A Three-component Approach to Model and Forecast Age-at-death Distributions

Ugofilippo Basellini[*,1,2] and Carlo Giovanni Camarda[1]

[1]*Institut national d'études démographiques (INED), Paris*
[2]*Interdisciplinary Centre on Population Dynamics (CPOP), Department of Public Health, University of Southern Denmark, Odense*

June 7, 2019

Abstract

Mortality forecasting has recently received growing interest, as accurate projections of future lifespans are needed to ensure the solvency of insurance and pension providers. Several innovative stochastic methodologies have been proposed in most recent decades, the majority of them being based on age-specific mortality rates or on summary measures of the life table. The age-at-death distribution is an informative life-table function that provides readily available information on the mortality pattern of a population, yet it has been mostly overlooked for mortality projections. In this chapter, we propose to analyse and forecast mortality developments over age and time by introducing a novel methodology based on age-at-death distributions. Our approach starts from a nonparametric decomposition of the mortality pattern into three independent components corresponding to Childhood, Early-adulthood and Senescence, respectively. We then model the evolution of each component-specific death density with a relational model that associates a time-invariant standard to a series of observed distributions by means of a transformation of the age axis. Our approach allows us to capture mortality developments over age and time, and forecasts can be derived from parameters' extrapolation using standard time series models. We illustrate our methods by estimating and forecasting the mortality pattern of females and males in two high-longevity countries using data of the Human Mortality Database. We compare the forecast accuracy of our model and its projections until 2050 with three other forecasting methodologies.

Keywords: Mortality forecasting · Mortality modelling · Relational models · Smoothing · Mortality decomposition · Life expectancy · Lifespan variability

1 Introduction

Population projections and mortality forecasts have been studied since the beginning of the twentieth century. The seminal works of Whelpton (1928, 1936) and Lotka (1939) on the cohort component method and the stable population contributed significantly to the development and application of population projections. Mortality forecasts go back at least to the beginning of the twentieth century, as actuaries were concerned about the financial effects of mortality improvements on life annuities and pensions (Pollard, 1987). It is however in the last three decades that mortality forecasting flourished, owing to the introduction and development of stochastic methodologies to project mortality.

---

[*]Corresponding author: `ugofilippo.basellini@ined.fr`

Three functions can be used to analyse human mortality and its developments over age and time: the hazard, the survival and the probability density function (Klein and Moeschberger, 2005). These functions describe the same stochastic phenomenon and are uniquely related between each other: one can derive any two of them by knowing the third one, without the need of additional information.

Despite the complementarity of the mortality functions, the majority of forecasting techniques is based on age-specific mortality rates or death probabilities (for comprehensive reviews, see Booth and Tickle, 2008; Cairns et al., 2009; Shang et al., 2011; Stoeldraijer et al., 2013). Most of these models take advantage of the regularities typically found in age- and time-patterns, such as the predominantly downward trend in age-specific mortality observed in many developed countries during the last 60 years, and they extrapolate the trends in the future using statistical methods (Haberman and Renshaw, 2011).

Nevertheless, the inspection of the other two functions can provide additional insights on mortality developments that one might not directly discern from a rate-based analysis. It is well known that the remarkable mortality improvements observed in these countries during the twentieth century are generally divided into two stages of mortality changes: compression and shifting dynamics (see, for example, Fries, 1980; Wilmoth and Horiuchi, 1999; Kannisto, 2000; Bongaarts, 2005; Canudas-Romo, 2008). Broadly speaking, the first stage took place in the first part of the century, as significant reductions in infant and childhood mortality resulted in grater equality in lengths of life. In the second part of the century, mortality improvements at older ages became more prominent, resulting in higher average lifespans with stagnating equality.

The age-at-death distribution is an excellent function to inspect these dynamics of mortality changes. Mortality compression can be detected from the reduction in the variability of the distribution, while shifting corresponds to a translation of the distribution to higher ages without relevant changes in its shape. In addition, the distribution provides immediate information on key questions in mortality studies, such as the longevity of the population, and the inequality in ages at death.

Figure 1 shows changes in the age-at-death distribution of Swiss males between 1950 and 2016. The graphical inspection of the death distribution readily provides information on the population's longevity, which is typically measured by life expectancy at birth or, in low mortality countries, by the modal age at death (Kannisto, 2001; Horiuchi et al., 2013). Additionally, the variability of lifespans within the population can be directly assessed from the spread of the distribution or its interquartile range. The increase in longevity as well as the reduction of lifespan variability for Swiss males during this period clearly emerge from Figure 1. Moreover, changes in the distribution over time highlight the two dynamics of mortality: for example, it is evident that the shifting dynamic of mortality started around the 1970-1980s, becoming more prominent in most recent decades, while the compression dynamic had been strongest in the decades 1950-1970 and 1990-2010.

Despite providing direct information on mortality patterns and trends over time, surprisingly few methods have been proposed to forecast mortality from age-at-death distributions. Among the firsts to abandon the conventional approach of using mortality rates, Oeppen (2008) and Oeppen and Camarda (2013) proposed to forecast the density of single and multiple-decrement life tables, using methodologies borrowed from compositional data analysis. Bergeron-Boucher et al. (2017a) expanded on this work, suggesting a coherent model based on life-table deaths of fifteen Western European countries. Furthermore, Basellini and Camarda (2019) proposed a relational model to forecast adult mortality from age-at-death distributions. Finally, Pascariu et al. (2019) suggested a vector autoregressive model to forecast the statistical moments of the death distribution.
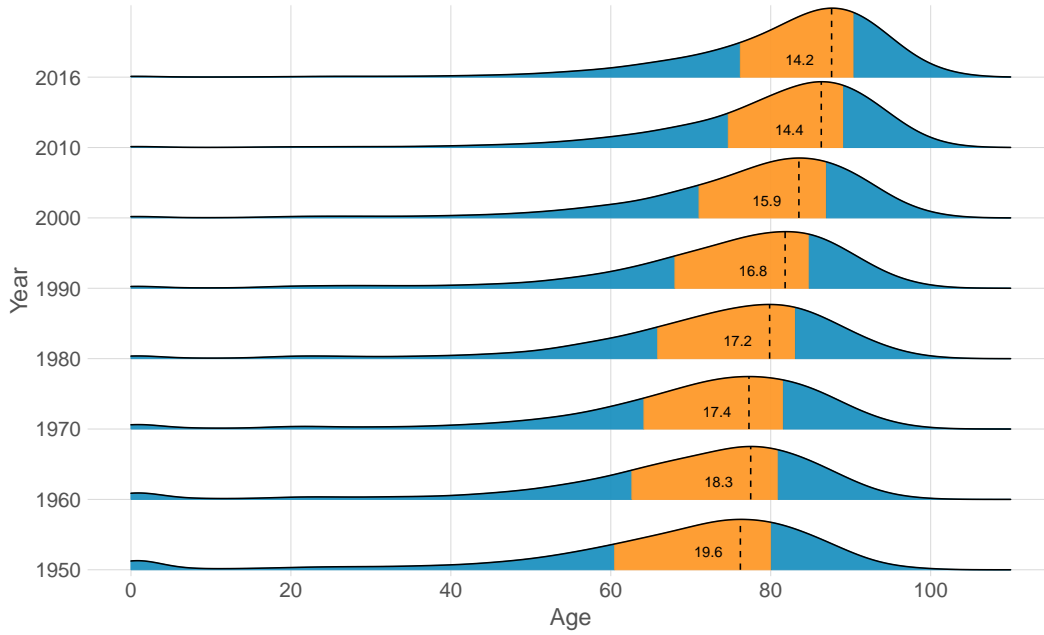
Figure 1. Changes in the age-at-death distribution for Swiss males at selected years between 1950 and 2016. The orange area corresponds to the interquartile range of the distribution, whose value is reported in print. The dashed line depicts the modal age at death. Data have been smoothed for illustrative purposes.

Source: authors' own elaborations on data retrieved from the Human Mortality Database (2019).

In this chapter, we contribute to the growing literature of forecasting the age-pattern of mortality from age-at-death distributions. Specifically, we extend the Segmented Transformation Age-at-death Distributions (STAD) model proposed by Basellini and Camarda (2019), which focuses on adult mortality only, to obtain mortality forecasts for the entire age range. While retaining the underlying methodology of the STAD model, here we introduce significant novelties to achieve our goal. In particular, our approach is based on two steps. First, we decompose the observed death counts into three additive mortality components, namely Childhood, Early-adulthood and Senescent mortality. We perform this decomposition via the nonparametric approach proposed by Camarda et al. (2016). Secondly, we model and forecast each component-specific age-at-death distribution employing specialized versions of the STAD model. As such, the Three-Component STAD (3C-STAD) model allows us to capture mortality developments over the entire age range, and forecasts are obtained from the extrapolation of the model's parameters using standard time-series techniques.

This chapter is organized as follows. In Section 2, we overview the methods that we introduce as well as the data that we employ. In Section 3, we provide two illustrations of our methodology by forecasting female and male mortality in two high-longevity countries. In particular, we first assess the accuracy of point and interval forecasts of the 3C-STAD model by performing three out-of-sample validation exercises. We then present the 3C-STAD forecasts until the year 2050. In both cases, we compare the 3C-STAD with three other well-known forecasting methodologies. Finally, in Section 4 we summarize and discuss our results.

