City Research Online

City, University of London Institutional Repository

Forecasting age distribution of death counts: An application to annuity pricing

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

Steven Haberman
Cass Business School
City, University of London

July 31, 2019

Abstract

We consider a compositional data analysis approach to forecasting the age distribution of death counts. Using the age-specific period life-table death counts in Australia obtained from the Human Mortality Database, the compositional data analysis approach produces more accurate one- to 20-step-ahead point and interval forecasts than Lee-Carter method, Hyndman-Ullah method, and two naïve random walk methods. The improved forecast accuracy of period life-table death counts is of great interest to demographers for estimating survival probabilities and life expectancy, and to actuaries for determining temporary annuity prices for various ages and maturities. Although we focus on temporary annuity prices, we consider long-term contracts which make the annuity almost lifetime, in particular when the age at entry is sufficiently high.

Keywords: Compositional data analysis; Life-table death counts; Log-ratio transformation; Principal component analysis; Single-premium temporary annuity

---

[*]Correspondence to: Han Lin Shang, Research School of Finance, Actuarial Studies and Statistics, Level 4, Building 26C, Kingsley Street, Australian National University, Kingsley Street, Acton, Canberra, ACT 2601, Australia; Telephone: +61(2) 612 50535; Fax: +61(2) 612 50087; Email: hanlin.shang@anu.edu.au

1 Introduction

Actuaries have produced forecasts of mortality since the beginning of the 20th century, in response to the adverse financial effects of mortality improvements over time on life annuities and pensions (Pollard 1987). Thus, projected mortality tables for annuitants was one of the topics discussed at the 5th International Congress of Actuaries, Berlin in 1906 (Cramér and Wold 1935). Several authors have proposed new approaches for forecasting age-specific central mortality rates using statistical models (see Booth 2006, Booth and Tickle 2008, Cairns et al. 2008, Shang et al. 2011, for reviews). Instead of modelling central mortality rates, we consider a compositional data analysis (CoDa) approach for modelling and forecasting the age-specific numbers of deaths in period life tables. Both central mortality rates or life-table death counts can be derived from the other based on standard life-table relations (for detail on the life table and its indicators, see Preston et al. (2001), Chapter 3, or Dickson et al. (2009), Chapters 2-3). By using the life-table death distribution, we could model and forecast a redistribution of the density of life-table deaths, where deaths at younger ages are shifted towards older ages. Alternatively, we may consider a cohort life table which depicts the life history of a specific group of individuals but is dependent on projected mortality rates for those cohorts born more recently. Instead, we choose to study the period life table which represents the mortality conditions in a period of time (see also Oeppen 2008, Bergeron-Boucher et al. 2017).

In the field of demography, Oeppen (2008) and Bergeron-Boucher et al. (2017) have put forward a principal component approach to forecast life-table death counts within a CoDa framework by considering age-specific life-table death count (d_x) as compositional data. As with compositional data, the data are constrained to vary between two limits (e.g., 0 and a constant upper bound), which conditions their covariance structure. This feature can represent great advantages in a forecasting context (Lee 1998). Thus, Oeppen (2008) demonstrated that using CoDa to forecast age-specific mortality does not lead to more pessimistic results than forecasting age-specific mortality (see, e.g., Wilmoth 1995). Apart from providing an informative description of the mortality experience of a population, the age-at-death distribution yields readily available information on the "central longevity indicators" (e.g., mean, median and modal age at death, see Cheung et al. 2005, Canudas-Romo 2010) as well as lifespan variability (e.g., Robine 2001, Vaupel et al. 2011, van Raalte and Caswell 2013, van Raalte et al. 2014, Aburto and van Raalte 2018).

Compositional data arise in many other scientific fields, such as geology (geochemical elements), economics (income/expenditure distribution), medicine (body composition), food industry (food composition), chemistry (chemical composition), agriculture (nutrient balance bionomics), environmental sciences (soil contamination), ecology (abundance of different

species), and demography (life-table death counts). In the field of statistics, Scealy et al. (2015) use CoDa to study the concentration of chemical elements in sediment or rock samples. Scealy and Welsh (2017) applied CoDa to analyse total weekly expenditure on food and housing costs for households in a chosen set of domains. Delicado (2011) and Kokoszka et al. (2019) use CoDa to analyse density functions and implement dimension-reduction techniques on the constrained compositional data space.

Compositional data are defined as a random vector of K positive components $\boldsymbol{D} = [d_1, \ldots, d_K]$ with strictly positive values whose sum is a given constant, set typically equal to 1 (portions), 100 (percentage) and 10^6 for parts per million (ppm) in geochemical trace element compositions (Aitchison 1986, p. 1). The sample space of compositional data is thus the simplex

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

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

Compositional data are subject to a sum constraint, which in turn imposes unpleasant constraints upon the variance-covariance structure of the raw data. The standard approach involves breaking the sum constraint using a transformation of the raw data to remove the constraint, before applying conventional statistical techniques to the transformed data. Among all possible transformations, the family of log-ratio transformations is commonly used. This family includes the additive log-ratio, the multiple log-ratio, the centred log-ratio transformations (Aitchison and Shen 1980, Aitchison 1982, 1986), and the isometric log-ratio transformation (Egozcue et al. 2003).

The contributions of this paper are threefold: First, as the CoDa framework of Oeppen (2008) is an adaptation of the Lee-Carter model to compositional data, our work could be seen as an adaptation of Hyndman and Ullah's (2007) to a CoDa framework. Second, we apply the CoDa method of Oeppen (2008) to model and forecast the age distribution of life-table death counts, from which we obtain age-specific survival probabilities, and we determine immediate temporary annuity prices. Third, we propose a nonparametric bootstrap method for constructing prediction intervals for the future age distribution of life-table death counts.

Using the Australian age- and sex-specific life-table death counts from 1921 to 2014, we evaluate and compare the one- to 20-step-ahead point forecast accuracy and interval forecast accuracy among the CoDa, Hyndman-Ullah (HU) method, Lee-Carter (LC) method and two naïve random walk methods. To evaluate point forecast accuracy, we use the mean absolute percentage error (MAPE). To assess interval forecast accuracy, we utilise the interval score of Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014), see Section 5.1 for details. Regarding both point forecast accuracy and interval forecast accuracy, the CoDa method

performs the best overall among the three methods which we have considered. The improved forecast accuracy of life-table death counts is of great importance to actuaries for determining remaining life expectancies and pricing temporary annuities for various ages and maturities.

The remainder of this paper is organised as follows: Section 2 describes the data set, which is Australian age- and sex-specific life-table death counts from 1921 to 2014. Section 3 introduces the CoDa method for producing the point and interval forecasts of the age distribution of life-table death counts. Section 4 studies the goodness-of-fit of the CoDa method and provides an example for generating point and interval forecasts. Using the MAPE and mean interval score in Section 5, we evaluate and compare the point and interval forecast accuracies among the methods considered. Section 6 applies the CoDa method to estimate the single-premium temporary immediate annuity prices for various ages and maturities for a *female* policyholder residing in Australia. Conclusions are presented in Section 7, along with some reflections on how the methods presented here can be extended.

2 Australian age- and sex-specific life-table death counts

We consider Australian age- and sex-specific life-table death counts from 1921 to 2014, obtained from the Human Mortality Database (2019). Although we use all data that include the first and second World War periods, a feature of our proposed method is to use an automatic algorithm of Hyndman and Khandakar (2008) to select the optimal data-driven parameters in exponential smoothing forecasting method. As with exponential smoothing forecasting method, it assigns more weights to the recent data than distant past data. In practice, it is likely that only a few of the most distant past data may influence our forecasts, regardless of the starting point.

We study life-table death counts, where the life-table radix (i.e., a population experiencing 100,000 births annually) is fixed at 100,000 at age 0 for each year. For the life-table death counts, there are 111 ages, and these are age 0, 1, $\cdots$, 109, 110+. Due to rounding, there are zero counts for age 110+ at some years. To rectify this problem, we prefer to use the probability of dying (i.e., q_x) and the life-table radix to recalculate our estimated death counts (up to 6 decimal places). In doing so, we obtain more detailed death counts than the ones reported in the Human Mortality Database (2019).

To understand the principal features of the data, Figure 1 presents rainbow plots of the female and male age-specific life-table death counts in Australia from 1921 to 2014 in a single-year group. The time ordering of the curves follows the colour order of a rainbow, where data from the distant past are shown in red and the more recent data are shown in purple (see Hyndman and Shang 2010, for other examples). Both figures demonstrate a decreasing trend in infant death counts and a typical negatively-skewed distribution for the life-table death counts,
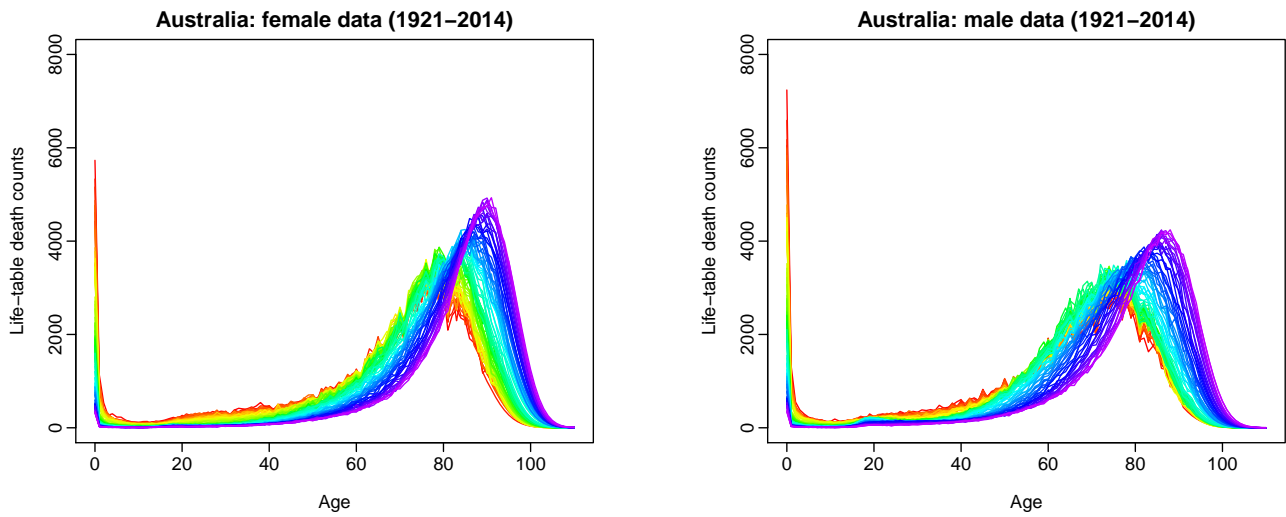
Figure 1: *Rainbow plots of age-specific life-table death count from 1921 to 2014 in a single-year group. The oldest years are shown in red, with the most recent years in violet. Curves are ordered chronologically according to the colours of the rainbow.*

