>18.031</b>	<b>0.018</b>	<b>0.004</b>	<b>0.004</b>	38.096	0.092	0.022	0.023
MLFTS	ARIMA	22.988	0.034	0.008	0.008	27.499	0.055	0.013	0.014
	ETS	32.837	0.054	0.013	0.013	39.211	0.079	0.019	0.020
	RW	34.762	0.054	0.013	0.013	39.797	0.078	0.019	0.019
	RWD	20.833	0.028	0.007	0.007	<b>27.049</b>	<b>0.053</b>	<b>0.013</b>	<b>0.013</b>
<u>RW</u>		34.209	0.042	0.010	0.010	43.151	0.086	0.021	0.021
<u>LC</u>	VECM	29.020	0.046	0.011	0.011	43.110	0.140	0.034	0.035
<u>LL</u>	RWD-AR	33.451	0.076	0.019	0.019	36.668	0.075	0.018	0.019
<u>PR</u>	ARIMA	35.114	0.062	0.015	0.016	40.699	0.108	0.026	0.027
<u>CBD-M6</u>	VAR	43.422	0.161	0.039	0.040	60.407	0.178	0.042	0.045
<u>CBD-M6</u>	VECM	31.029	0.088	0.022	0.022	41.486	0.107	0.026	0.027

317

318 From the aggregated error measures over the forecast horizon, we present the forecast errors  
319 obtained from the CoDa and a number of single- and multi-population benchmark models in  
320 Table 1, namely,

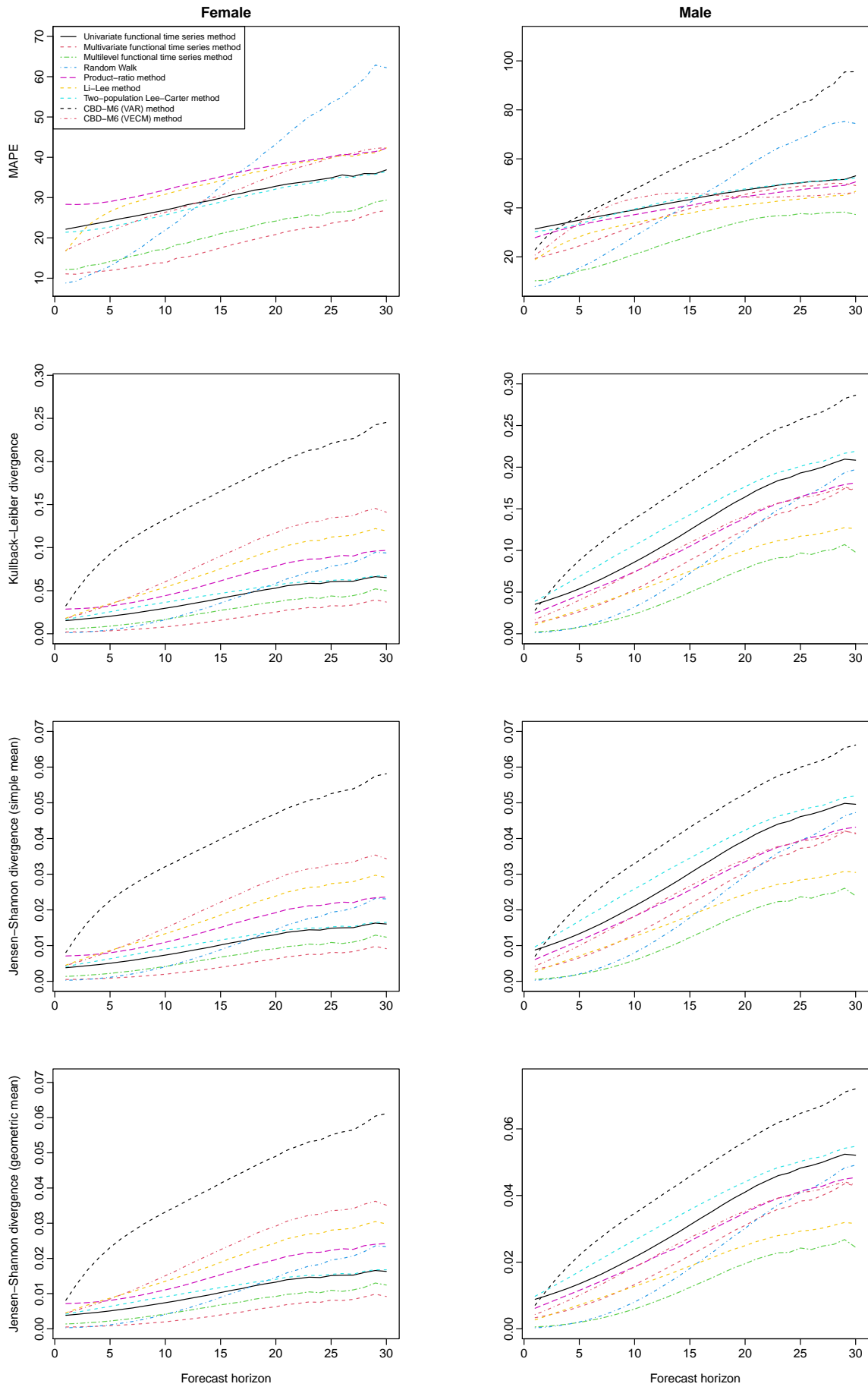
- 321 1) the naïve random walk model.
- 322 2) two-population Li-Carter model (LC) with a vector error correction model (VECM) for the  
323 evolution of period effects, discussed by Zhou et al. (2014).
- 324 3) Li-Lee model (LL) with RWD for common PC scores and autoregressive (AR) of order 1 for  
325 population-specific PC scores, considered by Li and Lee (2005).
- 326 4) product-ratio model (PR) with the FPC scores forecasted by ARIMA, which was introduced  
327 by Hyndman et al. (2013).
- 328 5) two-population Cairns-Blake-Dowd model with cohort effect (CBD-M6) with population-  
329 specific coefficients jointly predicted by vector autoregressive model (VAR) of order 1 or  
330 VECM (see, e.g., Li et al. 2015).