2 Methods

2.1 Mortality functions

Human mortality can be analysed by any one of three complementary functions: the hazard, the survival and the probability density function (Klein and Moeschberger, 2005). In demography, for a given calendar year t, these functions are generally known as the force of mortality $\mu(x,t)$ at age x, the probability of surviving from birth to age x, $\ell(x,t)$, and the age-at-death distribution $f(x,t)$.

The three mortality functions are uniquely related between each other, and knowing one of them allows to determine the other two. In the following, without loss of generality, let $\ell(0,t)$, commonly labelled as the life-table radix, be equal to one, and let us drop the time index t to ease notation. The relationship that exists between the three functions at any age x is given by:

$$f(x) = \ell(x)\,\mu(x)\,. \tag{1}$$

The probability of surviving $\ell(x)$ can be derived from the other two mortality functions:

$$\ell(x) = \exp\left(-\int_0^x \mu(a)da\right)\,, \quad \ell(x) = \int_x^\omega f(x)\,dx\,, \tag{2}$$

where ω is the highest age attained in the population. Thus, combining (1) and (2) demonstrates the complementarity of the three mortality functions.

Since Thiele (1871), demographers and actuaries described human mortality into three different components that operates principally, or almost exclusively, upon childhood, middle and old ages, respectively. The attempt to decompose those three components stimulated numerous approaches (cf. Section 4). In a general setting, the hypothesis can be expressed as follows:

$$\mu(x) = \mu_c(x) + \mu_e(x) + \mu_s(x)\,, \tag{3}$$

where the force of mortality $\mu(x)$ at age x is additively decomposed into three independent components, $\mu_c(x)$, $\mu_e(x)$, and $\mu_s(x)$. For ease of presentation, we labelled these mortality component with Childhood, Early-adulthood and Senescence, respectively. However, they theoretically operate over all ages x. Combining (1) and (3), the corresponding decomposition of the age-at-death distribution can be written as follows:

$$\begin{aligned} f(x) &= \ell(x)\,\mu_c(x) + \ell(x)\,\mu_e(x) + \ell(x)\,\mu_s(x) \\ &= f_c(x) + f_e(x) + f_s(x)\,. \end{aligned} \tag{4}$$

2.2 Data and mortality decomposition

Whereas risk of death acts continuously, mortality functions and models can be displayed only at particular ages and years. For modelling and forecasting mortality and for a specific sex and population, available data are thus observed death counts, $d_{x,t}$, and central exposures to the risk of death, $e_{x,t}$, with ages $x = 0, \ldots, \omega$ and years t. In the following, we analyse the female and male populations of two high-longevity countries, Sweden and Switzerland, choosing a common time period (1950-2016) and with $\omega = 110+$. While Sweden was selected for the high standard in data quality, even at the oldest ages (Vaupel and Lundström, 1994; Wilmoth and Lundström, 1996), Switzerland was chosen for its atypical mortality development, especially for males, related to the strong HIV epidemic during the 1980s (Csete and Grob, 2012). Data are taken from the Human Mortality Database (HMD, 2019).

We assume that the number of deaths at age x and year t is a random variable $D_{x,t}$ that follows a Poisson process (Brillinger, 1986):

$$D_{x,t} \sim \mathcal{P}(e_{x,t}\,\mu_{x,t}) \tag{5}$$

where the force of mortality $\mu_{x,t}$ is assumed to be constant over each year of age (i.e. from age x to $x+1$) and over each calendar year (i.e. from year t to $t+1$). This assumption implies that $\mu_{x,t}$ approximates the force of mortality at exact age $x+\frac{1}{2}$ and exact time $t+\frac{1}{2}$ (Cairns et al., 2009). Note that the notation $\mu_{x,t}$ is the discrete counterpart of the continuous notation $\mu(x,t)$ employed in Subsection 2.1. Moreover, death rates $m_{x,t} = d_{x,t}/e_{x,t}$ are the maximum likelihood estimators of the force of mortality $\mu_{x,t}$, if no structure is enforced over age and/or time.

The first step in the Three-Component Segmented Transformation Age-at-death Distributions (3C-STAD) model concerns the decomposition of the force of mortality into its three independent components $\mu_k(x)$, $k = c, e, s$. Instead of employing a parametric mortality model, we favour a non-parametric approach to avoid imposing a rigid structure and achieve a better fit to the observed data. For this purpose, we employ the Sum of Smooth Exponentials (SSE) model, which has been shown to provide insightful results for mortality analysis (Camarda et al., 2016; Remund et al., 2018). In the following, we provide a short overview of the SSE model; for a more detailed description of the model, we refer the interested reader to the original paper of Camarda et al. (2016).

The SSE belongs to the class of multiple-component models (also known as competing hazard models, Gage, 1993), as it proposes an additive decomposition of the expected value of counts in multiple (smooth) components. In a given year t, let $\boldsymbol{\mu}$, $\boldsymbol{d}$ and $\boldsymbol{e}$ denote vectors over age of overall force of mortality, death counts and exposures, respectively. Within the SSE, we can model the force of mortality as the sum of three components $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_c : \boldsymbol{\gamma}_e : \boldsymbol{\gamma}_s]'$. The expected value of the Poisson process $\boldsymbol{d} \sim \mathcal{P}(\boldsymbol{e}\,\boldsymbol{\mu})$ is expressed as a composition of exposures and mortality components, i.e. $\boldsymbol{e}\,\boldsymbol{\mu} = \boldsymbol{C}\,\boldsymbol{\gamma}$, where the composition matrix $\boldsymbol{C} = [\boldsymbol{E} : \boldsymbol{E} : \boldsymbol{E}]$ is a block matrix that includes three times the diagonal matrix of population exposures $\boldsymbol{E} = \mathrm{diag}(\boldsymbol{e})$ (one for each component of mortality). The composition matrix has the dual role of multiplying each component by the exposure times and of summing them to obtain the overall Poisson mean. The SSE model can be framed as a Composite Link Model (Thompson and Baker, 1981), and estimation of the model's parameters can be obtained by a modified version of the iterative reweighted least squares (IWLS) algorithm (Eilers, 2007).

The SSE model has several advantages over parametric decompositions of the force of mortality, which made it our favoured choice for the first step of the 3C-STAD. Although the SSE could accommodate parametric assumptions, it allows to model each component by assuming only smoothness over age (and eventually over time). We opted for this last more flexible setting. This can be achieved by expressing each component k as a linear combination of B-spline basis $\boldsymbol{B}_k$ and associated coefficients $\boldsymbol{\alpha}_k$:

$$\boldsymbol{\gamma}_k = \exp\left(\boldsymbol{B}_k\,\boldsymbol{\alpha}_k\right), \quad k = c, e, s. \tag{6}$$

Smoothness of $\boldsymbol{\gamma}_k$ is obtained by combining a large number of B-splines and a roughness penalty on the coefficients vector $\boldsymbol{\alpha}_k$ (Eilers and Marx, 1996). Note that the exponential in (6) guarantees positive component-specific force of mortality, as one would expect. Furthermore, component-specific shape constraints can be easily specified and included in the estimation procedure by additional asymmetric penalties. Here, we enforce monotonic decreasing and increasing constraints on the Childhood and Senescent components, respectively, and a log-concave shape for the Early-adulthood component. These constraints further aid the identifiability of the model by ensuring that the three components are not interchangeable.
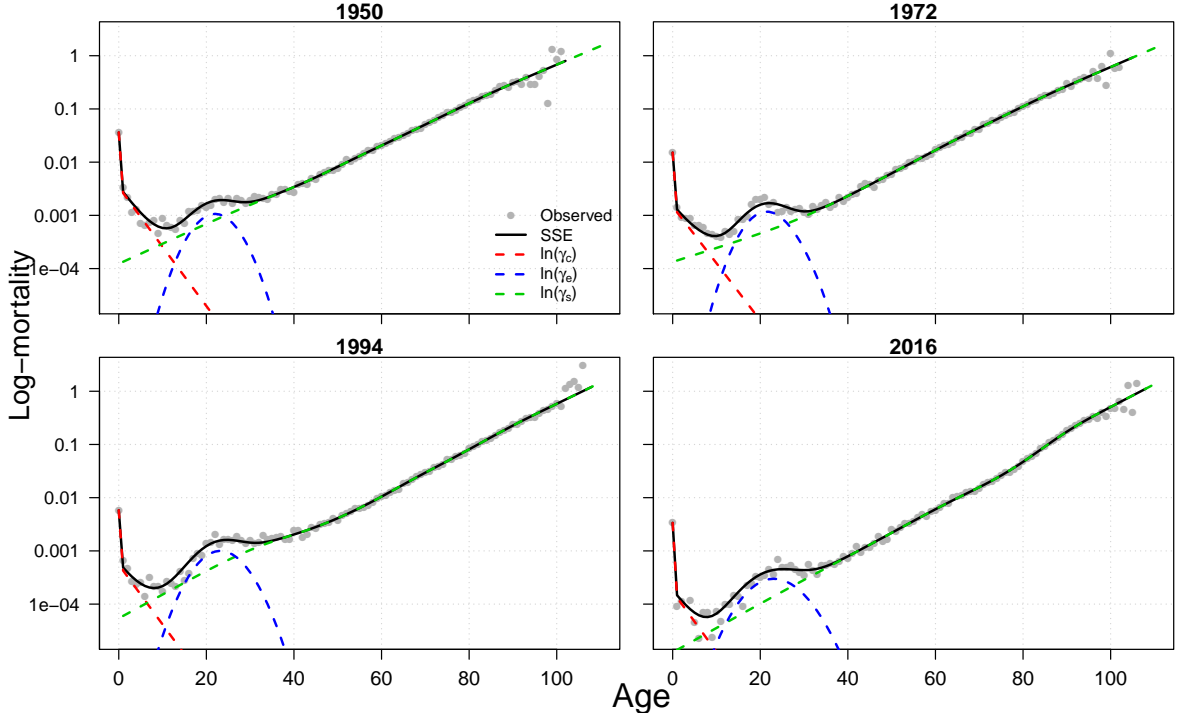
Figure 2. Observed and fitted mortality rates (in log scale) for Swiss males at selected years between 1950 and 2016. The force of mortality is decomposed into Childhood (γ_c), Early-adulthood (γ_e) and Senescent (γ_s) components via the two-dimensional SSE model. *Source*: as for Figure 1.