where the peaks shift to higher ages for both females and males. This shift is a primary source of the longevity risk, which is a major issue for insurers and pension funds, especially in the selling and risk management of annuity products (see Denuit et al. 2007, for a discussion). Moreover, the spread of the distribution indicates lifespan variability. A decrease in variability over time can be observed directly and can be measured, for example with the interquartile range of life-table ages at death or the Gini coefficient (for comprehensive reviews, see Wilmoth and Horiuchi 1999, van Raalte and Caswell 2013, Debón et al. 2017). Age-at-death distributions thus provide critical insights on longevity and lifespan variability that cannot be grasped directly from either mortality rates or the survival function.

3 Forecasting method

3.1 Compositional data analysis (CoDa)

In the field of demography, Oeppen (2008) and Bergeron-Boucher et al. (2017) have laid the foundations by presenting a modelling and forecasting framework. Following Oeppen (2008) and Bergeron-Boucher et al. (2017), a CoDa method can be summarised as follows:

1) We begin from a data matrix D of size $n \times K$ of the life-table death counts $(d_{t,x})$ with n rows representing the number of years and K columns representing the age x. The sum of each row adds up to the life-table radix, such as 100,000. Since we are working with life-table death counts, it is not necessary to have population-at-risk estimates.

2) We compute the geometric mean at each age, given by

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

For a given year t, we divide $(d_{t,1},\ldots,d_{t,K})$ by the corresponding geometric means $(\alpha_1,\ldots,\alpha_K)$, then standardise all elements so as to sum up to unity. As with compositional data, it is more important to know a relative proportion of each component (i.e., age) than the sum of all components. Although the life-table radix is 100,000 customarily, we model and forecast relative proportions, which are then multiplied by 100,000 to obtain forecast life-table death counts. Via standardisation, this is expressed as

$$f_{t,x} = \frac{d_{t,x}/\alpha_x}{\sum_{x=1}^{K} d_{t,x}/\alpha_x},$$

where $f_{t,x}$ denotes de-centred data.

3) Log-ratio transformation: Aitchison (1982, 1986) showed that compositional data are represented in a restricted space where the components can only vary between 0 and a positive constant. Therefore, Aitchison (1982, 1986) proposed a log-ratio transformation to transform the data into a real-valued space. We apply the centred log-ratio transformation, given by

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

where g_t denotes the geometric mean over the age at time t, given by

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

The transformed data matrix is denoted as $\mathbf{Z}$ with elements $z_{t,x} \in R$, where R denotes real-valued space.

4) Principal component analysis: Principal component analysis is then applied to the matrix $\mathbf{Z}$, to obtain the estimated principal components and their scores,

$$z_{t,x} = \sum_{\ell=1}^{\min(n,K)} \widehat{\beta}_{t,\ell}\widehat{\phi}_{\ell,x} = \sum_{\ell=1}^{L} \widehat{\beta}_{t,\ell}\widehat{\phi}_{\ell,x} + \widehat{\omega}_{t,x}, \tag{2}$$

where $\widehat{\omega}_{t,x}$ denotes model residual term for age x in year t, $\{\widehat{\phi}_{1,x},\ldots,\widehat{\phi}_{L,x}\}$ represents the first L sets of estimated principal components, $\{\widehat{\beta}_{t,1},\ldots,\widehat{\beta}_{t,L}\}$ represents the first L sets of estimated principal component scores for time t, and L denotes the number of retained principal components.

5) Forecast of principal component scores: Via a univariate time series forecasting method, such as exponential smoothing (ETS), we obtain the h-step-ahead forecast of the ℓth principal component score $\widehat{\beta}_{n+h,\ell}$, where h denotes forecast horizon (see also Hyndman and Ullah

2007). We utilise an automatic algorithm developed by Hyndman and Khandakar (2008) to determine the optimal ETS model based on the corrected Akaike Information Criterion (Hurvich and Tsai 1993). Conditioning on the estimated principal components and observed data, the forecast of $z_{n+h,x}$ can be obtained by

$$\widehat{z}_{n+h|n,x} = \sum_{\ell=1}^{L} \widehat{\beta}_{n+h|n,\ell} \widehat{\phi}_{\ell,x}. \tag{3}$$

In (3), the principal component scores can be modelled and forecasted by an ETS method. To select the optimal ETS parameter, we use the automatic search algorithm of Hyndman and Khandakar (2008) by the smallest corrected Akaike information criterion. In Table 1, we also consider three other forecasting methods, namely autoregressive integrated moving average (ARIMA), random walk with drift (RWD) and random walk without drift (RW).

6) Transform back to the compositional data: we take the inverse centred log-ratio transformation, given by

$$\widehat{f}_{n+h|n} = \left[\frac{\exp^{\widehat{z}_{n+h|n,1}}}{\sum_{x=1}^{K} \exp^{\widehat{z}_{n+h|n,x}}}, \frac{\exp^{\widehat{z}_{n+h|n,2}}}{\sum_{x=1}^{K} \exp^{\widehat{z}_{n+h|n,x}}}, \cdots, \frac{\exp^{\widehat{z}_{n+h|n,K}}}{\sum_{x=1}^{K} \exp^{\widehat{z}_{n+h|n,x}}} \right],$$

where $\widehat{z}_{n+h|n,x}$ denotes the forecasts in (3).

7) Finally, we add back the geometric means, to obtain the forecasts of the life-table death matrix d_{n+h},

$$\widehat{d}_{n+h|n} = \left[\frac{\widehat{f}_{n+h|n,1} \times \alpha_1}{\sum_{x=1}^{K} \widehat{f}_{n+h|n,x} \times \alpha_x}, \frac{\widehat{f}_{n+h|n,2} \times \alpha_2}{\sum_{x=1}^{K} \widehat{f}_{n+h|n,x} \times \alpha_x}, \cdots, \frac{\widehat{f}_{n+h|n,K} \times \alpha_K}{\sum_{x=1}^{K} \widehat{f}_{n+h|n,x} \times \alpha_x} \right],$$

where α_x is the age-specific geometric mean given in (1).

To determine the number of components L in (2) and (3), we consider a criterion known as cumulative percentage of variance (CPV), i.e., define the value of L as the minimum number of components that reaches a certain level of the proportion of total variance explained by the L leading components such that

$$L = \underset{L:L \geq 1}{\operatorname{argmin}} \left\{ \sum_{\ell=1}^{L} \widehat{\lambda}_{\ell} \bigg/ \sum_{\ell=1}^{n} \widehat{\lambda}_{\ell} \mathbb{1}_{\{\widehat{\lambda}_{\ell} > 0\}} \geq \delta \right\},$$

where $\delta = 85\%$ (see also Horváth and Kokoszka 2012, p.41) and $\mathbb{1}_{\{\cdot\}}$ denotes the binary indicator function which excludes possible zero eigenvalues. For the Australian female and male data, the chosen number of components $L = 1$ and $L = 2$, respectively.

We highlight some similarities and differences between Oeppen's (2008) approach and Lee and Carter's (1992) approach. From Steps 3 to 5, Oeppen's (2008) approach uses the Lee and Carter's (1992) approach to model and forecast the log-ratio of life-table death counts.

The difference is that the Oeppen's (2008) approach works with life-table death counts in a constrained space, whereas the Lee and Carter's (1992) approach works with real-valued log mortality rates.

When the number of components $L = 1$, our proposed CoDa method corresponds to the one presented by Oeppen (2008) and Bergeron-Boucher et al. (2017). However, we allow the possibility of using more than one pair of principal component and principal component scores in Steps 4 and 5. As a sensitivity analysis, we also consider setting the number of components to be $L = 6$ (see also Hyndman and Booth 2008). From this aspect, our proposal shares some similarity with the Hyndman and Ullah's (2007) approach. The difference is that our proposal works with life-table death counts in a constrained space, whereas the Hyndman and Ullah's (2007) approach works with real-valued log mortality rates.

3.2 Construction of prediction interval for the CoDa

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. The forecast uncertainty stems from both systematic deviations (e.g., due to parameter or model uncertainty) and random fluctuations (e.g., due to model error term). As was emphasised by Chatfield (1993, 2000), it is essential to provide interval forecasts as well as point forecasts to

1) assess future uncertainty levels;

2) enable different strategies to be planned for a range of possible outcomes indicated by the interval forecasts;

3) compare forecasts from different methods more thoroughly; and

4) explore different scenarios based on various assumptions.