331 The benchmark models do not consider the constraints. Therefore, to obtain a fair comparison  
332 with our proposed methods, we slightly modify the benchmark approaches by applying them  
333 to the unconstrained  $\beta_t^j(u)$  after the centred log-ratio transformation to adapt the positivity and  
334 summability constraints in the age-specific life table.

335 We consider the RW method for each sub-population for the single-population benchmark  
336 separately, where the one-step-ahead forecast density is the same as the most recently observed  
337 density. While this method does not consider temporal changes, it does not require the CoDa  
338 transformation and thus is computationally fast. Regarding the coherent forecasting of mortality  
339 rates for multiple sub-populations, most such models extend the well-known single-population  
340 models, such as the Lee-Carter model and Cairns-Blake-Dowd (CBD) model, by specifying the  
341 correlation and interaction between the involved populations (Villegas et al. 2017). The LC, LL  
342 and PR methods are extensions of the Lee-Carter model, while CBD-M6 is the extension of the  
343 CBD model. Li et al. (2015) proposed several two-population extensions of the CBD model,  
344 among which the CBD-M6 model is the final recommended one.

345 We find that the CoDa with the multivariate functional time series method using the RWD  
346 method produces the smallest point forecast errors for females. The CoDa with the multilevel  
347 functional time series method using the RWD method produces the smallest point forecast  
348 errors for males. While Table 1 presents an aggregated measure, we present horizon-specific  
349 forecast errors in Figure 4. In general, as the forecast horizon increases, the differences in  
350 forecast errors among the CoDa variants become more apparent.

351 The overall interval forecast error results of the CoDa method with various functional time  
352 series methods are presented in Table 2, where we average over the 30 forecast horizons. We  
353 find that the CoDa with the multivariate functional time series method using the automatically



**Figure 4:** A comparison of point forecast accuracy between the CoDa with functional time series methods using the holdout samples of the England & Wales period data. In the CoDa variants, we use the RWD forecasting method. The six benchmark methods are the random walk method, two-population Lee-Carter method of Zhou et al. (2014), Li-Lee method of Li and Lee (2005), the product-ratio method of Hyndman et al. (2013), two-population CBD-M6 model with VAR (or VECM) forecast of Li et al. (2015).

354 selected ARIMA of Hyndman and Khandakar (2008) produces the smallest interval forecast  
355 errors for females. In contrast, the CoDa with the multilevel functional time series method  
356 with the RWD produces the smallest interval forecast errors for male period data. Among the  
357 four univariate time series forecasting methods and three functional time series methods, the  
358 multilevel functional time series method with RWD produces the smallest interval forecast  
359 errors averaged over females and males, and thus is recommended to obtain both the point and  
360 interval forecasts.

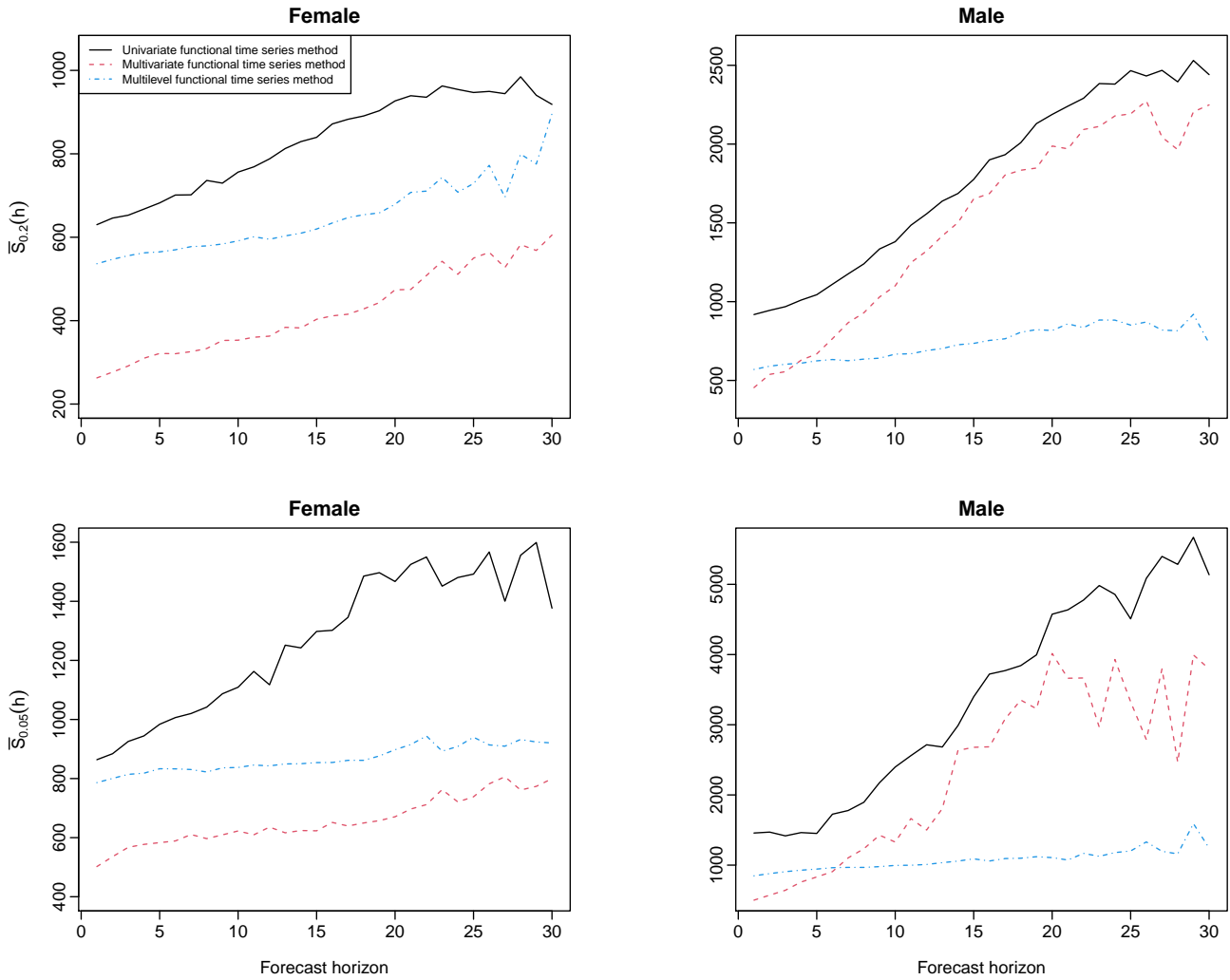
**Table 2:** A comparison of the interval forecast accuracy, as measured by the overall CPD and mean interval score among the CoDa methods using the holdout sample of the female and male data in England & Wales. Further, we consider four univariate time series forecasting methods for the CoDa.