Another advantage of the SSE methodology is that it adequately blends the transitions between components, without imposing sharp delimitations where one stops and another one continues. Moreover, we employ the two-dimensional extension of the SSE model. In this way we both account for the significant age-time interactions and avoid abrupt changes over time in the interaction of the components. A detailed description of year-to-year mortality fluctuations is relevant in a forecasting perspective. In the SSE model, at the cost of overfitting, this flexibility is achieved by a large number of *B*-splines with a low smoothing parameter in the time dimension.

Figure 2 shows an example of fitting the two-dimensional SSE model to Swiss males between 1950 and 2016: the three components of mortality clearly emerge, each one featuring the expected shape. Unlike the original SSE model, we start our analysis from age 0 which is treated in a specific manner. This particular age represents a clear discontinuity in the age-pattern of mortality, as mortality of newborns is sharply higher than death rates at later infant ages due to malformations, pre-term births and birth-related complications (Chiang, 1984; Camarda et al., 2016). Hence, we incorporate the discontinuity in the first age of life by including, for the Childhood component, a specialized coefficient for this age, which is not penalized over age.

Outcomes from the SSE model allow us to obtain (i) the age-at-death distribution of each component over time (using standard life-table construction, Preston et al., 2001), and (ii) the expected number of deaths separated by component, $\hat{\boldsymbol{d}}_k = \boldsymbol{e}\,\hat{\boldsymbol{\gamma}}_k$. This allows us to model and forecast age-at-death distributions independently for each component.

6

2.3 Modelling component-specific distributions

The second step of the 3C-STAD consists in modelling the component-specific age-at-death distributions. Since different features characterize the three components, we deal differently with each one of them.

2.3.1 Senescent mortality

We start by presenting the model employed for the Senescent component, originally proposed and described in greater details in Basellini and Camarda (2019). The Segmented Transformation Age-at-death Distributions (STAD) is a relational model that relates a fixed time-invariant reference distribution, denoted standard, to a series of observed distributions via a segmented transformation of the age axis. In general, consider two age-at-death distributions $f(x)$ and $g(x)$, where the former is the standard, and the latter any observed distribution. The STAD model can be expressed as $g(x) = f[t(x; \boldsymbol{\theta})]$, where the transformation function $t(x; \boldsymbol{\theta})$ is characterized by three parameters $\boldsymbol{\theta}$ that depend on: (i) the difference in modal ages at death between the two distributions, and (ii) the change in the variability of the two distributions *before* and *after* their modal ages.

Let $\nu_s = M_s^g - M_s^f$ denotes the difference between the mode of the Senescent distributions $g_s(x)$ and $f_s(x)$. The transformation function of the STAD model for the Senescent component, $t_s(\cdot)$, can then be written as:

$$t_s(x;\ \nu_s, b_s^\ell, b_s^u) = \begin{cases} M_s^f + b_s^\ell\, \tilde{x} & \text{if } x \le M_s^g \\ M_s^f + b_s^u\, \tilde{x} & \text{if } x > M_s^g \end{cases} \tag{7}$$

where $\tilde{x} = x - \nu_s - M_s^f$, and b_s^ℓ and b_s^u denote the change in the variability of $g_s(x)$ with respect to $f_s(x)$ before and after the mode, respectively. Note that the superscript ℓ and u refer to the lower and upper segments of the age range (i.e. before and after the modal age at death).

The top panels in Figure 3 explain graphically the mechanisms underlying the STAD model for the Senescence component. Given a standard distribution (black lines in the graphs), let us consider the simpler case in which we vary the parameter ν_s but keep the variability parameters equal to 1, that is, $b_s^\ell = b_s^u = 1$. The transformation function in Equation (7) then simplifies to $t_s(x) = x - \nu_s$, and the resulting distribution is shifted along the x-axis by an amount equal to ν_s. This case corresponds to a shifting mortality scenario (blue lines in the graphs): the new distribution has the same shape and variability of the standard, but it is translated by the shifting parameter.

A more general development of mortality can be described by different values of the variability parameters, which act jointly with ν_s to modify the age-pattern of the standard distribution. When the two parameters are greater (lower) than 1, the variability of the segmented distribution is compressed (expanded) before and after the modal age at death with respect to the standard. In the top right panel of Figure 3, the segmented distribution has a lower variability ($b_s^\ell > 1$) before the mode and a higher variability ($b_s^u < 1$) above the mode as compared to the standard distribution. As such, increases in the two parameters capture the compression dynamic of mortality, distinguishing between changes that occur before and after the modal age at death.
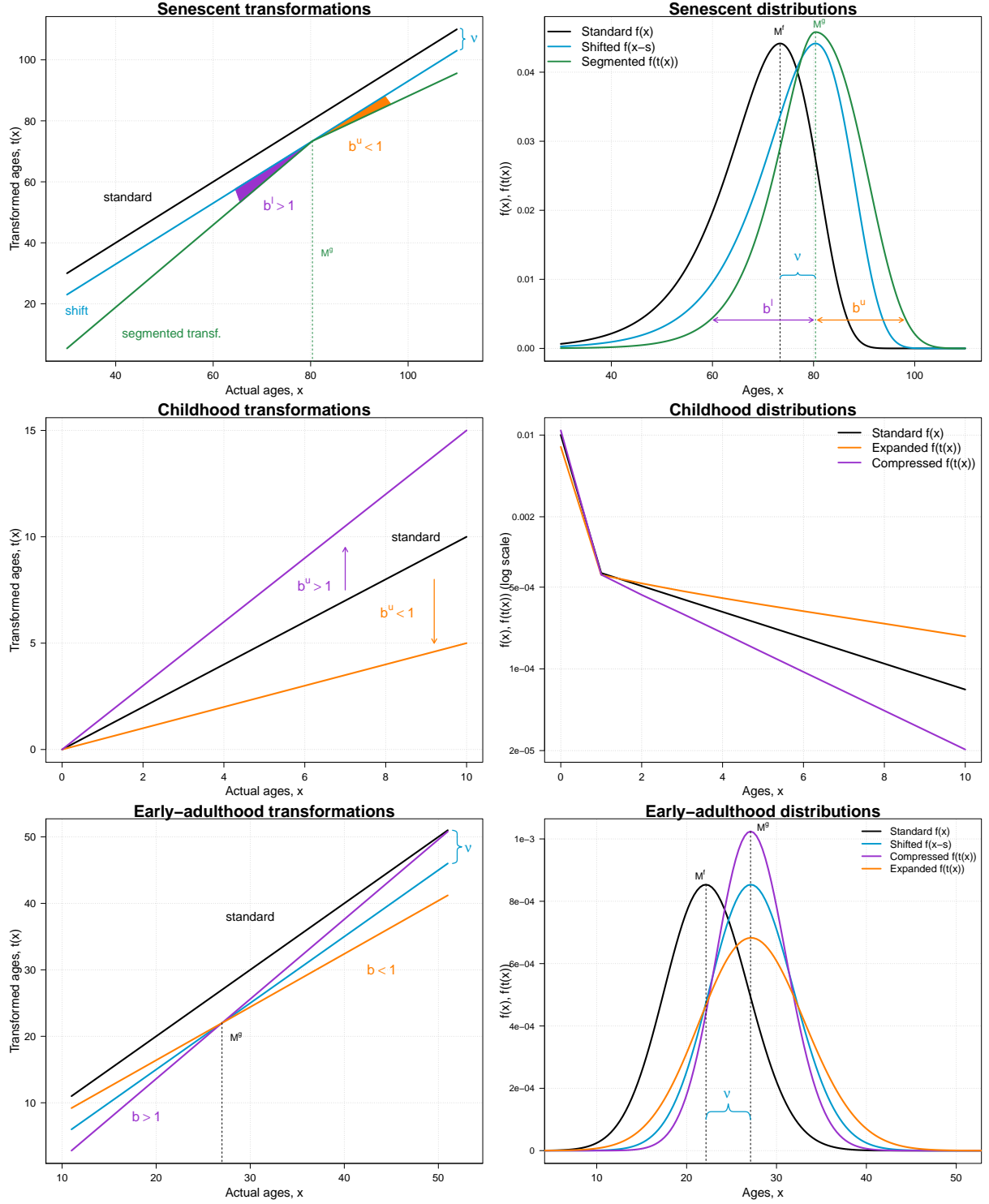
Figure 3. A graphical representation of the transformation functions (left panels) for the three components of the 3C-STAD model, and their effects on the corresponding component-specific age-at-death distributions (right panels).
Source: authors' own elaborations.