3.2.1 Proposed bootstrap method

We consider two sources of uncertainty: truncation errors in the principal component decomposition and forecast errors in the projected principal component scores. Since principal component scores are regarded as surrogates of the original functional time series, these principal component scores capture the temporal dependence structure inherited in the original functional time series (see also Salish and Gleim 2015, Paparoditis 2018, Shang 2018). By adequately bootstrapping the forecast principal component scores, we can generate a set of bootstrapped $\boldsymbol{Z}^*$, conditional on the estimated mean function and estimated functional principal components from the observed Z in (3).

Using a univariate time series model, we can obtain multi-step-ahead forecasts for the principal component scores, $\{\widehat{\beta}_{1,\ell}, \ldots, \widehat{\beta}_{n,\ell}\}$ for $l = 1, \ldots, L$. Let the h-step-ahead forecast errors be given by $\widehat{\vartheta}_{t,h,\ell} = \widehat{\beta}_{t,\ell} - \widehat{\beta}_{t|t-h,\ell}$ for $t = h + 1, \ldots, n$. These can then be sampled with replacement to give a bootstrap sample of $\beta_{n+h,\ell}$:

$$\widehat{\beta}_{n+h|n,\ell}^{(b)} = \widehat{\beta}_{n+h|n,\ell} + \widehat{\vartheta}_{*,h,\ell}^{(b)}, \qquad b = 1, \ldots, B,$$

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

Assuming the first L principal components approximate the data $\boldsymbol{Z}$ relatively well, the model residual should contribute nothing but random noise. Consequently, we can bootstrap the model fit errors in (2) by sampling with replacement from the model residual term $\{\widehat{\omega}_{1,x}, \ldots, \widehat{\omega}_{n,x}\}$ for a given age x.

Adding all two components of variability, we obtain B variants for $z_{n+h,x}$,

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

where $\widehat{\beta}_{n+h,\ell}^{(b)}$ denotes the forecast of the bootstrapped principal component scores. The construction of the bootstrap samples $\{\widehat{z}_{n+h|n,x}^{(1)}, \ldots, \widehat{z}_{n+h|n,x}^{(B)}\}$ is in the same spirit as the construction of the bootstrap samples for the LC method in (5) and for the HU method in Hyndman and Shang (2010).

With the bootstrapped $\widehat{z}_{n+h|n,x}^{(b)}$, we follow steps 6) and 7) in Section 3.1, in order to obtain the bootstrap forecasts of $d_{n+h,x}$. At the $100(1 - \gamma)\%$ nominal coverage probability, the pointwise prediction intervals are obtained by taking $\gamma/2$ and $1 - \gamma/2$ quantiles based on $\{\widehat{d}_{n+h|n,x}^{(1)}, \ldots, \widehat{d}_{n+h|n,x}^{(B)}\}$.

3.2.2 Existing bootstrap method

The construction of the prediction interval for CoDa has also been considered by Bergeron-Boucher et al. (2017). After applying the centred log-ratio transformation, Bergeron-Boucher et al. (2017) fit a nonparametric model to estimate the regression mean and obtain the residual matrix, from which bootstrap residual matrices can be obtained by randomly sampling with replacement from the original residual matrix. With the bootstrapped residuals, they then add them to the estimated regression mean to form the bootstrapped data samples. With each replication of the bootstrapped data samples, one could then re-apply singular value decomposition to obtain the first set of principal component and its associated scores. With the bootstrapped first set of principal component scores, one could then extrapolate them to the future using a univariate time-series forecasting method. By multiplying the bootstrapped

forecast of the principal component scores with the bootstrap principal component, bootstrap forecasts of life-table death counts can be obtained via back-transformation.

In Table 1, we compare the interval forecast accuracy, as measured by the mean interval score, between the two ways of constructing prediction intervals (as described in Section 3.2.1 and 3.2.2.

3.3   Other forecasting methods

3.3.1   Lee-Carter (LC) method

As a comparison, we revisit Lee and Carter's (1992) method. To stabilise the higher variance associated with mortality at advanced old ages, it is necessary to transform the raw data by taking the natural logarithm. For those missing death counts at advanced old ages, we use a simple linear interpolation method to approximate the missing values. We denote by $m_{t,x}$ the observed mortality rate at year t at age x calculated as the number of deaths in the calendar year t at age x, divided by the corresponding mid-year population aged x. The model structure is given by

$$\ln(m_{t,x}) = a_x + b_x \kappa_t + \epsilon_{t,x}, \tag{4}$$

where a_x denotes the age pattern of the log mortality rates averaged across years; b_x denotes the first principal component reflecting relative change in the log mortality rate at each age; κ_t is the first set of principal component scores by year t and measures the general level of the log mortality rates; and $\epsilon_{t,x}$ denotes the residual at year t and age x.

The LC model in (4) is over-parametrised in the sense that the model structure is invariant under the following transformations:

$$\{a_x, b_x, \kappa_t\} \mapsto \{a_x, b_x/c, c\kappa_t\},$$
$$\{a_x, b_x, \kappa_t\} \mapsto \{a_x - cb_x, b_x, \kappa_t + c\}.$$

To ensure the model's identifiability, Lee and Carter (1992) imposed two constraints, given as:

$$\sum_{t=1}^{n} \kappa_t = 0, \qquad \sum_{x=x_1}^{x_p} b_x = 1,$$

where n is the number of years, and p is the number of ages in the observed data set.

Also, the LC method adjusts κ_t by refitting to the total number of deaths. This adjustment gives more weight to high rates, thus roughly counterbalancing the effect of using a log transformation of the mortality rates. The adjusted k_t is then extrapolated using RW models. Lee and Carter (1992) used an RWD model, which can be expressed as:

$$\kappa_t = \kappa_{t-1} + d + e_t,$$

where d is known as the drift parameter and measures the average annual change in the series, and e_t is an uncorrelated error. It is notable that the RWD provides satisfactory results in many cases (see, e.g., Tulkapurkar et al. 2000, Lee and Miller 2001). From this forecast of the principal component scores, the forecast age-specific log mortality rates are obtained using the estimated age effects a_x and b_x in (4).

Note that the LC model can also be formulated within a Generalised Linear Model framework with a generalised error distribution (see, e.g., Tabeau 2001, Brouhns et al. 2002, Renshaw and Haberman 2003). In this setting, the LC model parameters can be estimated by maximum likelihood methods based on the choice of the error distribution. Thus, in line with traditional actuarial practice, this approach assumes that the age- and period-specific number of deaths are independent realisations from a Poisson distribution. In the case of Gaussian error, the estimation method based on either the singular value decomposition or maximum likelihood method leads to the same parameter estimates.

Two sources of uncertainty are considered in the LC model: errors in the parameter estimation of the LC model and forecast errors in the forecasted principal component scores. Using a univariate time series model, we can obtain multi-step-ahead forecasts for the principal component scores, $\{\widehat{\kappa}_1, \ldots, \widehat{\kappa}_n\}$. Let the h-step-ahead forecast errors be given by $v_{t,h} = \widehat{\kappa}_t - \widehat{\kappa}_{t|t-h}$ for $t = h+1, \ldots, n$. These can then be sampled with replacement to give a bootstrap sample of κ_{n+h}:

$$\widehat{\kappa}^{(b)}_{n+h|n} = \widehat{\kappa}_{n+h|n} + \widehat{v}^{(b)}_{*,h}, \qquad b = 1, \ldots, B,$$

where $\widehat{v}^{(b)}_{*,h}$ are sampled with replacement from $\{\widehat{v}_{t,h}\}$.

Assuming the first principal component approximates the data $\ln m$ relatively well, the model residual should contribute nothing but random noise. Consequently, we can bootstrap the model fit error in (4) by sampling with replacement from the model residual term $\{\widehat{e}_{1,x}, \ldots, \widehat{e}_{n,x}\}$ for a given age x.

Adding the two components of variability, we obtain B variants of $\ln m_{n+h,x}$:

$$\ln \widehat{m}^{(b)}_{n+h|n,x} = \widehat{\kappa}^{(b)}_{n+h|n} \widehat{b}_x + \widehat{e}^{(b)}_{n+h,x'} \tag{5}$$

where $\widehat{\kappa}^{(b)}_{n+h|n}$ denotes the forecast of the bootstrapped principal component scores.

Since the LC model forecasts age-specific central mortality rate, we convert forecast mortality rate to the probability of dying via a simple approximation. The formula is given as

$$\widehat{q}_{n+h|n,x} = 1 - \exp^{-\widehat{m}_{n+h|n,x}},$$

where $\widehat{q}_{n+h|n,x}$ denotes the probability of dying in age x and year $n+h$, and $\widehat{m}_{n+h|n,x}$ denotes the forecast of central mortality rate in age x and year $n+h$. With an initial life-table death count of 100,000, the forecast of $\widehat{d}_{n+h|n,x}$ can be obtained from $\widehat{q}_{n+h|n,x}$.

3.3.2 Hyndman-Ullah (HU) method

The HU method differs from the LC method in the following three aspects:

1. Instead of modelling the original mortality rates, the HU method uses a *P*-spline with a monotonic constraint to smooth log mortality rates.

2. Instead of using only one principal component and its associated scores, the HU method uses six sets of principal components and their scores.

3. Instead of forecasting each set of principal component scores by an RWD method, the HU method uses an automatic algorithm of Hyndman and Khandakar (2008) to select the optimal model and estimate the parameters in a univariate time series forecasting method.