Modelling method	Forecasting method	$\overline{\text{CPD}}$ (Female)		$\bar{S}_\alpha$ (Female)		$\overline{\text{CPD}}$ (Male)		$\bar{S}_\alpha$ (Male)	
		$\alpha = 0.2$	0.05	0.2	0.05	0.2	0.05	0.2	0.05
FTS	ARIMA	0.288	0.164	829.766	1267.737	0.339	0.241	1830.273	3954.051
	ETS	0.247	0.107	718.468	931.781	0.292	0.234	1958.649	4437.786
	RW	0.118	0.070	926.145	1437.361	0.284	0.229	1942.471	4306.580
	RWD	0.163	0.034	797.098	994.477	0.308	0.201	1781.903	3393.862
MFTS	ARIMA	<b>0.079</b>	<b>0.037</b>	<b>421.524</b>	<b>657.474</b>	0.331	0.293	2022.347	4731.384
	ETS	0.080	0.041	480.040	676.078	0.309	0.282	2120.377	5435.254
	RW	0.112	0.045	459.755	673.594	0.273	0.261	2037.320	5095.628
	RWD	0.158	0.049	443.686	796.824	0.215	0.148	1503.861	2344.673
MLFTS	ARIMA	0.094	0.045	650.225	866.905	0.139	0.045	941.893	1221.611
	ETS	0.139	0.046	701.098	949.348	0.119	0.047	1009.098	1330.439
	RW	0.141	0.045	711.635	918.514	0.122	0.046	1000.434	1327.669
	RWD	0.195	0.050	655.530	1056.286	<b>0.115</b>	<b>0.044</b>	<b>738.543</b>	<b>1077.238</b>


361 In contrast to the point forecasts, the RW method does not provide a prediction interval.  
362  
363 For the multi-population benchmark models, there is no established prediction interval under  
364 the CoDa framework in the literature. Thus, we report the interval forecast accuracy based on  
365 the functional time series methods using the nonparametric bootstrap approach described in  
366 Section 4 in Table 2.

367 While Table 2 presents the average over 30 forecast horizon, we show the one-step-ahead to  
368 30-step-ahead interval forecast errors in Figure 5. The differences in forecast accuracy among  
369 the CoDa methods become wider as the forecast horizon becomes longer. In the relatively

370 longer forecast horizon, the errors associated with all the methods become larger, but they are  
 371 relatively smaller for our recommended multilevel functional time series method with RWD.



**Figure 5:** A comparison of interval forecast accuracy between the CoDa with functional time series methods using the holdout samples of the England & Wales period data. We use the ARIMA forecasting method for females and the RWD forecasting method for males.

372 Among the four univariate time series forecasting methods, the RWD is generally recom-  
 373 mended for producing point forecasts. For producing interval forecasts, the RWD is recom-  
 374 mended for males but not so for females. The difference is how well the forecasts generated  
 375 from a statistical model capture the holdout data. For females, we found the ARIMA method  
 376 produces the smallest interval scores using the holdout samples of the England & Wales data  
 377 set. The ARIMA method can include higher-order lags than the RWD counterpart, the latter  
 378 method being a special case of AR(1). Since we don't know the underlying data generating  
 379 process, the most accurate method is data-driven, subject to the fitting period. Although a  
 380 model is better when it is a good proxy of the data generating process, it can not be gener-  
 381 alised. To facilitate reproducibility, the  code for implementing all the methods is available at  
 382 <https://github.com/hanshang/LTDC>.

## 383 7 Results on Swedish period life-table death counts

384 To check the robustness of our proposed methods, we perform the point and interval estimation  
 385 on age- and sex-specific period life-table death counts from 1861 to 2018 in Sweden, obtained  
 386 from the [Human Mortality Database \(2022\)](#). The comparison of the point and interval forecast  
 387 accuracy are presented in Table 3 and 4, respectively. In addition, we present horizon-specific  
 388 point and interval forecast errors in Figures 6 and 7, respectively. As the forecast horizon  
 389 increases, the differences in forecast errors among the CoDa variants become more apparent.