2.3.2 Childhood mortality

The modal age at death for the Childhood component is invariably at age 0. The STAD is thus simplified and we drop from the transformation in (7) the part below the mode, i.e., we consider a left-truncated distribution with a constant mode at age 0. For the Childhood component,

changes between the standard distribution, $f_c(x)$, and any observed distributions, $g_c(x)$, are modelled by varying the slope of the associated transformation of the age axis. In formulas, since $M_c^g = M_c^f = 0$, we can express the transformation of the age-axis as:

$$t_c(x;\ b_c^u) = b_c^u\, x\,. \tag{8}$$

The parameter b_c^u captures the change in the variability of the observed (left-truncated) distribution with respect to the standard distribution. The middle panels in Figure 3 present this case. A parameter b_c^u larger than 1 will reduce the variability of the Childhood age-at-death distribution with respect to the standard one (purple lines). Vice versa, a slope smaller than 1 will lead to an increase of the variance of the associated distribution (orange lines).

2.3.3 Early-adulthood mortality

The Early-adulthood component of mortality is a typical and distinguishable feature of the human mortality pattern, which has been observed and modelled since the very first approaches to mortality decomposition (e.g. Thiele, 1871; Lexis, 1878; Pearson, 1897). Cause-of-death investigations of young excess mortality have often provided relevant policy recommendations (Heuveline, 2002; Remund et al., 2018) [XXX: for Giancarlo: what do you think??]. As such, including this mortality component enhances the goodness-of-fit of the 3C-STAD, while increasing the plausibility of fitted and forecast age-profiles .

Transformations for the Early-adulthood component account for changes in the component-specific modal age-at-death and for the variability of the observed distribution, $g_e(x)$, always with respect to the standard one, $f_e(x)$. Unlike the original STAD model, a linear transformation of the age axis without segmentation has been proven adequate for describing changes of the Early-adulthood component over years. Therefore we do not differentiate between variability before and after the mode. This adaptation of the STAD can be thought as an Accelerated Failure Time model for age-at-death distributions, where the aging process is first shifted and then uniformly accelerated/decelerated with respect to the standard distribution.

Formally, we can write the transformation function for the Early-adulthood component as:

$$t_e(x;\ \nu_e, b_e) = M_e^f + b_e\, \tilde{x} \tag{9}$$

where $\tilde{x} = x - \nu_e - M_e^f$, $\nu_e = M_e^g - M_e^f$ and the parameter b_e captures the change in the variability of the observed distribution $g_e(x)$ with respect to the standard $f_e(x)$. Bottom panels in Figure 3 illustrates the effect of $t_e(\cdot)$ on a theoretical standard distribution. A shifting mortality scenario for Early-adulthood could be achieved by different values of the parameter ν_e, keeping $b_e = 1$ (blue lines). Alternatively, a b_e smaller than 1 leads to an increase of the variability of the distribution, simultaneously before and after the observed mode (orange lines). A shrinkage of the age axis is achieved by a b_e larger than 1, and it prompts a $g(x)$ with lower variability with respect to the standard $f_e(x)$ (purple lines).

2.4 Estimating and forecasting the 3C-STAD parameters

Being equipped with the component-specific transformation functions, we can move from the theoretical description of the 3C-STAD model to its actual application for modelling and forecasting a series of age-at-death distributions over time. The first step needed to achieve this goal is the choice of the standard distribution $f_k(x)$ for each component. For the Senescent component, we start by aligning the observed distributions to a common modal age at death, using a landmark registration approach frequently employed in Functional Data Analysis (Ramsay

Component, k	Transformation, $t_k(\cdot)$	Parameters Shift	Parameters Variability	Standard, $f_k(x)$
Senescence	Segmented at the mode	ν_s	b_s^ℓ, b_s^u	Mean of aligned distr.
Childhood	Left-Truncated, no shift	–	b_c^u	Mean of distr.
Early-adulthood	Linear, shift at the mode	ν_e	b_e	Mean of distr.

Table 1. Summary of the 3C-STAD model by component: type of transformation of the age axis, associated parameters and choice of the standard distribution.

and Silverman, 2005). The alignment procedure corresponds to a simple shifting transformation of the observed densities, which preserve all their features except the modal value. The standard is then computed as the mean of the aligned distributions. This approach increases the representativeness of the standard, which does not conflate features of the distributions that occur at different distances with respect to the mode (for additional details and an explicative illustration, see Basellini and Camarda, 2019, pp. 122–124). For the Childhood and Early-adulthood components, we choose the standard as simple means of the observed distributions, as the alignment procedure is not required for the former, and it does not significantly improve the fit for the latter.

Table 1 summarizes all hypotheses made in the 3C-STAD model about each component, and the associated parameters that are necessary to estimate and forecast. Given the component-specific standard distributions, parameters of the transformation functions $t_k(\cdot)$ are estimated from the data by maximum likelihood. Here we make use of the outcomes of the SSE model (cf. Subsection 2.2), and expected number of deaths over age and time due to each component k, $d_{x,t}^k$ are modelled by the 3C-STAD. Given the actual exposures $e_{x,t}$ and assuming that component-specific expected deaths are Poisson distributed counts as in (5), we maximize the following log-likelihood function for each year t:

$$\ln \mathcal{L} \left(\boldsymbol{\theta}_{k,t} \,|\, d_{x,t}^k, e_{x,t}, \nu_{k,t} \right) \propto \sum_x \left[d_{x,t}^k \ln \left(\hat{\mu}_{x,t}^k \right) - e_{x,t}\, \hat{\mu}_{x,t}^k \right] , \quad k = c, a, s \qquad (10)$$

where $\hat{\mu}_{x,t}^k$ denotes the hazard of component k corresponding to the transformed distribution derived from $t_k(\cdot)$ applied in year t to the associated standard $f_k(x)$. In particular, the hazard $\hat{\mu}_{x,t}^k$ is derived from the age-at-death distribution $f_k(t_k(\cdot))$ using standard life-table formulas (Preston et al., 2001).[1] Note that the vector $\boldsymbol{\theta}_{k,t}$ contains only the variability parameter(s). For each year t, the shifting parameters ν_s and ν_e of the Senescent and Early-adulthood components are computed as differences in the modal age at death between standard and observed distributions, as estimated by the SSE model.

Once the parameters have been estimated over all years t, we can model their trends using standard time-series methods. Mortality forecasts of the 3C-STAD model are then obtained by extrapolating the model's parameters and the time-fixed standard distributions. We combine univariate and multivariate approaches to achieve our goal. For the Senescent component, we employ the best fitting ARIMA(p,d,q) model for ν_s, and a VAR(1) model for b_s^ℓ and b_s^u (as in Basellini and Camarda, 2019). For the Childhood component, the parameter b_c^u is modelled with the best fitting ARIMA(p,d,q) model, while for the Early-adulthood parameters ν_e and b_e we employ a VAR(1) model.

The 3C-STAD acts directly on age-at-death distributions, therefore we must ensure that the sum over ages x of the three component-specific probability masses is equal to 1, that

---

[1]One readily implemented approach to derive the hazard from age-at-death distribution in R is provided by the function `convertFx` in the `MortalityLaws` package (Pascariu, 2018).

is:

$$\sum_x f_{x,t} = \sum_x \left(f_{x,t}^c + f_{x,t}^e + f_{x,t}^s \right) = 1 \tag{11}$$

for each year t. Consequently and in addition to the shifting/variability parameters, it is necessary to forecast the probability masses of the three components. Specifically, we recognize the compositional nature of a set of component-specific age-at-death distributions: we are dealing with three non-negative components that always sum to a constant. We thus employ a Compositional Data methodology to model and forecast the time series of component-specific probability masses (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011). Specifically, we transform the probability masses for each component obtained by the SSE model using an additive log-ratio transformation. This procedure produces two time-series that are unconstrained (i.e. they take values on the entire set of real numbers). The two transformed time-series are modelled and forecast with a VAR(1). We finally back-transform the results to obtain forecasts of the original time-series. For each forecast year, these back-transformed series sum up to 1 because they have been treated as compositional data. Note that this approach reduces the dimensionality of the forecasting problem for the probability masses by one dimension, i.e. from three to two time-series.

Finally, the complexity of our methodology requires a bootstrapping procedure to produce prediction intervals (PI, Efron and Tibshirani, 1994). We take into account the uncertainty of the 3C-STAD parameters by simulating 1000 new time-series of all parameters from randomly resampled residual values. For each simulation, we then forecast mortality patterns and associated summary measures. From the obtained distribution of forecast simulations, we took the median as central forecast, and the lowest and highest deciles to construct 80% PI. Residual bootstrap of this type has already been employed to construct PI in mortality models (Bergeron-Boucher et al., 2017a; Basellini and Camarda, 2019).

Routines for estimating and forecasting the parameters of the 3C-STAD model were implemented in R (R Development Core Team, 2018) and are available online[2]. Our routines take advantage of the R packages `forecast`, `demography`, `MortalitySmooth`, `MortalityLaws` and `vars` (Pfaff, 2008a,b; Hyndman and Khandakar, 2008; Camarda, 2012; Hyndman et al., 2018a,b; Pascariu, 2018).