3.3.3 Random walk with and without drift

Given the findings of linear life expectancy (White 2002, Oeppen and Vaupel 2002) and the debate about its continuation (Bengtsson 2003, Lee 2003), it is pertinent to compare the forecast accuracy of the CoDa method with a linear extrapolation method (see Alho and Spencer 2005, pp. 274-276 for an introduction). Using the centred log-ratio transformation, the linear extrapolation of $z_{t,x}$ in (2) is achieved by applying the RW and RWD models for each age:

$$z_{t+1,x} = \zeta + z_{t,x} + e_{t+1,x}, \tag{6}$$

where $z_{t,x}$ represents the centred log ratio transformed data at age $x = 0, 1, \ldots, p$ in year $t = 1, 2, \ldots, n - 1$, ζ denotes the drift term, and $e_{t+1,x}$ denotes model error term. The h-step-ahead point and interval forecasts are given by

$$\widehat{z}_{n+h|n,x} = E\left[z_{n+h,x} | z_{1,x}, \ldots, z_{n,x}\right] = \zeta h + z_{n,x}$$
$$\text{var}(\widehat{z}_{n+h|n,x}) = \text{var}[z_{n+h,x} | z_{1,x}, \ldots, z_{n,x}] = \text{var}(z_{n,x}) + \text{var}(e_{n+h,x}). \tag{7}$$

Based on (6) and (7), the RW can be obtained by omitting the drift term. Computationally, the forecasts of the RW and RWD are obtained by the `rwf` function in the *forecast* package (Hyndman et al. 2019) in R (R Core Team 2019). Then, by way of the back-transformation of the centred log ratio, naïve forecasts of age-specific life-table death counts can be obtained.

4 CoDa model fitting

For the life-table death counts, we examine the goodness-of-fit of the CoDa model to the observed data. The number of retained components L in (2) and (3) is determined by explaining at least 85% of the total variation. For the Australian female and male data, the chosen number

of components $L = 1$ and $L = 2$, respectively. We attempt to interpret the first component for the Australian female and male age-specific life-table death counts.

Based on the historical death counts from 1921 to 2014 (i.e., 94 observations), in Figure 2, we present the geometric mean of female and male life-table death counts, given by α_x, transformed data matrix Z, and the first set of estimated principal component obtained by applying the principal component analysis to Z. The shape of the first estimated principal component appears to be similar to the first estimated principal component in the Lee and Carter's (1992) and Hyndman and Ullah's (2007) methods. Because Z is unbounded, the principal component analysis captures a similar projection direction. By using the automatic ETS forecasting method, we produce the 20-steps-ahead forecast of principal component scores for the year between 2015 and 2034.

Apart from the graphical display, we measure the in-sample goodness-of-fit via an R^2 criterion given as

$$R^2 = 1 - \frac{\sum_{x=1}^{111} \sum_{t=1}^{94} \left(d_{t,x} - \widehat{d}_{t,x} \right)^2}{\sum_{x=1}^{111} \sum_{t=1}^{94} \left(d_{t,x} - \overline{d}_x \right)^2},$$

where $d_{t,x}$ denotes the observed age-specific life-table death count for age x in year t, and $\widehat{d}_{t,x}$ denotes the fitted age-specific life-table death count obtained from the CoDa model. For the Australian female and male data, the R^2 values are 0.9946 and 0.9899 using the CPV criterion while the R^2 values are 0.9987 and 0.9987 based on $L = 6$, respectively. Using the CPV criterion, we select $L = 1$ for both female and male data. One remark is that choosing the number of principal components $L = 6$ does not improve greatly the R^2 values, but it improves the forecast accuracy greatly as will be shown in Table 1.

5   Comparisons of point and interval forecast accuracy

5.1   Forecast error criteria

An expanding window analysis of a time-series model is commonly used to assess model and parameter stability over time, and prediction accuracy. The expanding window analysis determines the constancy of a model's parameter by computing parameter estimates and their resultant forecasts over an expanding window of a fixed size through the sample (For details, Zivot and Wang 2006, pp. 313-314). Using the first 74 observations from 1921 to 1994 in the Australian female and male age-specific life-table death counts, we produce one- to 20-step-ahead forecasts. Through a rolling-window approach, we re-estimate the parameters in the time series forecasting models using the first 75 observations from 1921 to 1995. Forecasts from the estimated models are then produced for one- to 19-step-ahead. We iterate this process
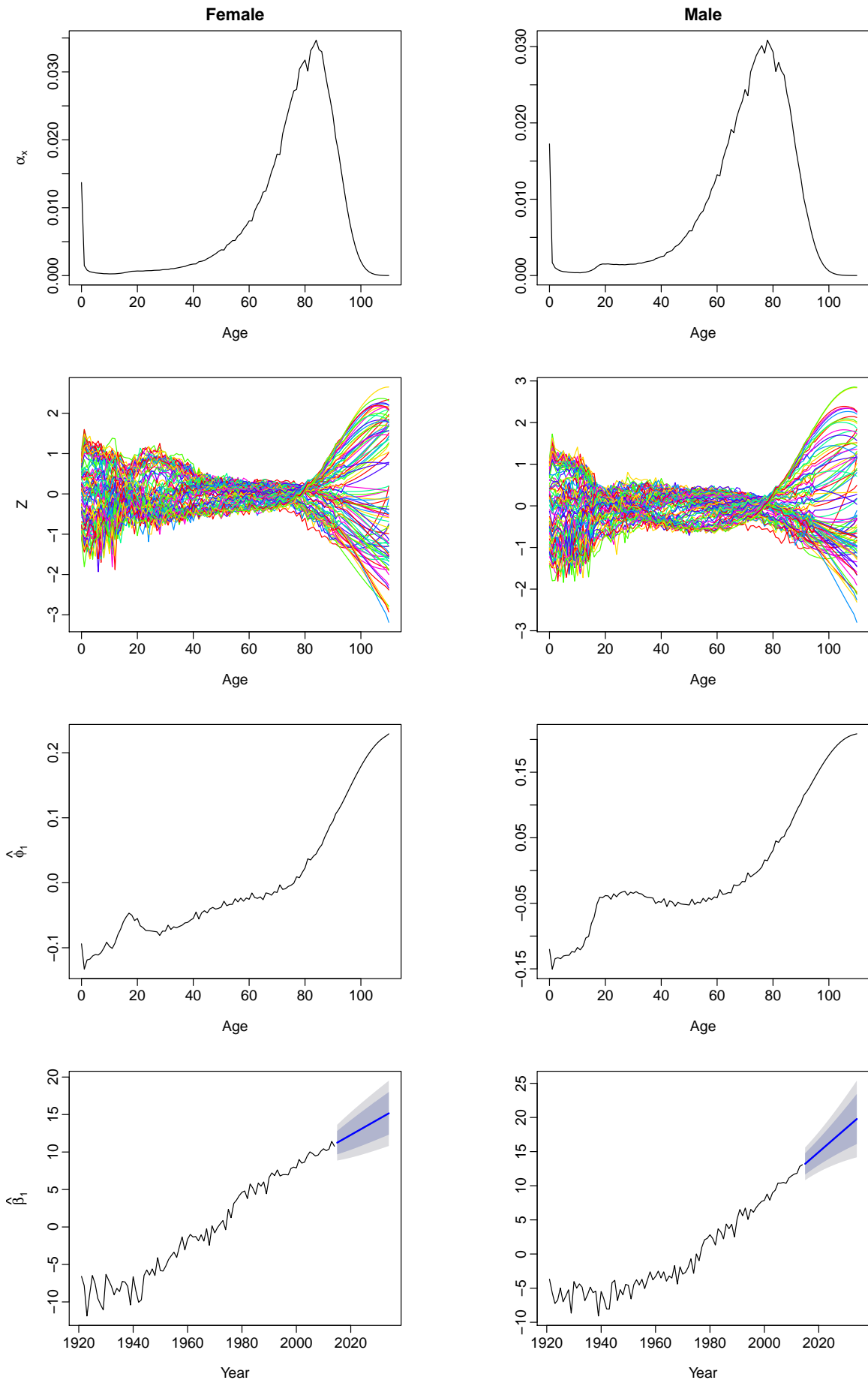
Figure 2: *Elements of the CoDa method for analysing the female and male age-specific life-table death counts in Australia. We present the first principal component and its scores, although we use L = 6 for fitting.*

by increasing the sample size by one year until reaching the end of the data period in 2014. This process produces 20 one-step-ahead forecasts, 19 two-step-ahead forecasts, ..., and one 20-step-ahead forecast. We compare these forecasts with the holdout samples to determine the out-of-sample point forecast accuracy.

To evaluate the point forecast accuracy, we consider the MAPE which measures how close the forecasts are in comparison to the actual values of the variable being forecast, regardless of the direction of forecast errors. The error measure can be written as

$$\text{MAPE}(h) = \frac{1}{111 \times (21-h)} \sum_{\varsigma=h}^{20} \sum_{x=1}^{111} \left| \frac{d_{n+\varsigma,x} - \widehat{d}_{n+\varsigma,x}}{d_{n+\varsigma,x}} \right| \times 100,$$

where $d_{n+\varsigma,x}$ denotes the actual holdout sample for the xth age and ς^{th} forecasting year, while $\widehat{d}_{n+\varsigma,x}$ denotes the point forecasts for the holdout sample.