**Table 3:** *A comparison of the point forecast accuracy, as measured by the overall MAPE, KLD and JSD with simple and geometric means, among the CoDa with independent and joint modelling approaches and several benchmark methods using a holdout sample of the period life-table death counts for the female and male data in Sweden. Further, we consider four univariate time series forecasting methods for our functional time-series forecasting methods within the CoDa. ETS denotes the automatic exponential smoothing method. ARIMA denotes the automatic autoregressive integrated moving average method of Hyndman and Khandakar (2008). FTS denotes the univariate functional time series method. MFTS denotes the multivariate functional time series method. MLFTS denotes the multilevel functional time series method. The smallest errors are highlighted in bold.*

Modelling	Forecasting	Female				Male			
method	method	MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>	MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>
<u>CoDa</u>									
FTS	ARIMA	26.721	0.028	0.007	0.010	28.625	0.041	0.010	0.010
	ETS	28.385	0.020	0.005	0.011	29.431	0.042	0.011	0.011
	RW	37.155	0.061	0.015	0.018	48.421	0.073	0.018	0.018
	RWD	27.587	0.032	0.008	0.011	30.020	0.043	0.011	0.011
MFTS	ARIMA	26.085	0.032	0.008	0.014	34.673	0.058	0.014	0.014
	ETS	24.070	0.029	0.007	0.009	29.891	0.038	0.009	0.009
	RW	42.589	0.049	0.012	0.018	54.852	0.072	0.018	0.018
	RWD	<b>22.285</b>	0.019	0.005	0.009	29.806	0.036	0.009	0.009
MLFTS	ARIMA	26.737	0.020	0.005	0.008	26.521	0.033	0.008	0.008
	ETS	24.597	<b>0.017</b>	<b>0.004</b>	<b>0.004</b>	<b>22.568</b>	<b>0.017</b>	<b>0.004</b>	<b>0.004</b>
	RW	48.286	0.054	0.013	0.015	48.987	0.062	0.015	0.015
	RWD	25.907	0.025	0.006	0.007	26.678	0.028	0.007	0.007
<u>RW</u>		43.945	0.028	0.007	0.007	51.057	0.061	0.015	0.015
<u>LC</u>	VECM	27.022	0.029	0.007	0.007	27.216	0.035	0.009	0.009

Continued on next page

**Table 3:** A comparison of the point forecast accuracy, as measured by the overall MAPE, KLD and JSD with simple and geometric means, among the CoDa with independent and joint modelling approaches and several benchmark methods using a holdout sample of the period life-table death counts for the female and male data in Sweden. Further, we consider four univariate time series forecasting methods for our functional time-series forecasting methods within the CoDa. ETS denotes the automatic exponential smoothing method. ARIMA denotes the automatic autoregressive integrated moving average method of Hyndman and Khandakar (2008). FTS denotes the univariate functional time series method. MFTS denotes the multivariate functional time series method. MLFTS denotes the multilevel functional time series method. The smallest errors are highlighted in bold.

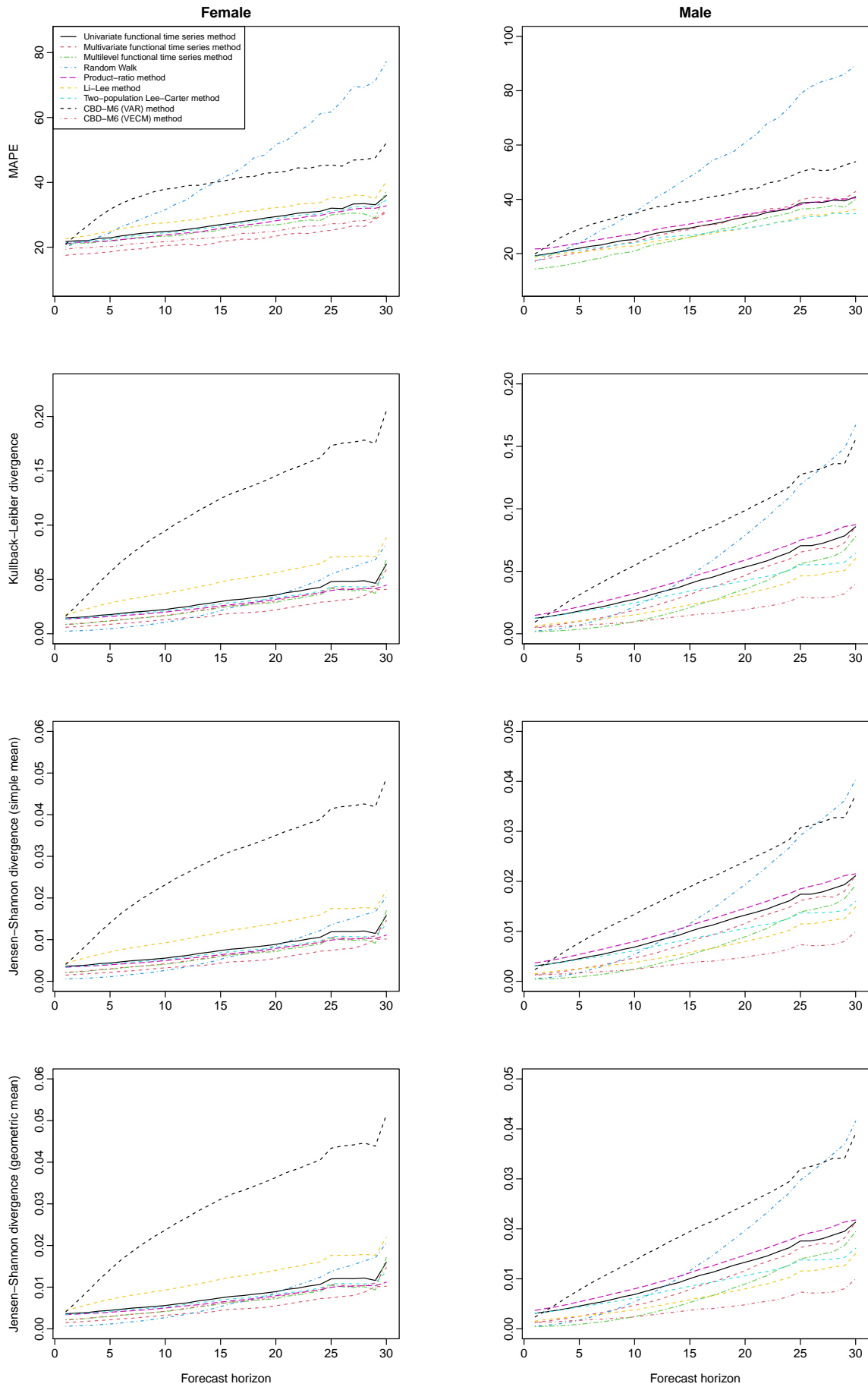
Modelling method	Forecasting method	Female				Male			
		MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>	MAPE	KLD	JSD <sup>s</sup>	JSD <sup>g</sup>
<u>LL</u>	RWD-AR	30.137	0.048	0.012	0.012	26.932	0.027	0.007	0.007
<u>PR</u>	ARIMA	26.373	0.027	0.012	0.012	31.226	0.048	0.012	0.012
<u>CBD-M6</u>	VAR	39.356	0.118	0.029	0.030	39.349	0.079	0.019	0.020
<u>CBD-M6</u>	VECM	23.890	0.026	0.006	0.007	26.002	<b>0.017</b>	<b>0.004</b>	<b>0.004</b>