3 Results

3.1 Out-of-sample validation

Here, we assess the predictive performance of the 3C-STAD model using out-of-sample validation. Specifically, we employ data of the Human Mortality Database (2019) for the female and male populations of Sweden and Switzerland for the period 1950-2016. For each population, we perform three exercises, corresponding to validation periods of 10 years (training period 1950-2006), 20 years (training period 1950-1996) and 30 years (training period 1950-1986). The common starting year of analysis, 1950, was chosen in order to have training periods longer than validation horizons for each exercise.

To assess the performance of our forecasts, we employ the standard life-table functions: life expectancy at birth (e_t^0) as measure of population's longevity in year t, and age-specific mortality rates (in log scale, $\ln(m_{x,t})$), which measure the age-pattern and intensity of mortality.

---

[2]`R` codes to replicate all results presented in this chapter are available at [the GitHub repository will be made public upon eventual acceptance of the chapter]

Additionally, we use the Gini coefficient (g_t^0), a measure of lifespan inequality, whose importance for evaluating mortality forecasts has been recently highlighted (Bohk-Ewald et al., 2017).

We compare the performance accuracy of the 3C-STAD model with three other forecasting methodologies. First, given its prominence and wide application, we employ the original model of Lee and Carter (LC, 1992). Second, since one limitation of the LC model is the lack of smoothness in the fitted and forecast mortality rates, we use the functional data model of Hyndman and Ullah (HU, 2007), which overcomes this limitation by smoothing the starting data as a first step. Third, we choose the CODA model proposed by Oeppen (2008): this model is indeed closer in spirit to the 3C-STAD, as it models and forecasts the age-at-death distribution. The LC and HU models were estimated and forecast with the R packages `forecast` and `demography` (Hyndman et al., 2018a,b; Hyndman and Khandakar, 2008). The CODA model was fitted and forecast using the R codes provided in the Supplementary Material of Bergeron-Boucher et al. (2017a).

Our evaluations of mortality forecasts are based on the accuracy of both point predictions and calibration of prediction intervals (PI), as both measures are relevant for the validation of probabilistic projections (Chatfield, 2000). Greater accuracy in point forecasts occurs when point predictions are closer to the observed data. To evaluate point forecasts, we employ the mean absolute error (MAE), which is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{t \in T} |\hat{y}_t - y_t|$$

where $\hat{y}_t$ is the point forecast at time t for either life expectancy at birth, mortality rates or Gini coefficient. Associated out-of-sample observed values are denoted by y_t. The set of validation years is T, and N is the total number of data used for validation. Note that for mortality rates, mean is computed over ages, too.

Greater calibration of PI is achieved when the proportion of out-of-sample data that falls within the calculated PI is closer to the given nominal level (for example, 80% or 95%). To evaluate interval forecasts, we compute the empirical coverage probability (ECP) of the 80% PI for each model (as in, for example, Shang et al., 2011; Raftery et al., 2013). For the sake of consistency and fairness, we computed the PI for all models by the same bootstrapping procedure, i.e. residual bootstrap of the time-series of the model's parameters (cf. Subsection 2.4).

In addition to the MAE and ECP, proper scoring rules can be used to assess calibration and sharpness of probabilistic forecasts simultaneously (for a review, see Gneiting and Katzfuss, 2014). Scoring rules allow to jointly assess point and interval predictions by providing a summary measure of the predictive performance that forecasters aim to minimize. Here, we employ the proper Dawid-Sebastiani score (DSS) (Dawid and Sebastiani, 1999), which is given by:

$$\text{DSS} = \frac{(y - \mu_F)^2}{\sigma_F^2} + 2 \ln \sigma_F$$

where μ_F and σ_F^2 are the first two central moments of the probabilistic forecast, and y is the associated out-of-sample observed value.

Table 2 reports the point, interval and probabilistic forecast accuracy of the four models in the three out-of-sample scenarios as well as for all the four populations analysed here. Bold values correspond to better performances. In terms of point forecast, the 3C-STAD is the most accurate model, as its forecasts are more or as precise as those of the other models. Out of 36 indicators, the 3C-STAD outperforms 20 times. The HU is the second most precise model with 9 indicators, followed by the LC and CODA models, each with 8 and 3 indicators, respectively.

Note that the sum does not add up to the total number of indicators due to the draw of some models for some specific measures (for example, both the 3C-STAD and LC models are equally best performers for the indicator g^0 for Swedish females in the 30y exercise). In terms of interval forecast, the CODA outperforms all other models, being more accurate for 15 indicators over 36. The 3C-STAD, LC and HU follow, each with 12, 11 and 7 indicators, respectively. Finally, if we consider point and prediction accuracy simultaneously using the DSS measure, we find that the HU and 3C-STAD models are the best performer, each one outperforming the others for 12 and 11 indicators, respectively. The LC model follows with 9, while the CODA is the best performer for only 4 indicators.

3.2 Forecast to 2050

Having assessed and compared the forecast accuracy of the 3C-STAD model, we now present its mortality forecasts for the four populations analysed until 2050. As in the previous Subsection, we compare projections based on the 3C-STAD model with those of LC, HU and CODA models.

Figure 4 shows the observed and forecast life expectancy at birth (e^0) and Gini coefficient (g^0) in the four populations for the years 1950-2050. In terms of e^0, the 3C-STAD forecasts are always more optimistic than those of the LC and HU model. With respect to CODA, the 3C-STAD is more optimistic for males and less optimistic for females. In terms of lifespan inequality, CODA forecasts are the most egalitarian in 2050 (lower values of g^0) for the female populations, while the 3C-STAD predicts more equality for males.

In Figure 5, we compare the age-specific mortality rates forecasts in 2050 for all populations. Several differences emerge between the models from this age-pattern analysis. Mortality rates of the 3C-STAD are smooth, lacking the jagged features visible in the LC and CODA forecasts. This is a great advantage for long-term mortality projections (Li et al., 2013). Additionally, the Swedish projections of the 3C-STAD do not display an unexpected S-shape displayed by other models in the age range 60-100.

Finally, Figure 6 shows the observed age-at-death distribution for the four populations in 2016, along with the 2050 forecasts of the four models. With respect to the other models, the 3C-STAD forecasts are characterized by greatest shift for all the populations. In addition to this, the 3C-STAD projections are also less compressed than those of other models, with the exception of Swedish males.

Forecast Accuracy

Country	Training period	Validation period	Measure	Point forecast (MAE)				Interval forecast (ECP for 80% PI)				Probabilistic forecast (mean DSS)			
				3C-STAD	LC	CODA	HU	3C-STAD	LC	CODA	HU	3C-STAD	LC	CODA	HU
Switzerland Females	1950-2006	10y	e^0	**0.11**	0.29	0.40	0.16	100%	100%	100%	**90%**	1.10	1.40	1.61	**1.08**
			g^0	0.08	0.11	0.09	**0.06**	100%	**80%**	90%	**80%**	0.92	1.13	**0.71**	1.12
			$\ln(m_{x,t})$	**0.19**	0.24	0.24	0.22	38%	35%	**47%**	37%	25.25	40.51	74.56	**11.61**
	1950-1996	20y	e^0	0.13	0.11	0.39	**0.12**	100%	100%	100%	**95%**	1.69	1.65	1.70	**0.91**
			g^0	0.08	**0.06**	0.21	0.10	90%	100%	95%	**85%**	0.97	**0.63**	1.22	1.34
			$\ln(m_{x,t})$	**0.19**	0.24	0.23	0.23	51%	48%	**54%**	38%	15.13	48.47	32.40	**10.13**
	1950-1986	30y	e^0	0.74	0.43	0.45	**0.20**	**87%**	100%	97%	93%	2.48	2.18	1.75	**1.44**
			g^0	0.21	0.40	0.30	**0.13**	70%	63%	37%	**77%**	2.54	2.96	2.43	**1.30**
			$\ln(m_{x,t})$	0.22	0.28	0.23	**0.20**	54%	50%	**58%**	43%	12.16	20.08	22.54	**8.00**
Switzerland Males	1950-2006	10y	e^0	**0.19**	0.41	0.34	0.78	100%	**90%**	100%	0%	**1.17**	1.70	1.84	6.95
			g^0	0.13	0.21	**0.12**	0.26	**80%**	70%	100%	10%	1.15	1.84	**0.85**	5.83
			$\ln(m_{x,t})$	**0.17**	0.23	0.20	0.20	37%	32%	**53%**	31%	13.70	39.52	53.54	**8.54**
	1950-1996	20y	e^0	**1.39**	1.59	1.61	2.23	10%	20%	25%	0%	5.09	5.31	**4.80**	46.32
			g^0	**0.46**	0.74	0.65	1.01	15%	0%	**20%**	0%	**4.45**	9.83	4.49	56.09
			$\ln(m_{x,t})$	**0.25**	0.29	0.28	0.32	29%	29%	**43%**	17%	**20.32**	168137	71.00	21.54
	1950-1986	30y	e^0	**1.55**	1.63	2.27	1.94	37%	**47%**	30%	17%	6.17	**4.52**	6.23	34.29
			g^0	0.98	**0.64**	0.88	0.71	20%	27%	**27%**	17%	6.05	**4.72**	6.07	43.39
			$\ln(m_{x,t})$	0.31	**0.27**	0.31	**0.27**	25%	35%	**42%**	21%	**19.86**	103.94	22.85	23.81
Sweden Females	1950-2006	10y	e^0	0.12	**0.10**	0.38	0.15	**100%**	100%	100%	100%	**0.82**	0.99	1.66	0.99
			g^0	0.07	**0.05**	0.13	0.12	**80%**	90%	100%	50%	1.30	**0.70**	0.82	2.93
			$\ln(m_{x,t})$	**0.16**	0.18	0.22	**0.16**	37%	39%	**55%**	44%	25.12	19.83	93.65	**9.91**
	1950-1996	20y	e^0	0.46	**0.17**	0.73	0.46	**90%**	100%	95%	20%	1.79	**1.53**	2.42	2.90
			g^0	**0.06**	0.10	0.20	**0.06**	95%	90%	95%	**85%**	**0.65**	1.06	1.16	0.99
			$\ln(m_{x,t})$	**0.18**	0.19	0.22	0.19	43%	45%	**54%**	28%	24.70	**11.06**	52.18	12.26
	1950-1986	30y	e^0	0.69	0.53	1.43	**0.43**	23%	**97%**	7%	30%	3.62	**2.08**	10.81	3.04
			g^0	**0.07**	**0.07**	0.34	0.09	**80%**	87%	20%	70%	1.34	**1.03**	6.13	1.64
			$\ln(m_{x,t})$	**0.18**	0.21	0.27	0.19	**39%**	39%	34%	29%	**15.92**	33.39	87.08	16.73
Sweden Males	1950-2006	10y	e^0	**0.11**	0.38	0.41	0.81	**100%**	100%	100%	0%	**1.05**	1.57	1.62	7.75
			g^0	0.10	0.09	**0.03**	0.04	70%	90%	100%	100%	1.29	1.06	0.42	**0.38**
			$\ln(m_{x,t})$	**0.17**	0.23	0.19	0.18	43%	26%	**49%**	27%	25.41	120.36	127.46	**11.36**
	1950-1996	20y	e^0	**0.72**	1.35	1.34	1.69	**100%**	30%	15%	0%	**2.61**	4.48	7.82	18.73
			g^0	0.32	0.37	**0.26**	0.39	15%	10%	**40%**	0%	4.00	3.36	**3.20**	5.42
			$\ln(m_{x,t})$	**0.20**	0.29	0.23	0.21	**51%**	21%	32%	27%	15.73	120.98	70.90	**12.10**
	1950-1986	30y	e^0	2.65	**2.43**	2.59	3.14	7%	**13%**	7%	0%	20.92	**9.79**	75.35	153.06
			g^0	**0.40**	0.83	0.93	1.02	**23%**	17%	10%	10%	**3.54**	13.67	72.39	80.96
			$\ln(m_{x,t})$	**0.29**	0.38	0.31	0.31	18%	**28%**	25%	16%	**21.87**	98.09	56.84	50.36