To evaluate and compare the interval forecast accuracy, we consider the interval score of Gneiting and Raftery (2007). For each year in the forecasting period, the h-step-ahead prediction intervals are calculated at the $100(1-\gamma)\%$ nominal coverage probability. We consider the common case of the symmetric $100(1-\gamma)\%$ prediction intervals, with lower and upper bounds that are predictive quantiles at $\gamma/2$ and $1-\gamma/2$, denoted by $\widehat{d}^l_{n+\varsigma,x}$ and $\widehat{d}^u_{n+\varsigma,x}$. As defined by Gneiting and Raftery (2007), a scoring rule for the interval forecasts at time point $d_{n+\varsigma,x}$ is

$$S_{\gamma,\varsigma} \left[\widehat{d}^l_{n+\varsigma,x}, \widehat{d}^u_{n+\varsigma,x}, d_{n+\varsigma,x} \right] = \left(\widehat{d}^u_{n+\varsigma,x} - \widehat{d}^l_{n+\varsigma,x} \right) + \frac{2}{\gamma} \left(\widehat{d}^l_{n+\varsigma,x} - d_{n+\varsigma,x} \right) \mathbb{1} \left\{ d_{n+\varsigma,x} < \widehat{d}^l_{n+\varsigma,x} \right\}$$
$$+ \frac{2}{\gamma} \left(d_{n+\varsigma,x} - \widehat{d}^u_{n+\varsigma,x} \right) \mathbb{1} \left\{ d_{n+\varsigma,x} > \widehat{d}^u_{n+\varsigma,x} \right\},$$

where $\mathbb{1}\{\cdot\}$ represents the binary indicator function, and γ denotes the level of significance, customarily $\gamma = 0.2$ or $\gamma = 0.05$. The interval score rewards a narrow prediction interval, if and only if the true observation lies within the prediction interval. The optimal interval score is achieved when $d_{n+\varsigma,x}$ lies between $\widehat{d}^l_{n+\varsigma,x}$ and $\widehat{d}^u_{n+\varsigma,x}$, and the distance between $\widehat{d}^l_{n+\varsigma,x}$ and $\widehat{d}^u_{n+\varsigma,x}$ is minimal.

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

$$\bar{S}_\gamma(h) = \frac{1}{111 \times (21-h)} \sum_{\varsigma=h}^{20} \sum_{x=1}^{111} S_{\gamma,\varsigma} \left[\widehat{d}^l_{n+\varsigma,x}, \widehat{d}^u_{n+\varsigma,x}; d_{n+\varsigma,x} \right],$$

where $S_{\gamma,\varsigma} \left[\widehat{d}^l_{n+\varsigma,x}, \widehat{d}^u_{n+\varsigma,x}; d_{n+\varsigma,x} \right]$ denotes the interval score at the ςth curve in the forecasting year.

5.2   Forecast results

Using the expanding window approach, we compare the point and interval forecast accuracies among the CoDa, HU, LC and two naïve RW methods based on the MAPE and mean interval

score. Also, we consider forecasting each set of the estimated principal component scores by the ARIMA, ETS, RWD and RW. The overall point and interval forecast error results are presented in Table 1, where we average over the 20 forecast horizons. From Table 1, we find that the CoDa method with the ETS forecasting method generally performs the best among a range of methods considered. Among the four univariate time series forecasting methods, the ETS method produces the smallest errors and thus is recommended to be used in practice. Also, setting $L = 6$, including additional principal component decomposition pairs, generally provides more accurate point and interval forecast accuracies than setting $L = 1$ or $L = 2$ for the Australian female and male data.

Table 1: *A comparison of the point and interval forecast accuracy, as measured by the overall MAPE and mean interval score, among the CoDa, HU, LC and two naïve RW methods (in italic) using the holdout sample of the Australian female and male data. Further, we consider four univariate time series forecasting methods for the CoDa and HU methods. In terms of interval forecast accuracy, we compare the finite-sample performance between the proposed bootstrap approach and an existing bootstrap approach of Bergeron-Boucher et al. (2017). The smallest errors are highlighted in bold.*

Series	L	Method	Forecasting method		Criteria			
					Proposed bootstrap		Existing bootstrap	
				MAPE	$\overline{S}_{\gamma=0.2}$	$\overline{S}_{\gamma=0.05}$	$\overline{S}_{\gamma=0.2}$	$\overline{S}_{\gamma=0.05}$
Female	CPV	CoDa	ARIMA	20.51	423.69	653.17	575.89	670.54
			ETS	20.65	492.77	870.10	529.32	655.10
			RWD	21.32	476.77	782.52	711.58	859.78
			RW	32.18	512.30	629.45	1302.91	2368.32
	$L = 6$	CoDa	ARIMA	17.02	298.93	431.48	578.12	812.75
			ETS	**14.60**	**232.10**	**369.76**	277.51	407.85
			RWD	16.46	369.77	640.87	427.40	689.46
			RW	29.25	985.94	2512.42	1056.94	2422.63
		HU	ARIMA	19.60	348.84	462.21		
			ETS	16.71	262.36	376.59		
			RWD	16.49	349.00	472.72		
			RW	30.95	350.32	442.68		
		LC		26.54	516.79	667.54		
		RWD		16.59	703.09	1185.52		

Series	L	Method	Forecasting method	MAPE	Proposed bootstrap $\overline{S}_{\gamma=0.2}$	$\overline{S}_{\gamma=0.05}$	Existing bootstrap $\overline{S}_{\gamma=0.2}$	$\overline{S}_{\gamma=0.05}$
		RW		30.04	703.79	1186.35		
Male	CPV	CoDa	ARIMA	31.93	1122.74	2079.19	1624.23	3889.32
			ETS	25.48	644.38	752.16	1139.60	2184.71
			RWD	32.16	690.67	896.75	1750.66	4434.10
			RW	44.82	1428.70	3142.54	2115.80	5683.19
	$L=6$	CoDa	ARIMA	27.18	648.85	860.78	1484.00	4067.55
			ETS	**18.37**	**371.22**	**516.23**	699.92	1417.62
			RWD	23.59	776.14	1221.44	1124.12	2854.69
			RW	37.48	457.71	576.84	1638.70	4721.93
		HU	ARIMA	27.22	641.30	944.56		
			ETS	24.64	726.86	1281.29		
			RWD	25.14	806.16	1703.06		
			RW	39.48	965.43	2288.80		
		LC		38.61	1273.36	2692.18		
		RWD		23.60	783.15	1166.38		
		RW		38.18	783.30	1166.34		

The CoDa method tends to provide less bias forecasts than the Lee-Carter method as it allows the rate of mortality improvements to change over time (see Bergeron-Boucher et al. 2017). With respect to the positivity and summability constraints, the CoDa method can adopt temporal changes of the age distribution of the life-table death counts over the years. The unsatisfactory performance of the HU and LC methods may because of the approximation of converting the forecasts of central mortality rate to the probability of dying at higher ages. The inferior performance of the two naïve RW methods may because of their slow responses to adapt to the change of age distribution of death counts. Between the two naïve RW methods, we found that the RW is preferable for producing point forecasts, while the RWD is preferable for producing interval forecasts.

While Table 1 presents the average over 20 forecast horizons, we show the one-step-ahead

to 20-step-ahead point and interval forecast errors in Figure 3. Since it is advantageous to set $L = 6$, we report the CoDa method with the ETS forecasting method and $L = 6$ in Figure 3. The difference in forecast accuracy between the CoDa and the other methods is widening over the forecast horizon. We suspect that in the relatively longer forecast horizon, the errors associated with all the methods become larger, but they are relatively smaller for the CoDa method with the ETS forecasting method and $L = 6$.

6   Application to a single-premium temporary immediate annuity

An important use of mortality forecasts for those individuals at ages over 60 is in the pension and insurance industries, whose profitability and solvency crucially rely on accurate mortality forecasts to appropriately hedge longevity risks. When a person retires, an optimal way of guaranteeing one individual's financial income in retirement is to purchase an annuity (as demonstrated by Yaari 1965). An annuity is a financial contract offered by insurers guaranteeing a steady stream of payments for either a temporary or the lifetime of the annuitants in exchange for an initial premium fee.

Following Shang and Haberman (2017), we consider temporary annuities, which have grown in popularity in a number of countries (e.g., Australia and United States of America), because lifetime immediate annuities, where rates are locked in for life, have been shown to deliver poor value for money (i.e., they may be expensive for the purchasers; see for example Cannon and Tonks 2008, Chapter 6). These temporary annuities pay a pre-determined and guaranteed level of income which is higher than the level of income provided by a lifetime annuity for a similar premium. Fixed-term annuities offer an alternative to lifetime annuities and allow the purchaser the option of also buying a deferred annuity at a later date.

Using the CoDa method, we obtain forecasts of age-specific life-table death counts and then determine the corresponding survival probabilities. In Figure 4, we present the age-specific life-table death count forecasts from 2015 to 2064 for Australian females.