390

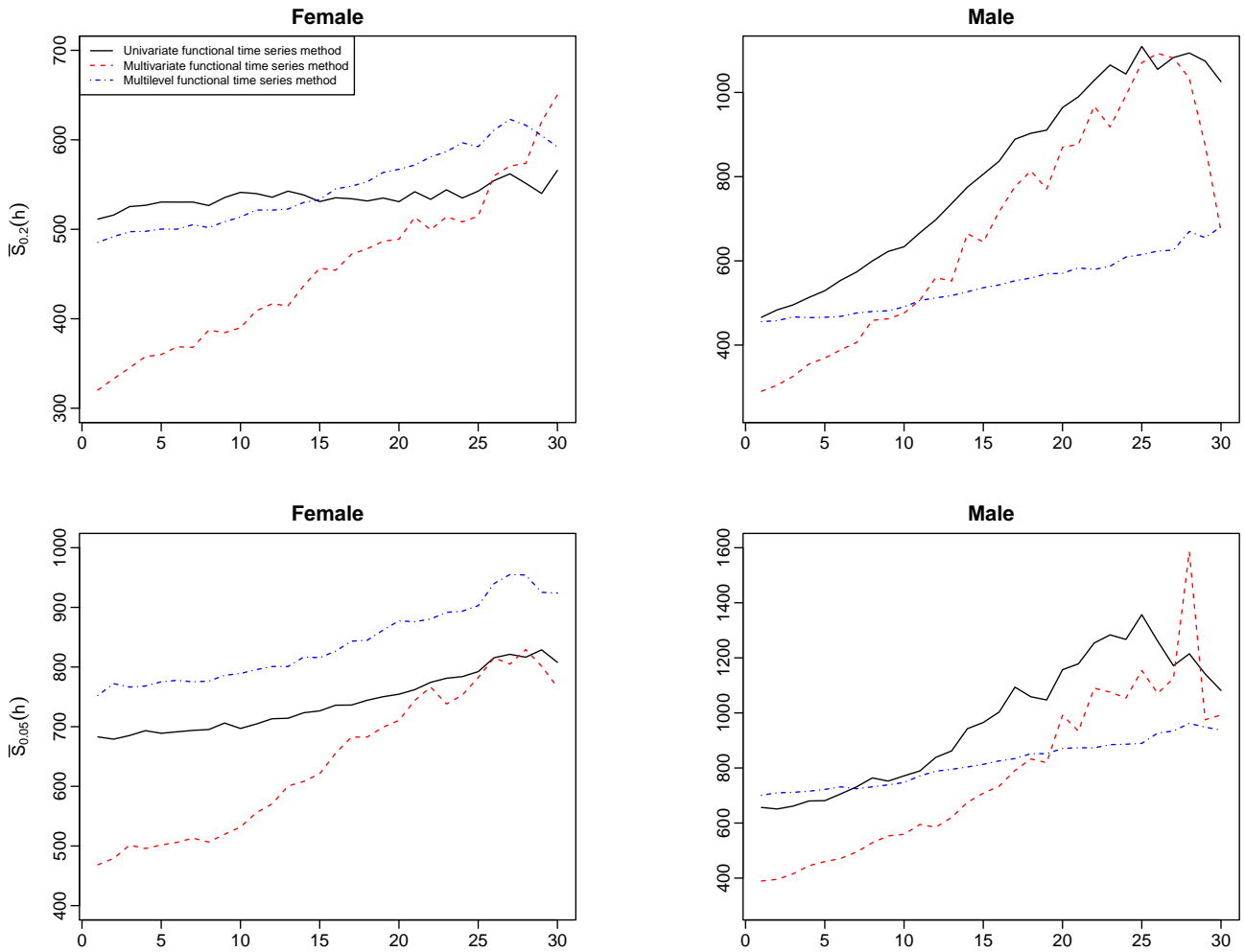
**Table 4:** A comparison of the interval forecast accuracy, as measured by the overall CPD and mean interval score among the CoDa methods using the holdout sample of the female and male data in Sweden. Further, we consider four univariate time series forecasting methods for the CoDa.

Modelling method	Forecasting method	$\overline{\text{CPD}}$ (Female)		$\bar{S}_\alpha$ (Female)		$\overline{\text{CPD}}$ (Male)		$\bar{S}_\alpha$ (Male)	
		$\alpha = 0.2$	0.05	0.2	0.05	0.2	0.05	0.2	0.05
FTS	ARIMA	0.142	0.098	555.147	729.946	0.164	0.057	727.857	828.368
	ETS	0.103	0.074	536.641	739.982	0.166	0.072	807.424	967.387
	RW	0.097	0.044	748.397	947.080	0.195	0.111	957.126	1408.082
	RWD	0.056	0.028	575.866	900.978	0.038	0.036	666.834	934.014
MFTS	ARIMA	<b>0.031</b>	<b>0.022</b>	466.119	629.841	0.176	0.103	919.075	1198.767
	ETS	0.054	0.027	<b>455.101</b>	640.374	0.096	0.036	676.307	<b>770.708</b>
	RW	0.034	0.023	478.209	<b>625.414</b>	0.178	0.082	865.841	1003.854
	RWD	0.101	0.036	553.178	1007.872	<b>0.023</b>	<b>0.032</b>	668.155	860.760
MLFTS	ARIMA	0.095	0.027	545.464	840.192	0.084	0.039	619.514	810.734
	ETS	0.099	0.030	547.777	837.673	0.110	0.039	<b>544.588</b>	822.576
	RW	0.112	0.040	595.128	865.717	0.103	0.037	664.062	835.116
	RWD	0.152	0.044	717.787	1244.656	0.155	0.047	724.160	1169.407