Table 2. Mean absolute error (MAE), empirical coverage probability (ECP) for the 80% PI, and mean Dawid-Sebastiani score (DSS) of the 3C-STAD, LC, CODA and HU forecasts of e^0, g^0 and $\ln(m_{x,t})$ for females and males in two countries and three out-of-sample exercises: validation periods of 10 years (training period 1950-2006), 20 years (1950-1996) and 30 years (1950-1986). Lower values of the MAE and of the DSS (in bold) correspond to greater forecast accuracy. Values of the ECP closer to the 80% nominal level (in bold) correspond to greater interval forecast accuracy.

Source: as for Figure 1.

14
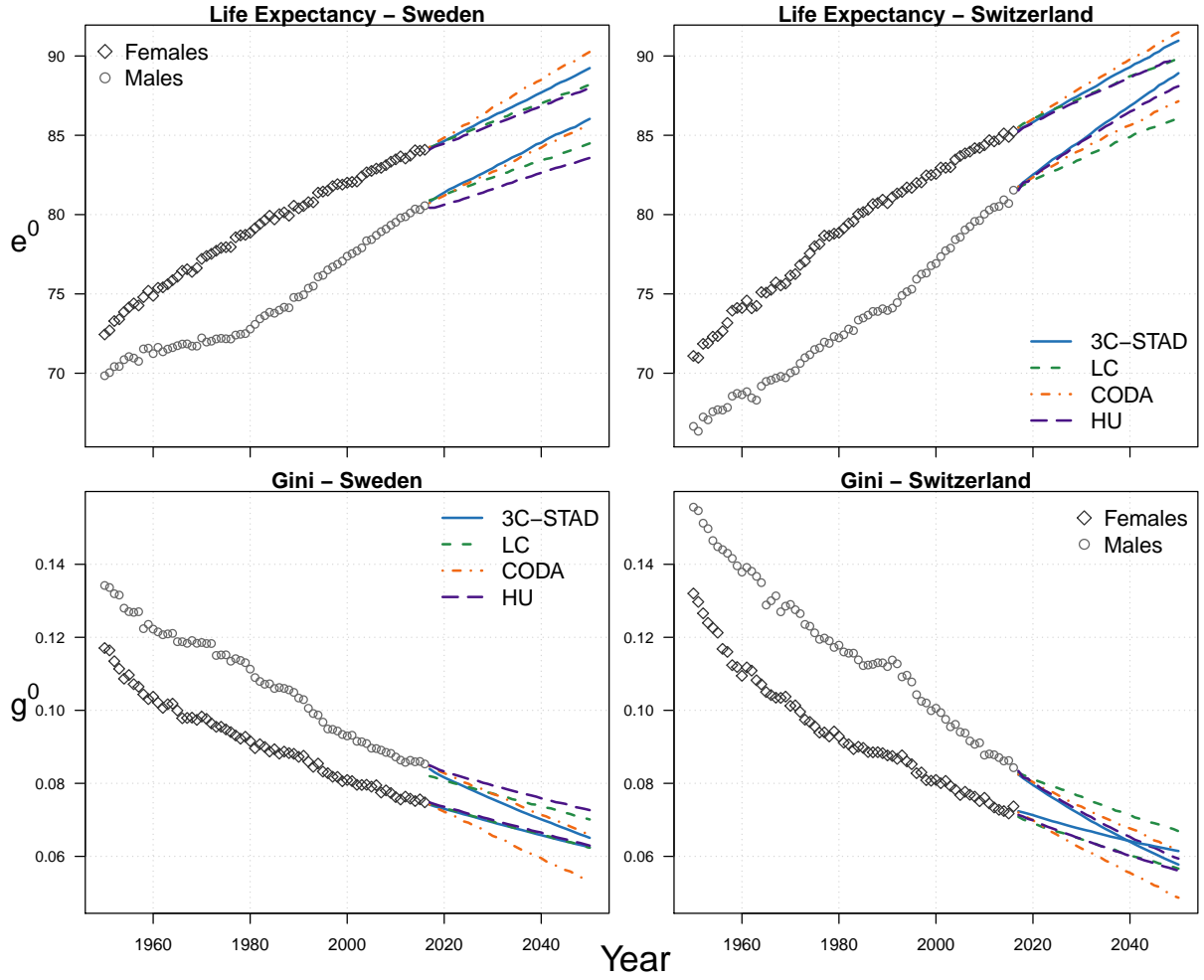
Figure 4. Observed and forecast life expectancy at birth (e^0, top panels) and Gini coefficient (g^0, bottom panels) females and males in Sweden and Switzerland, 1950-2050.
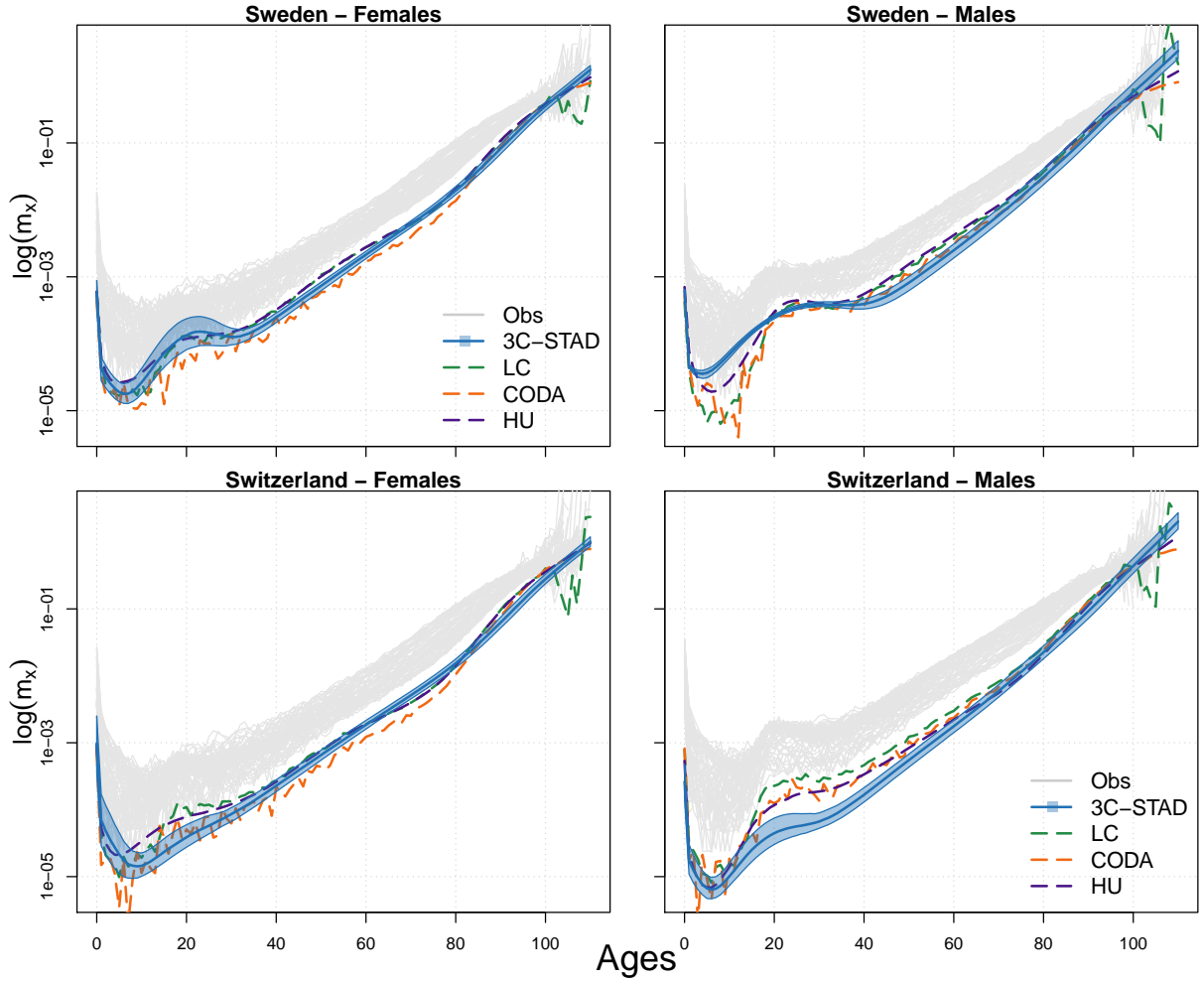*Source*: as for Figure 1.