With the mortality forecasts, we then input the forecasts of death counts to the calculation of single-premium term immediate annuities (see Dickson et al. 2009, p. 114), and we adopt a cohort approach to the calculation of the survival probabilities. The τ year survival probability of a person aged x currently at $t = 0$ (or year 2016) is determined by

$$\tau p_x = \prod_{j=1}^{\tau} p_{x+j-1} = \prod_{j=1}^{\tau} \left(1 - q_{x+j-1}\right) = \prod_{j=1}^{\tau} \left(1 - \frac{d_{x+j-1}}{l_{x+j-1}}\right),$$

where d_{x+j-1} denotes the number of death counts between ages $x + j - 1$ and $x + j$; and l_{x+j-1} denotes the number of lives alive at age $x + j - 1$. Note that τp_x is a random variable given that death counts for $j = 1, \ldots, \tau$ are the forecasts obtained by the CoDa method.
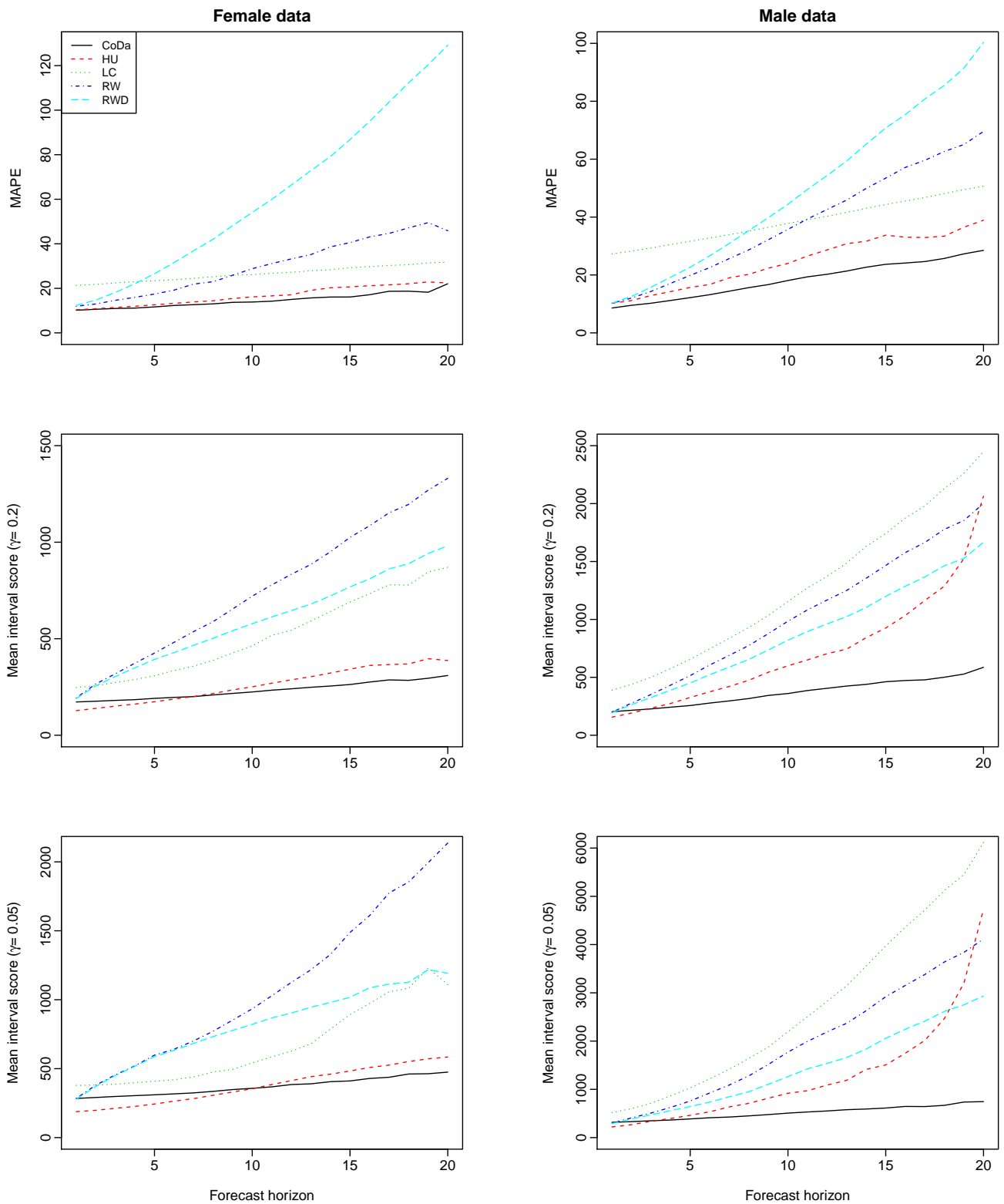
Figure 3: *A comparison of the point and interval forecast accuracy, as measured by the MAPE and mean interval score, among the CoDa, HU, LC and two naïve RW methods using the holdout sample of the Australian female and male data. In the CoDa method, we use the ETS forecasting method with the number of principal components L = 6.*
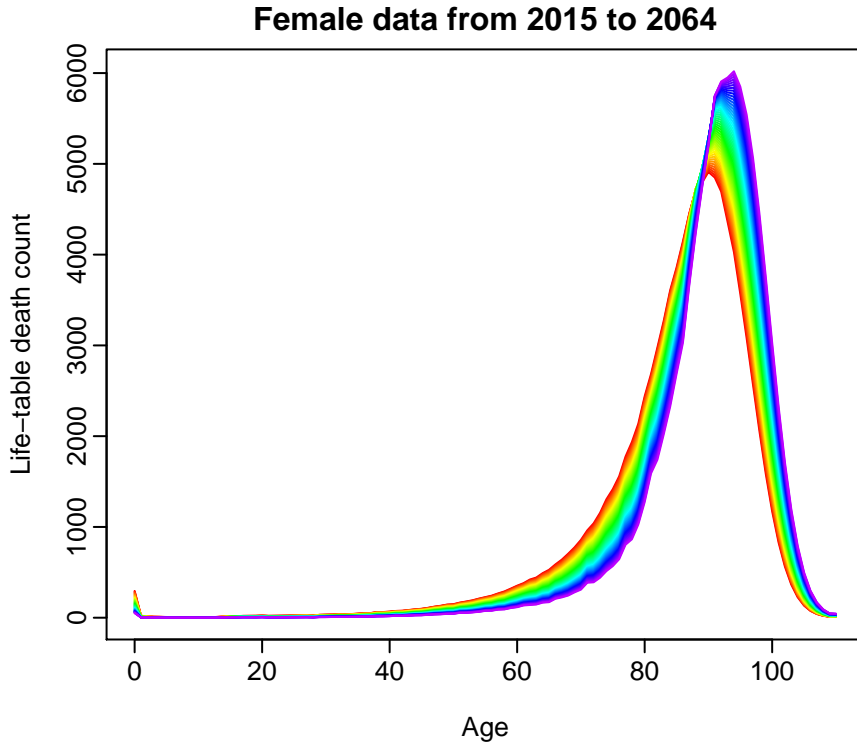
Figure 4: *Age-specific life-table death count forecasts from 2015 to 2064 for Australian females.*

The price of an annuity with a maturity term of a T year is a random variable, as it depends on the value of zero-coupon bond price and future mortality. The annuity price can be written for an x-year-old with benefit one Australian dollar per year is given by

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

where $B(0, \tau)$ is the τ-year bond price and $_\tau p_x$ denotes the survival probability.

In Table 2, to provide an example of the annuity calculations, we compute the best estimate of the annuity prices for different ages and maturities for a female policyholder residing in Australia. We assume a constant interest rate at $\eta = 3\%$ and hence zero-coupon bond is given as $B(0, \tau) = \exp^{-\eta\tau}$. Although the difference in annuity price might appear to be small, any mispricing can involve a significant risk when considering a large annuity portfolio. Given that an annuity portfolio consists of N policies where the benefit per year is B, any underpricing of $\tau\%$ of the actual annuity price will result in a shortfall of $NBa_x^T\tau/100$, where a_x^T is the estimated annuity price being charged with benefit one Australian dollar per year. For example $\tau = 0.1\%$, $N = 10,000$ policies written to 80-year-old female policyholders with maturity $\tau = 20$ years and benefit $20,000$ Australian dollars per year will result in a shortfall of $10,000 \times 20,000 \times 7.8109 \times 0.1\% = 1.5622$ million.

To measure forecast uncertainty, we construct the bootstrapped age-specific life-table death counts, derive the survival probabilities and calculate the corresponding annuities associated with different ages and maturities. Given that we consider ages from 60 to 110, we construct

Table 2: *Estimates of annuity prices with different ages and maturities (T) for a female policyholder residing in Australia. These estimates are based on forecast mortality rates from 2015 to 2064. We consider only contracts with maturity so that age + maturity ≤ 110. If age + maturity > 110, NA will be shown in the table.*

Age	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
CPV						
60	4.5109	8.2748	11.3625	13.8221	15.6564	16.8572
65	4.4831	8.1608	11.0904	13.2752	14.7054	15.4242
70	4.4314	7.9614	10.5940	12.3173	13.1833	13.4609
75	4.3445	7.5847	9.7056	10.7715	11.1132	11.1684
80	4.1440	6.8566	8.2198	8.6568	8.7274	8.7322
85	3.7975	5.7060	6.3177	6.4166	6.4232	NA
90	3.1874	4.2090	4.3742	4.3852	NA	NA
95	2.4261	2.8183	2.8445	NA	NA	NA
100	1.6751	1.7870	NA	NA	NA	NA
105	1.1283	NA	NA	NA	NA	NA
$L = 6$						
60	4.5197	8.3091	11.4419	13.9673	15.8921	17.1909
65	4.4972	8.2151	11.2122	13.4965	15.0380	15.8363
70	4.4538	8.0441	10.7806	12.6271	13.5835	13.8961
75	4.3788	7.7163	9.9683	11.1348	11.5160	11.5764
80	4.2042	7.0411	8.5105	8.9907	9.0668	9.0715
85	3.8680	5.8714	6.5261	6.6299	6.6364	NA
90	3.2348	4.2919	4.4595	4.4699	NA	NA
95	2.4406	2.8276	2.8515	NA	NA	NA
100	1.6466	1.7486	NA	NA	NA	NA
105	1.0859	NA	NA	NA	NA	NA

50-steps-ahead bootstrap forecasts of age-specific life-table death counts. In Table 3, we present the 95% pointwise prediction intervals of annuities for different ages and maturities, where age + maturity $\leq$ 110.