391



**Figure 6:** A comparison of point forecast accuracy between the CoDa with functional time series methods using the holdout samples of the Swedish period data. In the CoDa variants, we use the ETS forecasting method. The six benchmark methods are the random walk method, two-population Lee-Carter method of Zhou et al. (2014), Li-Lee method of Li and Lee (2005), the product-ratio method of Hyndman et al. (2013), two-population CBD-M6 model with VAR (or VECM) forecast of Li et al. (2015).



**Figure 7:** A comparison of interval forecast accuracy between the CoDa with functional time series methods using the holdout samples of the Swedish period data. In the CoDa variants, we use the ETS forecasting method.

392 For the point estimates, our proposed CoDa methods outperform the all the benchmark  
 393 models except CBD-M6 model with VECM forecasts, which performs on par with our proposed  
 394 methods for the male group. Among our proposed CoDa method, the CoDa with the multilevel  
 395 functional time series and ETS generally produces the smallest forecasting error for Swedish  
 396 period data. The multivariate functional times series model generally performs best regarding  
 397 the interval forecast accuracy. At the same time, there is no universally best univariate time-  
 398 series forecasting method to forecast the FPC scores.

## 399 8 Conclusion

400 We extend the CoDa from univariate to multivariate and multilevel functional time series  
 401 methods for modelling multiple populations. Within the CoDa framework, we take the log-ratio  
 402 transformation to obtain unconstrained data. Then, we apply the multivariate and multilevel  
 403 functional time series methods to model the unconstrained data for multiple populations. After

404 applying the back-transformation, we obtain the age-specific life-table death counts for multiple  
405 populations.

406 We compare our proposed CoDa methods with six benchmarks in the single-population  
407 forecast (simple random walk) and multi-population forecasts (two-population Lee-Carter  
408 model, Li-Lee method, product-ratio method and two-population CBD model with cohort  
409 effects using VAR/VECM for prediction) on both period and cohort life-table death counts  
410 in England & Wales and Sweden. Our proposed CoDa methods outperform the existing  
411 benchmarks in most cases. In contrast, the product-ratio method and CBD-M6 (with VECM  
412 forecasts) perform on par or with a slight edge under certain error measurements for the  
413 Swedish male life-table death counts. Among our proposed methods, it is difficult to single out  
414 a universal univariate time-series forecasting method that outperforms the rest in all cases.

415 There are a few ways in which this paper could be extended, and we briefly discuss four.  
416 1) To obtain more stable forecasts, simple averaging forecasts from three functional time series  
417 models may result in improved forecast accuracy. In addition, we recommend selecting the  
418 univariate time series forecasting method via an information criterion before fitting the model.  
419 2) A robust CoDa method proposed by [Filzmoser et al. \(2009\)](#) may be utilised in the presence  
420 of outlying years. These outlying years are from a different data generating process than the  
421 non-outlying observations. 3) We could consider some of the other methods in the literature for  
422 extending data sets for non-extinct cohorts (e.g., [Basellini et al. 2020](#)) and suggest examining  
423 the sensitivity of the results to the choice of method. 4) We could also extend from considering  
424 two populations in the current paper to considering life-table data for multiple populations and  
425 for multiple subsets of a population (for example, subdivided by socio-economic status). In  
426 these scenarios, the number of populations could exceed the number of functional curves in a  
427 population, and this then leads to the high-dimensional functional time series analysis (see, e.g.,  
428 [Gao et al. 2019](#), [Tang and Shi 2021](#)).

## 429 **Acknowledgment**

430 The authors thanks insightful comments and suggestions from two reviewers. The first author  
431 acknowledges comments and suggestions from participants of a one-world actuarial research  
432 seminar in 2020, and a seminar at the Australian Government Actuary in 2022.

## 433 **References**

- 434 Aburto, J. M. and van Raalte, A. A. (2018), 'Lifespan dispersion in times of life expectancy  
435 fluctuation: The case of central and eastern Europe', *Demography* **55**, 2071–2096.
- 436 Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors',  
437 *Econometrica* **81**(3), 1203–1227.
- 438 Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical*  
439 *Society: Series B* **44**(2), 139–177.
- 440 Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman & Hall, London.
- 441 Aitchison, J. and Shen, S. M. (1980), 'Logistic-normal distributions: Some properties and uses',  
442 *Biometrika* **67**(2), 261–272.
- 443 Aue, A., Norinho, D. D. and Hörmann, S. (2015), 'On the prediction of stationary functional  
444 time series', *Journal of the American Statistical Association: Theory and Methods* **110**(509), 378–392.
- 445 Bagchi, P., Characiejus, V. and Dette, H. (2018), 'A simple test for white noise in functional time  
446 series', *Journal of Time Series Analysis* **39**(1), 54–74.
- 447 Basellini, U., Kjaergaard, S. and Camarda, C. G. (2020), 'An age-at-death distribution approach  
448 to forecast cohort mortality', *Insurance: Mathematics & Economics* **91**, 129–143.
- 449 Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J. and Vaupel, J. W. (2017), 'Coherent  
450 forecasts of mortality with compositional data analysis', *Demographic Research* **37**, 527–566.
- 451 Bergeron-Boucher, M.-P., Simonacci, V., Oeppen, J. and Gallo, M. (2018), 'Coherent modeling and  
452 forecasting of mortality patterns for subpopulations using multiway analysis of compositions:  
453 An application to Canadian provinces and territories', *The North American Actuarial Journal*  
454 **22**(1), 92–118.
- 455 Booth, H. (2006), 'Demographic forecasting: 1980 to 2005 in review', *International Journal of*  
456 *Forecasting* **22**(3), 547–581.
- 457 Booth, H. and Tickle, L. (2008), 'Mortality modelling and forecasting: A review of methods',  
458 *Annals of Actuarial Science* **3**(1-2), 3–43.
- 459 Cairns, A. J. G., Blake, D. and Dowd, K. (2008), 'Modelling and management of mortality risk:  
460 A review', *Scandinavian Actuarial Journal* **2008**(2-3), 79–113.