4 Discussion

Age-at-death distributions have generally been neglected for modelling and forecasting mortality, despite providing insightful information on mortality age-patterns and trends over time. In this chapter, we introduced a novel stochastic methodology to forecast mortality that is based on changes in age-at-death distributions. Our proposed Three-Component Segmented Transformation Age-at-death Distributions (3C-STAD) model captures and forecasts mortality developments over age and time by: (i) decomposing mortality into three independent components, namely Childhood, Early-adulthood and Senescence, and (ii) modelling and forecasting changes in each component-specific age-at-death distributions.

The decomposition of the mortality age-pattern into multiple components has a long history in demographic analysis. In 1871, Thiele pioneered this decomposition by expressing the force of mortality as the sum of three independent components that operate principally, or almost exclusively, upon childhood, middle and old ages, respectively. Shortly afterwards, Lexis (1878) theorized a similar three-component decomposition, but he shifted the attention from the force of mortality to the age-at-death distribution. His ideas were followed upon and further elaborated by Pearson (1897), who divided the death density into five components, each one

Figure 5. Observed age-specific mortality rates in 1950–2016 (grey lines) and forecast rates of four models in 2050 for females and males in Sweden and Switzerland. Shaded areas correspond to 80% PI for the 3C-STAD model.
Source: as for Figure 1.

with its own distribution with different masses and degree of skewness. Finally, more recently, different parametric approaches have been proposed to decompose human mortality patterns (Siler, 1979; Heligman and Pollard, 1980; Kostaki, 1992; de Beer and Janssen, 2016; Mazzuco et al., 2018).

For our purposes, we performed a non-parametric decomposition using the Sum of Smooth Exponentials (SSE) model (Camarda et al., 2016). We favour this over other parametric approaches because it allows us to achieve a good fit to the observed data without imposing a rigid parametric structure, hence adapting the decomposition to a large and diverse range of mortality developments. Moreover, via the SSE model, we obtain smooth components with specific shape constraints, and a two-dimensional age-time perspective is incorporated into the mortality decomposition. Component-specific age-at-death distributions derived by the SSE model are then isolated to model and forecast their changes. To do so, we employ modified versions of the relational model proposed by Basellini and Camarda (2019), originally designed for forecasting only adult distributions of deaths.

We have applied the 3C-STAD model to the female and male populations of Sweden and Switzerland using data retrieved from the Human Mortality Database (2019). First, we assessed the point and interval forecast accuracy of the model by performing three out-of sample
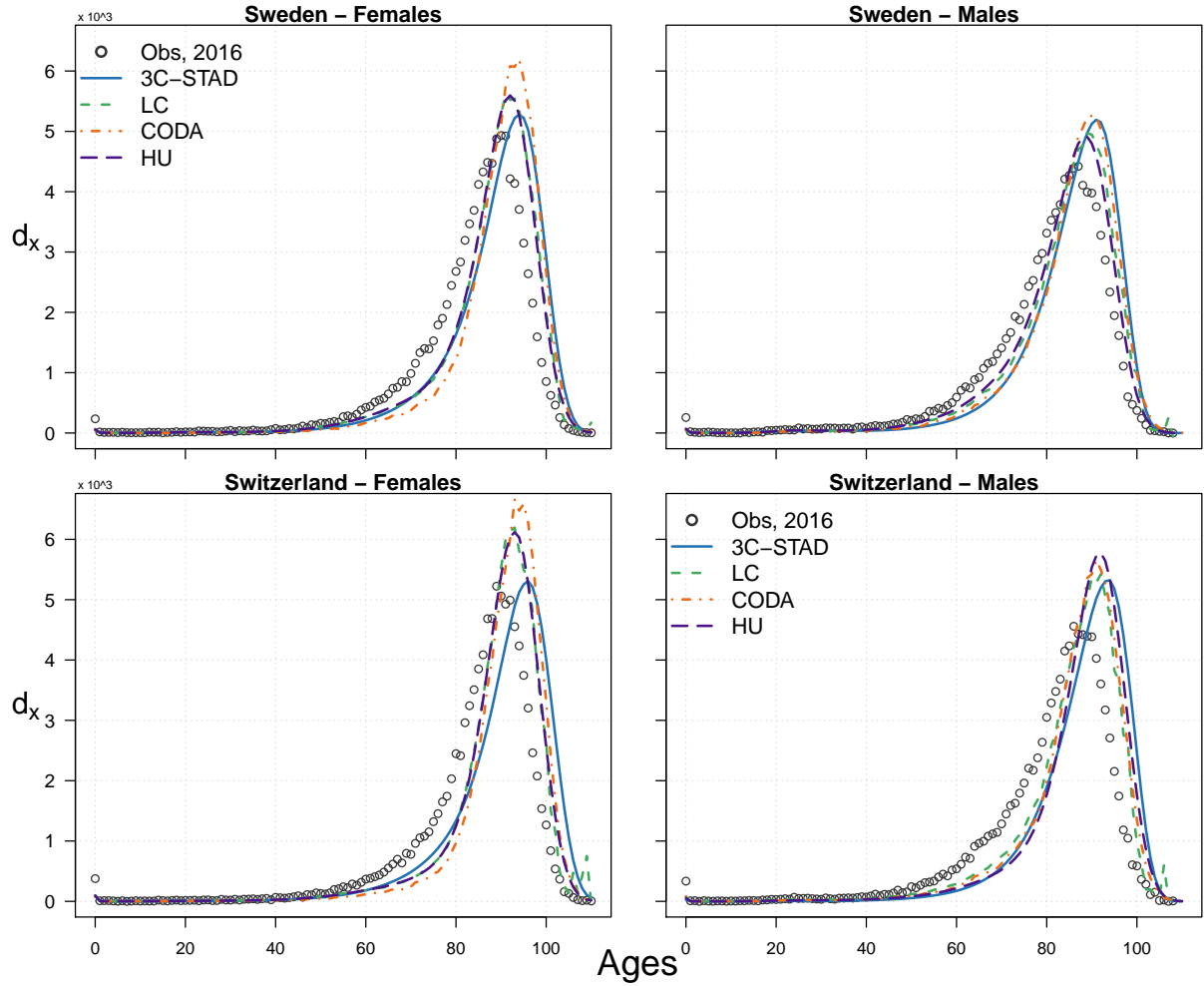
Figure 6. Observed age-at-death distribution in 2016 (grey points) and forecasts of four models in 2050 for females and males in Sweden and Switzerland.
Source: as for Figure 1.

validation exercises. We have then forecast mortality for each population until 2050. In both cases, we compared the 3C-STAD projections with those of three well-known and employed methodologies: the Lee and Carter (LC, 1992), the CODA (Oeppen, 2008) and the Hyndman and Ullah (HU, 2007) models. We compare forecasts of summary measures, such as life expectancy as birth (e^0) and lifespan inequality (as measured by the Gini coefficient, g^0), as well as age-specific functions, such as death rates or age-at-death distributions.

The results of the out-of sample validation exercises show that the 3C-STAD produces accurate mortality forecasts, both in terms of point forecasts and prediction intervals (PI). In particular, the 3C-STAD was the most accurate model for point forecasts with respect to other models. Additionally, the 3C-STAD PI outperformed the other models for about one indicator out of four (see Table 2).

Concerning interval forecasts, CODA was found relatively more accurate, a result that might be related to the fact that "the PI are wider with a CODA method than with an LC method" (Bergeron-Boucher et al., 2017a, pp. 546). The evidence that (very) wide PI (as those produced by CODA) correspond to higher interval forecasts accuracy requires some further analysis. Within 3C-STAD framework, a possibility to improve estimation of PI would be to include the uncertainty related to the SSE decomposition. However, preliminary analyses showed that this approach raises computational burden without a significant widening of the

17

forecast variability. It is likely that the reason is due to our usage of the SSE model. In the decomposition procedure, we aim to follow mortality data as close as possible, consequently the SSE model presents extremely small uncertainty. Nonetheless, we envisage alternative procedures to further improve estimation of the interval accuracy of the 3C-STAD model.