Table 3: *95% pointwise prediction intervals of annuity prices with different ages and maturities (T) for a female policyholder residing in Australia. These estimates are based on forecast mortality rates from 2015 to 2064. We consider only contracts with maturity so that age + maturity $\leq$ 110. If age + maturity > 110, NA will be shown in the table.*

Age	T = 5	T = 10	T = 15	T = 20	T = 25	T = 30
CPV						
60	(4.497, 4.571)	(8.270, 8.495)	(11.363, 11.848)	(13.828, 14.678)	(15.722, 17.012)	(16.977, 18.800)
65	(4.454, 4.569)	(8.115, 8.486)	(11.060, 11.813)	(13.276, 14.542)	(14.746, 16.650)	(15.498, 18.021)
70	(4.372, 4.563)	(7.865, 8.455)	(10.488, 11.669)	(12.234, 14.154)	(13.114, 15.736)	(13.423, 16.475)
75	(4.218, 4.552)	(7.400, 8.359)	(9.541, 11.328)	(10.591, 13.265)	(10.975, 14.076)	(11.041, 14.274)
80	(3.978, 4.500)	(6.601, 8.050)	(8.002, 10.362)	(8.474, 11.375)	(8.546, 11.579)	(8.551, 11.598)
85	(3.521, 4.354)	(5.295, 7.252)	(5.894, 8.569)	(5.992, 8.856)	(6.004, 8.881)	NA
90	(2.888, 3.919)	(3.888, 5.719)	(4.047, 6.128)	(4.058, 6.162)	NA	NA
95	(2.204, 3.267)	(2.582, 4.022)	(2.610, 4.082)	NA	NA	NA
100	(1.491, 2.258)	(1.598, 2.460)	NA	NA	NA	NA
105	(0.898, 1.561)	NA	NA	NA	NA	NA
L = 6						
60	(4.499, 4.572)	(8.286, 8.501)	(11.455, 11.864)	(14.033, 14.712)	(16.057, 17.097)	(17.460, 18.992)
65	(4.467, 4.571)	(8.174, 8.495)	(11.221, 11.839)	(13.547, 14.627)	(15.206, 16.825)	(16.164, 18.379)
70	(4.397, 4.569)	(7.983, 8.477)	(10.763, 11.768)	(12.699, 14.411)	(13.809, 16.234)	(14.281, 17.154)
75	(4.293, 4.563)	(7.596, 8.427)	(9.946, 11.544)	(11.258, 13.729)	(11.775, 14.881)	(11.892, 15.186)
80	(4.088, 4.537)	(6.953, 8.250)	(8.616, 10.851)	(9.271, 12.240)	(9.380, 12.659)	(9.389, 12.704)
85	(3.736, 4.453)	(5.873, 7.673)	(6.712, 9.376)	(6.897, 9.928)	(6.913, 9.971)	NA
90	(3.112, 4.150)	(4.322, 6.396)	(4.571, 7.038)	(4.592, 7.113)	NA	NA
95	(2.370, 3.523)	(2.901, 4.571)	(2.942, 4.694)	NA	NA	NA
100	(1.546, 2.507)	(1.682, 2.821)	NA	NA	NA	NA
105	(0.964, 1.617)	NA	NA	NA	NA	NA

The forecast uncertainties become larger as maturities increase from $T = 5$ to $T = 30$ for a given age. The forecast uncertainties also increase as the initial ages when entering contracts increase from 60 to 105 for a given maturity.

7 Conclusion

We proposed an adaptation of the Hyndman and Ullah's (2007) method to a CoDa framework. Using the Australian age-specific life-table death counts, we evaluate and compare the point and interval forecast accuracies among the CoDa, HU, LC and two naïve RW methods for forecasting the age distribution of death counts. Based on the MAPE and mean interval score, the CoDa method with the ETS forecasting method and $L = 6$ is recommended, as it outperforms the HU method, LC method and two naïve RW and RWD for forecasting the age distribution of death counts. The superiority of the CoDa method is driven by the use of singular value decomposition to model the age distribution of the transformed death counts and the summability *constraint* of the age distribution of death counts.

We apply the CoDa method to forecast age-specific life-table death counts from 2015 to 2064. We then calculate the cumulative survival probability and obtain temporary annuity prices. As expected, we find that the cumulative survival probability has a pronounced impact on annuity prices. Although annuity prices for an individual contract may be small, mispricing could have a dramatic effect on a portfolio, mainly when the yearly benefit is a great deal larger than one Australia dollar per year.

There are a few ways in which this paper could be extended, and we briefly discuss four. First, a robust CoDa method proposed by Filzmoser et al. (2009) may be utilised, in the presence of outlying years. Second, the methodology can be applied to calculate other types of annuity prices, such as the whole-life immediate annuity or deferred annuity. Thirdly, we can consider cohort life-table death counts for modelling a group of individuals. Finally, we may consider other density forecasting methods as in Kokoszka et al. (2019), such as the log quantile density transformation method.

Acknowledgements

References

Aburto, J. M. and van Raalte, A. A. (2018), 'LifeLife dispersion in times of life expectancy fluctuation: The case of central and eastern Europe', *Demography* **55**, 2071–2096.

Aitchison, J. (1982), 'The statistical analysis of compositional data', *Journal of the Royal Statistical Society: Series B* **44**(2), 139–177.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, Chapman & Hall, London.

Aitchison, J. and Shen, S. M. (1980), 'Logistic-normal distributions: Some properties and uses', *Biometrika* **67**(2), 261–272.

Alho, J. M. and Spencer, B. D. (2005), *Statistical Demography and Forecasting*, Springer, New York.

Bengtsson, T. (2003), The need for looking far back in time when predicting future mortality trends, *in* T. Bengtsson and N. Keilman, eds, 'Perspectives on Mortality Forecasting', Swedish National Social Insurance Agency, Stockholm.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J. and Vaupel, J. W. (2017), 'Coherent forecasts of mortality with compositional data analysis', *Demographic Research* **37**, 527–566.

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

Booth, H. and Tickle, L. (2008), 'Mortality modelling and forecasting: A review of methods', *Annals of Actuarial Science* **3**(1-2), 3–43.

Brouhns, N., Denuit, M. and Vermunt, J. (2002), 'A Poisson log-bilinear regression approach to the construction of projected life tables', *Insurance: Mathematics and Economics* **31**(3), 373–393.

Cairns, A. J. G., Blake, D. and Dowd, K. (2008), 'Modelling and management of mortality risk: A review', *Scandinavian Actuarial Journal* **2008**(2-3), 79–113.

Cannon, E. and Tonks, I. (2008), *Annuity Markets*, Oxford University Press, Oxford.

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

Chatfield, C. (1993), 'Calculating interval forecasts', *Journal of Business & Economics Statistics* **11**(2), 121–135.

Chatfield, C. (2000), *Time-Series Forecasting*, Chapman & Hall, Boca Raton, Florida.

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.

Cramér, H. and Wold, H. (1935), 'Mortality variations in Sweden', *Scandinavian Actuarial Journal* **1935**(3-4), 161–241.

Debón, A., Chaves, L., Haberman, S. and Villa, F. (2017), 'Characterization of between-group inequality of longevity in EU countries', *Insurance: Mathematics and Economics* **75**, 151–165.

Delicado, P. (2011), 'Dimensionality reduction when data are density functions', *Computational Statistics and Data Analysis* **55**(1), 401–420.

Denuit, M., Devolder, P. and Goderniaux, A.-C. (2007), 'Securitization of longevity risk: Pricing survivor bonds with Wang transform in the Lee-Carter framework', *The Journal of Risk and Insurance* **74**(1), 87–113.

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

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueraz, G. and Barceló-Vidal, C. (2003), 'Isometric logratio transformations for compositional data analysis', *Mathematical Geology* **35**(3), 279–300.

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

Gneiting, T. and Katzfuss, M. (2014), 'Probabilistic forecasting', *Annual Review of Statistics and Its Application* **1**, 125–151.

Gneiting, T. and Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction and estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Horváth, L. and Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer, New York.

Human Mortality Database (2019), *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. Accessed at May 8, 2017. URL: http://www.mortality.org.

Hurvich, C. M. and Tsai, C.-L. (1993), 'A corrected Akaike information criterion for vector autoregressive model selection', *Journal of Time Series Analysis* **14**(3), 271–279.

Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., Team, R. C., Ihaka, R., Reid, D., Shaub, D., Tang, Y. and Zhou, Z. (2019), *forecast: Forecasting Functions for Time Series and Linear Models*. R package version 8.7.

URL: https://CRAN.R-project.org/package=forecast

Hyndman, R. J. and Booth, H. (2008), 'Stochastic population forecasts using functional data models for mortality, fertility and migration', *International Journal of Forecasting* **24**, 323–342.

Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).

Hyndman, R. J. and Shang, H. L. (2010), 'Rainbow plots, bagplots, and boxplots for functional data', *Journal of Computational and Graphical Statistics* **19**(1), 29–45.

Hyndman, R. J. and Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics and Data Analysis* **51**(10), 4942–4956.

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* **revised and resubmitted**.

Lee, R. D. (1998), 'Probabilistic approach to population forecasting', *Population and Development Review* **24**(419), 156–190.

Lee, R. D. (2003), Mortality forecasts and linear life expectancy trends, Working paper 03-25, University of California, Berkeley.

URL: https://escholarship.org/uc/item/3sd9m7d5

Lee, R. D. and Carter, L. R. (1992), 'Modeling and forecasting U.S. mortality', *Journal of the American Statistical Association* **87**(419), 659–671.

Lee, R. D. and Miller, T. (2001), 'Evaluating the performance of the Lee-Carter method for forecasting mortality', *Demography* **38**(4), 537–549.

Oeppen, J. (2008), Coherent forecasting of multiple-decrement life tables: A test using Japanese cause of death data, *in* 'European Population Conference', Barcelona, Spain.

URL: http://epc2008.princeton.edu/papers/80611

Oeppen, J. and Vaupel, J. W. (2002), 'Broken limits to life expectancy', *Demography* **296**(5570), 1029–1031.

Paparoditis, E. (2018), 'Sieve bootstrap for functional time series', *The Annals of Statistics* **46**(6B), 3510–3538.