- 461 Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D. and Khalaf-Allah, M. (2011), 'Bayesian  
462 stochastic mortality modelling for two populations', *ASTIN Bulletin: The Journal of the IAA*  
463 **41**(1), 29–59.
- 464 Canudas-Romo, V. (2010), 'Three measures of longevity: Time trends and record values',  
465 *Demography* **47**(2), 299–312.
- 466 Chatfield, C. (2000), *Time-Series Forecasting*, 1st edn, Chapman and Hall/CRC, New York.
- 467 Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C. and Caselli, G. (2005), 'Three dimensions of the  
468 survival curve: Horizontalization, verticalization, and longevity extension', *Demography*  
469 **42**(2), 243–258.
- 470 Debón, A., Chaves, L., Haberman, S. and Villa, F. (2017), 'Characterization of between-group  
471 inequality of longevity in EU countries', *Insurance: Mathematics and Economics* **75**, 151–165.
- 472 Delicado, P. (2011), 'Dimensionality reduction when data are density functions', *Computational*  
473 *Statistics and Data Analysis* **55**(1), 401–420.
- 474 Denuit, M., Devolder, P. and Goderniaux, A.-C. (2007), 'Securitization of longevity risk: Pricing  
475 survivor bonds with Wang transform in the Lee-Carter framework', *The Journal of Risk and*  
476 *Insurance* **74**(1), 87–113.
- 477 Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009), 'Multilevel functional  
478 principal component analysis', *Annals of Applied Statistics* **3**(1), 458–488.
- 479 Dickson, D. C. M., Hardy, M. R. and Waters, H. R. (2009), *Actuarial Mathematics for Life Contingent*  
480 *Risks*, Cambridge University Press, Cambridge.
- 481 Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D. and Khalaf-Allah, M. (2011), 'A grav-  
482 ity model of mortality rates for two related populations', *North American Actuarial Journal*  
483 **15**(2), 334–356.
- 484 Filzmoser, P., Hron, K. and Reimann, C. (2009), 'Principal component analysis for compositional  
485 data with outliers', *Environmetrics* **20**(6), 621–632.
- 486 Fuglede, B. and Topsoe, F. (2004), Jensen-Shannon divergence and Hilbert space embedding, *in*  
487 'Proceedings of International Symposium on Information Theory', IEEE, Chicago, IL, USA.
- 488 Gao, Y., Shang, H. L. and Yang, Y. (2019), 'High-dimensional functional time series forecasting:  
489 An application to age-specific mortality rates', *Journal of Multivariate Analysis* **170**, 232–243.

- 490 Gneiting, T. and Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and*  
491 *Its Application* **1**, 125–151.
- 492 Gneiting, T. and Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction and estimation',  
493 *Journal of the American Statistical Association: Review Article* **102**(477), 359–378.
- 494 Goldstein, J. R. and Wachter, K. W. (2006), 'Relationships between period and cohort life  
495 expectancy: Gaps and lags', *Population Studies* **60**(3), 257–269.
- 496 Hatzopoulos, P. and Haberman, S. (2013), 'Common mortality modelling and coherent forecasts.  
497 An empirical analysis of worldwide mortality data', *Insurance: Mathematics and Economics*  
498 **52**(2), 320–337.
- 499 Horiuchi, S., Ouellette, N., Cheung, S. L. K. and Robine, J.-M. (2013), 'Modal age at death:  
500 lifespan indicator in the era of longevity extension', *Vienna Yearbook of Population Research*  
501 **11**, 37–69.
- 502 Human Mortality Database (2022), *University of California, Berkeley (USA), and Max Planck*  
503 *Institute for Demographic Research (Germany)*. Accessed on August 26, 2021.  
504 URL: <http://www.mortality.org>
- 505 Hyndman, R. J., Booth, H. and Yasmeen, F. (2013), 'Coherent mortality forecasting: the product-  
506 ratio method with functional time series models', *Demography* **50**(1), 261–283.
- 507 Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast  
508 package for R', *Journal of Statistical Software* **27**(3).
- 509 Hyndman, R. J. and Shang, H. L. (2009), 'Forecasting functional time series (with discussions)',  
510 *Journal of the Korean Statistical Society* **38**(3), 199–221.
- 511 Jarner, S. F. and Kryger, E. M. (2011), 'Modelling adult mortality in small populations: The Saint  
512 model', *ASTIN Bulletin: The Journal of the IAA* **41**(2), 377–418.
- 513 Kokoszka, P., Miao, H., Petersen, A. and Shang, H. L. (2019), 'Forecasting of density functions  
514 with an application to cross-sectional and intraday returns', *International Journal of Forecasting*  
515 **35**(4), 1304–1317.
- 516 Kullback, S. and Leibler, R. A. (1951), 'On information and sufficiency', *Annals of Mathematical*  
517 *Statistics* **22**(1), 79–86.
- 518 Li, D., Robinson, P. M. and Shang, H. L. (2020), 'Long-range dependent curve time series',  
519 *Journal of the American Statistical Association: Theory and Methods* **115**(530), 957–971.