Pollard, J. H. (1987), 'Projection of age-specific mortality rates', *Population Bulletin of the United Nations* **21-22**, 55–69.

Preston, S., Heuveline, P. and Guillot, M. (2001), *Demography: Measuring and Modeling Population Processes*, Blackwell Publishers, Oxford, U.K.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
   URL: https://www.R-project.org/

Renshaw, A. E. and Haberman, S. (2003), 'Lee-Carter mortality forecasting with age-specific enhancement', *Insurance: Mathematics and Economics* **33**(2), 255–272.

Robine, J.-M. (2001), 'Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: The case of France', *Population: An English Selection* **13**(1), 173–193.

Salish, N. and Gleim, A. (2015), Forecasting methods for functional time series, Technical report, University of Cologne.
   URL: http://www.eco.uc3m.es/temp/paper%20Salish.pdf

Scealy, J. L., de Caritat, P., Grunsky, E. C., Tsagris, M. T. and Welsh, A. H. (2015), 'Robust principal component analysis for power transformed compositional data', *Journal of the American Statistical Association* **110**(509), 136–148.

Scealy, J. L. and Welsh, A. H. (2017), 'A directional mixed effects model for compositional expenditure data', *Journal of the American Statistical Association* **112**(517), 24–36.

Shang, H. L. (2018), 'Bootstrap methods for stationary functional time series', *Statistics and Computing* **28**(1), 1–10.

Shang, H. L., Booth, H. and Hyndman, R. J. (2011), 'Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods', *Demographic Research* **25**(5), 173–214.

Shang, H. L. and Haberman, S. (2017), 'Grouped multivariate and functional time series forecasting: An application to annuity pricing', *Insurance Mathematics & Economics* **75**, 166–179.

Tabeau, E. (2001), A review of demographic forecasting models for mortality, *in* E. Tabeau, A. Van Den Berg Jeths and C. Heathcote, eds, 'Forecasting Mortality in Developed Countries:

Insights from a Statistical, Demographic and Epidemiological Perspective', Kluwer Academic Publishers, Dordrecht, pp. 1–32.

Tulkapurkar, S., Li, N. and Boe, C. (2000), 'A universal pattern of mortality decline in the G7 countries', *Nature* **405**(6788), 789–792.

van Raalte, A. A. and Caswell, H. (2013), 'Perturbation analysis of indices of lifespan variability', *Demography* **50**(5), 1615–1640.

van Raalte, A. A., Martikainen, P. and Myrskylä, M. (2014), 'Lifespan variation by occupational class: Compression or stagnation over time?', *Demography* **51**, 73–95.

Vaupel, J. W., Zhang, Z. and van Raalte, A. A. (2011), 'Life expectancy and disparity: An international comparison of life table data', *BMJ Open* **1**(1), e000128.

White, K. M. (2002), 'Longevity advances in high-income countries, 1955-96', *Population and Development Review* **28**(1), 59–76.

Wilmoth, J. R. (1995), 'Are mortality projections always more pessimistic when disaggregated by cause of death?', *Mathematical Population Studies* **5**(4), 293–319.

Wilmoth, J. R. and Horiuchi, S. (1999), 'Rectangularization revisited: Variability of age at death within human populations', *Demography* **36**(4), 475–495.

Yaari, M. E. (1965), 'Uncertain lifetime, life insurance, and the theory of the consumer', *The Review of Economic Studies* **32**(2), 137–150.

Zivot, E. and Wang, J. (2006), *Modeling Financial Time Series with S-PLUS*, Springer, New York.

Appendix: supplementary material

Annuity price calculation for the Australian male data In Section 6, we present the annuity price calculation for the Australian female data. Since the mortality rates are different between male and female, the estimate of annuity prices ought to be different. Here, we present the annuity price calculation for the Australian male data.

R code and data files The R code for implementing the compositional data analysis approach and for producing point and interval forecasts is provided, along with the Australian age- and sex-specific life-table death counts.

<div align="center">

Supplementary materials

"Forecasting age distribution of death counts: An application to annuity pricing"

by Han Lin Shang and Steven Haberman

</div>

S1. Annuity price calculation for the Australian male data

In Table A, to provide an example of the annuity calculations, we compute the best estimate of the annuity prices for different ages and maturities for a male policyholder residing in Australia. We assume a constant interest rate $\eta = 3\%$ and hence zero-coupon bond is given as $B(0, \tau) = \exp^{-\eta\tau}$.

Table A: *Estimates of annuity prices with different ages and maturities (T) for a male policyholder residing in Australia. These estimates are based on forecast mortality rates from 2015 to 2064. We consider only contracts with maturity so that age + maturity $\leq$ 110. If age + maturity $>$ 110, NA will be shown in the table.*

Age	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
CPV						
60	4.4821	8.1694	11.1358	13.4387	15.0911	16.1213
65	4.4417	8.0150	10.7890	12.7794	14.0204	14.5930
70	4.3779	7.7766	10.2151	11.7356	12.4372	12.6353
75	4.2734	7.3396	9.2514	10.1336	10.3827	10.4173
80	4.0425	6.5631	7.7262	8.0546	8.1001	8.1027
85	3.6839	5.3838	5.8637	5.9303	5.9341	NA
90	3.0140	3.8651	3.9832	3.9899	NA	NA
95	2.2655	2.5799	2.5978	NA	NA	NA
100	1.5593	1.6481	NA	NA	NA	NA
105	1.0617	NA	NA	NA	NA	NA
$K = 6$						
60	4.4851	8.1866	11.1898	13.5540	15.2963	16.4257
65	4.4520	8.0641	10.9077	13.0033	14.3617	15.0298
70	4.4041	7.8712	10.4262	12.0825	12.8970	13.1469
75	4.3124	7.4903	9.5503	10.5634	10.8742	10.9212
80	4.1171	6.7858	8.0983	8.5010	8.5618	8.5655
85	3.7732	5.6287	6.1980	6.2840	6.2893	NA
90	3.1297	4.0899	4.2350	4.2438	NA	NA
95	2.3655	2.7229	2.7448	NA	NA	NA
100	1.6179	1.7171	NA	NA	NA	NA
105	1.0895	NA	NA	NA	NA	NA

To measure forecast uncertainty, we construct the bootstrapped age-specific life-table death counts, derive the survival probabilities and calculate the corresponding annuities associated with different ages and maturities. Given that we consider ages from 60 to 110, we construct 50-steps-ahead bootstrap forecasts of age-specific life-table death counts. In Table B, we present the 95% pointwise prediction intervals of annuities for different ages and maturities, where age + maturity $\leq$ 110.

Table B: *95% pointwise prediction intervals of annuity prices with different ages and maturities (T) for a male policyholder residing in Australia. These estimates are based on forecast mortality rates from 2015 to 2064. We consider only contracts with maturity so that age + maturity $\leq$ 110. If age + maturity > 110, NA will be shown in the table.*

Age	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
CPV						
60	(4.540, 4.572)	(8.384, 8.500)	(11.582, 11.867)	(14.188, 14.739)	(16.146, 17.136)	(17.437, 19.004)
65	(4.510, 4.570)	(8.263, 8.487)	(11.319, 11.828)	(13.624, 14.612)	(15.147, 16.786)	(15.866, 18.174)
70	(4.461, 4.563)	(8.083, 8.453)	(10.814, 11.698)	(12.610, 14.239)	(13.460, 15.845)	(13.711, 16.480)
75	(4.367, 4.547)	(7.656, 8.341)	(9.809, 11.306)	(10.845, 13.191)	(11.137, 13.959)	(11.179, 14.106)
80	(4.137, 4.480)	(6.825, 7.989)	(8.109, 10.208)	(8.457, 11.104)	(8.502, 11.278)	(8.506, 11.290)
85	(3.727, 4.301)	(5.500, 7.047)	(6.015, 8.135)	(6.088, 8.332)	(6.092, 8.347)	NA
90	(2.959, 3.761)	(3.806, 5.258)	(3.930, 5.557)	(3.938, 5.575)	NA	NA
95	(2.157, 2.980)	(2.451, 3.523)	(2.465, 3.568)	NA	NA	NA
100	(1.460, 2.000)	(1.538, 2.137)	NA	NA	NA	NA
105	(0.811, 1.434)	NA	NA	NA	NA	NA
$K = 6$						
60	(4.542, 4.573)	(8.396, 8.505)	(11.631, 11.882)	(14.302, 14.773)	(16.375, 17.216)	(17.828, 19.201)
65	(4.520, 4.572)	(8.319, 8.498)	(11.437, 11.858)	(13.862, 14.700)	(15.560, 17.008)	(16.512, 18.656)
70	(4.477, 4.568)	(8.166, 8.479)	(11.039, 11.785)	(13.042, 14.475)	(14.159, 16.395)	(14.592, 17.442)
75	(4.406, 4.560)	(7.835, 8.416)	(10.235, 11.547)	(11.567, 13.793)	(12.075, 15.025)	(12.198, 15.418)
80	(4.230, 4.518)	(7.192, 8.201)	(8.848, 10.857)	(9.446, 12.278)	(9.564, 12.697)	(9.573, 12.784)
85	(3.923, 4.426)	(6.083, 7.594)	(6.893, 9.353)	(7.082, 9.900)	(7.091, 9.948)	NA
90	(3.275, 4.077)	(4.460, 6.292)	(4.689, 6.994)	(4.713, 7.097)	NA	NA
95	(2.423, 3.494)	(2.907, 4.582)	(2.955, 4.740)	NA	NA	NA
100	(1.611, 2.651)	(1.739, 3.006)	NA	NA	NA	NA
105	(1.055, 1.840)	NA	NA	NA	NA	NA