- 520 Li, J. S.-H., Zhou, R. and Hardy, M. (2015), 'A step-by-step guide to building two-population  
521 stochastic mortality models', *Insurance: Mathematics and Economics* **63**, 121–134.
- 522 Li, N. and Lee, R. (2005), 'Coherent mortality forecasts for a group of populations: An extension  
523 of the Lee–Carter method', *Demography* **42**(3), 575–594.
- 524 Oeppen, J. (2008), Coherent forecasting of multiple-decrement life tables: A test using Japanese  
525 cause of death data, in 'European Population Conference', Barcelona, Spain.  
526 URL: <http://epc2008.princeton.edu/papers/80611>
- 527 Pawlowsky-Glahn, V., Egozcue, J. and Tolosana-Delgado, R. (2015), *Modeling and Analysis of*  
528 *Compositional Data*, John Wiley & Sons, Ltd, Chichester.
- 529 Pollard, J. H. (1987), 'Projection of age-specific mortality rates', *Population Bulletin of the United*  
530 *Nations* **21-22**, 55–69.
- 531 Preston, S., Heuveline, P. and Guillot, M. (2001), *Demography: Measuring and Modeling Population*  
532 *Processes*, Blackwell Publishers, Oxford, U.K.
- 533 Rizzi, S., Kjærgaard, S., Boucher, M.-P. B., Camarda, C. G., Lindahl-Jacobsen, R. and Vaupel, J. W.  
534 (2021), 'Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized  
535 composite link model', *International Journal of Forecasting* **37**(1), 95–104.
- 536 Robine, J.-M. (2001), 'Redefining the stages of the epidemiological transition by a study of the  
537 dispersion of life spans: The case of France', *Population: An English Selection* **13**(1), 173–193.
- 538 Russolillo, M., Giordano, G. and Haberman, S. (2011), 'Extending the Lee-Carter model: A  
539 three-way decomposition', *Scandinavian Actuarial Journal* **2011**(2), 96–117.
- 540 Scealy, J. L., de Caritat, P., Grunsky, E. C., Tsagris, M. T. and Welsh, A. H. (2017), 'Robust  
541 principal component analysis for power transformed compositional data', *Journal of the*  
542 *American Statistical Association: Theory and Methods* **110**(509), 136–148.
- 543 Scealy, J. L. and Welsh, A. H. (2017), 'A directional mixed effects model for compositional  
544 expenditure data', *Journal of the American Statistical Association: Applications and Case Studies*  
545 **112**(517), 24–36.
- 546 Shang, H. L. (2016), 'Mortality and life expectancy forecasting for a group of populations  
547 in developed countries: A multilevel functional data method', *Annals of Applied Statistics*  
548 **10**(3), 1639–1672.

- 549 Shang, H. L., Booth, H. and Hyndman, R. J. (2011), 'Point and interval forecasts of mortality  
550 rates and life expectancy: A comparison of ten principal component methods', *Demographic  
551 Research* **25**(5), 173–214.
- 552 Shang, H. L. and Haberman, S. (2020), 'Forecasting age distribution of death counts: An  
553 application to annuity pricing', *Annals of Actuarial Science* **14**(1), 150–169.
- 554 Shang, H. L. and Kearney, F. (2022), 'Dynamic functional time-series forecasts of foreign ex-  
555 change implied volatility surfaces', *International Journal of Forecasting* **38**(3), 1025–1049.
- 556 Shang, H. L., Smith, P. W. F., Bijak, J. and Wiśniowski, A. (2016), 'A multilevel functional data  
557 method for forecasting population, with an application to the United Kingdom', *International  
558 Journal of Forecasting* **32**(3), 629–649.
- 559 Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell Labs Technical Journal*  
560 **27**(3), 379–423.
- 561 Shkolnikov, V. M., Andreev, E. E. and Begun, A. Z. (2003), 'Gini coefficient as a life table function:  
562 computation from discrete data, decomposition of differences and empirical examples',  
563 *Demographic Research* **8**, 305–358.
- 564 Tang, C. and Shi, Y. (2021), 'Forecasting high-dimensional financial functional time series: An  
565 application to constituent stocks in Dow Jones index', *Journal of Risk and Financial Management*  
566 **14**(343).
- 567 van Raalte, A. A. and Caswell, H. (2013), 'Perturbation analysis of indices of lifespan variability',  
568 *Demography* **50**(5), 1615–1640.
- 569 van Raalte, A. A., Martikainen, P. and Myrskylä, M. (2014), 'Lifespan variation by occupational  
570 class: Compression or stagnation over time?', *Demography* **51**, 73–95.
- 571 Vaupel, J. W., Zhang, Z. and van Raalte, A. A. (2011), 'Life expectancy and disparity: An  
572 international comparison of life table data', *BMJ Open* **1**(1), e000128.
- 573 Villegas, A., Haberman, S., Kaishev, V. and Millosovich, P. (2017), 'A comparative study of two  
574 population models for the assessment of basis risk in longevity hedges', *ASTIN Bulletin: The  
575 Journal of the IAA* **47**(3), 631–679.
- 576 Wilmoth, J. R. and Horiuchi, S. (1999), 'Rectangularization revisited: Variability of age at death  
577 within human populations', *Demography* **36**(4), 475–495.

- 578 Zhang, C., Kokoszka, P. and Petersen, A. (2022), 'Wasserstein autoregressive models for density  
579 time series', *Journal of Time Series Analysis* **43**(1), 30–52.
- 580 Zhou, R., Wang, Y., Kaufhold, K., Li, J. S.-H. and Tan, K. S. (2014), 'Modeling period effects in  
581 multi-population mortality models: Applications to Solvency II', *North American Actuarial*  
582 *Journal* **18**(1), 150–167.
- 583 Zivot, E. and Wang, J. (2006), *Modeling Financial Time Series with S-PLUS*, Springer, New York.