Mortality forecasts until 2050 for the four populations highlighted additional differences between models. The 3C-STAD and CODA forecasts of e^0 are generally more optimistic than those of the LC and HU models. Forecasting age-at-death distributions instead of mortality rates here translates into more optimistic forecast of life expectancy, a preliminary finding already observed elsewhere (Bergeron-Boucher et al., 2017b). This could be an advantage, given that the LC forecasts have often under-predicted future gains in life expectancy (Lee and Miller, 2001). Significant differences further emerge from an age-specific analysis of the different projections. On one side, the 3C-STAD forecast rates are inherently smooth, which is a desirable property especially for longer-term projections (Li et al., 2013). On the other side, the 3C-STAD forecast age-at-death distributions are characterized by greater shifting and smaller compression than those of other models. These projections seem more plausible, given that the shifting mortality dynamic has replaced the compression one in high-longevity countries in the most recent decades (Canudas-Romo, 2008; Bergeron-Boucher et al., 2015; Janssen and de Beer, 2019).

In general, we regard three characteristics as desirable for any forecasting methodology. First, the model should be able to capture and forecast mortality trends that can move in different directions across ages. Second, the relevant dynamics of mortality changes observed during the last century, i.e. shift and compression, should be appropriately accounted for. Third, the forecast age-profile of mortality rates should be smooth, without implausible jaggedness where rates of adjacent age groups have very different and volatile values. Despite being one of the most employed forecasting methodology by public and private companies, the seminal LC model does not satisfy any of these properties. The single time index regulates the direction of change for mortality rates at all ages, i.e. mortality improvements occur in the same direction at all ages. Furthermore, the model cannot account for the two mortality dynamics, and forecasts age-pattern are very volatile and jagged (see Figures 5 and 6).

Conversely, the 3C-STAD model meets all these three requirements. On one hand, the mortality decomposition allows us to capture and forecast mortality improvements across ages without rigid assumptions. Smoothness in the fitted and forecast age-profiles is a by-product of the non-parametric decomposition that we have employed. On the other hand, the 3C-STAD parameters capture and disentangle the shifting and compression mortality dynamics. The recently proposed model of Bardoutsos et al. (2018) is another example of projection methodology that satisfies these features.

Obviously, the 3C-STAD is not free of shortcomings, and neither we claim here that it outperforms all other forecasting methodologies. In addition to the width of the PI mentioned before, the computational time needed to produce mortality forecasts could be improved. The estimation of the two-dimensional SSE model in fact generally requires around thirty minutes, and speeding this step up will be required to shorten computational times. Future mortality values are obtained by forecasting eight time-series. Although this feature might pose issues in other situations, all of these series have clear demographic meanings and rather intelligible trends. Combination of univariate and multivariate time-series approaches has thus provided a reliable tool for overcoming this seemingly critical drawback of the 3C-STAD model. Different approaches in extrapolating the eight time-series will be pursuit, also for assessing consequences of specific future demographic scenarios. Moreover, in line with recent literature (Li and Lee, 2005; Hyndman et al., 2013; Janssen et al., 2013; Bergeron-Boucher et al., 2017a), future research will be directed towards the inclusion of coherence as an additional factor to improve forecasts

for a group of (sub)populations.

To conclude, we have shown that the proposed 3C-STAD model offers great prospects for modelling and forecasting human mortality. In light of the generally pessimistic forecasts of the widely employed LC model (Li et al., 2013; Seligman et al., 2016), forecasting methodologies, such as the 3C-STAD, should be explored by pension and insurance providers to better assess their solvency needs, and by statistical bureaus to produce alternative population projections.

References

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

Bardoutsos, A., de Beer, J., and Janssen, F. (2018). Projecting delay and compression of mortality. *Genus*, 74(1).

Basellini, U. and Camarda, C. G. (2019). Modelling and forecasting adult age-at-death distributions. *Population Studies*, 73(1):119–138.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., and Vaupel, J. W. (2017a). Coherent forecasts of mortality with compositional data analysis. *Demographic Research*, 37(17):527–566.

Bergeron-Boucher, M.-P., Ebeling, M., and Canudas-Romo, V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research*, 33(14):391–424.

Bergeron-Boucher, M.-P., Kjærgaard, S., and Oeppen, J. (2017b). What to forecast? An exploratory study of the implication of using different indicators to forecast mortality. In *XXVIII International Population Conference*.

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan Disparity as an Additional Indicator for Evaluating Mortality Forecasts. *Demography*, 54(4):1559–1577.

Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography*, 42(1):23–49.

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

Brillinger, D. R. (1986). A biometrics invited paper with discussion: The natural variability of vital rates and associated statistics. *Biometrics*, 42(4):693–734.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.

Camarda, C. G. (2012). MortalitySmooth: An R Package for Smoothing Poisson Counts with *P*-Splines. *Journal of Statistical Software*, 50:1–24. Available on www.jstatsoft.org/v50/i01.

Camarda, C. G., Eilers, P. H., and Gampe, J. (2016). Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling*, 16(4):279–296.

Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19:1179–1204.

Chatfield, C. (2000). *Time-series forecasting*. Chapman & Hall/CRC: London.

Chiang, C. L. (1984). *The Life Table and its Application*. Malabar, FL: Krieger.

Csete, J. and Grob, P. J. (2012). Switzerland, HIV and the power of pragmatism: Lessons for drug policy development. *International Journal of Drug Policy*, 23(1):82–86.

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, 27(1):65–81.

de Beer, J. and Janssen, F. (2016). A new parametric model to assess delay and compression of mortality. *Population Health Metrics*, 14(1):46.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Eilers, P. H. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, 7(3):239–254.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with *B*-splines and Penalties (with discussion). *Statistical Science*, 11(2):89–102.

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303(3):130–135.

Gage, T. B. (1993). The Decline of Mortality in England and Wales 1861 to 1964: Decomposition by Cause of Death and Component of Mortality. *Population Studies*, 47(1):47–66.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151.

Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Heuveline, P. (2002). An International Comparison of Adolescent and Young Adult Mortality. *The ANNALS of the American Academy of Political and Social Science*, 580(1):172–200.

Horiuchi, S., Ouellette, N., Cheung, S. L. K., and Robine, J.-M. (2013). Modal age at death: lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, 11:37–69.

Human Mortality Database (2019). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 15 February 2019).

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., and Yasmeen, F. (2018a). *forecast: Forecasting functions for time series and linear models*. R package version 8.4.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3):1–22.

Hyndman, R. J., Booth, H., Tickle, L., and Maindonald, J. (2018b). *demography: Forecasting Mortality, Fertility, Migration and Population Data*. R package version 1.21.

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

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

Janssen, F. and de Beer, J. (2019). The timing of the transition from mortality compression to mortality delay in Europe, Japan and the United States. *Genus*, 75(1).

Janssen, F., van Wissen, L. J. G., and Kunst, A. E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, 50(4):1341–1362.

Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research*, 3(article 6):doi: 10.4054/DemRes.2000.3.6.

Kannisto, V. (2001). Mode and dispersion of the length of life. *Population: An English Selection*, 13:159–171.

Klein, J. P. and Moeschberger, M. L. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, second edition.

Kostaki, A. (1992). A nine-parameter version of the Heligman-Pollard formula. *Mathematical Population Studies*, 3(4):277–288.

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549.

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

Lexis, W. H. R. A. (1878). Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de Démographie Internationale*, 2(5):447–460.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3):575–594.

Li, N., Lee, R., and Gerland, P. (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051.

Lotka, A. J. (1939). Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particuliere al'espece humaine. *Actualités Scientifiques et Industrielles*, (780).

Mazzuco, S., Scarpa, B., and Zanotto, L. (2018). A mortality model based on a mixture distribution function. *Population Studies*, 72(2):191–200.

Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. In *Compositional Data Analysis Conference*.

Oeppen, J. and Camarda, C. G. (2013). Coherent forecasting of multiple-decrement life tables: compositional models for French Cause of Death data, 1925-2008. In *Joint Eurostat/UNECE Work Session on Demographic Projections*.

Pascariu, M. D. (2018). *MortalityLaws: Parametric Mortality Models, Life Tables and HMD*. R package version 1.7.0.

Pascariu, M. D., Lenart, A., and Canudas-Romo, V. (2019). The maximum entropy mortality model: forecasting mortality using statistical moments. *Scandinavian Actuarial Journal*, pages 1–25.

Pawlowsky-Glahn, V. and Buccianti, A., editors (2011). *Compositional Data Analysis: Theory and Applications*. Wiley.

Pearson, K. (1897). *The Chances of Death, and Other Studies in Evolution*, volume 1. E. Arnold.

Pfaff, B. (2008a). *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.

Pfaff, B. (2008b). VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27(4):1–32.

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

Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, 2nd edition.

Remund, A., Camarda, C. G., and Riffe, T. (2018). A cause-of-death decomposition of young adult excess mortality. *Demography*, 55(3):957–978.

Seligman, B., Greenberg, G., and Tuljapurkar, S. (2016). Convergence in male and female life expectancy: Direction, age pattern, and causes. *Demographic Research*, 34:1063–1074.

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

Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60(4):750–757.

Stoeldraijer, L., van Duin, C., van Wissen, L., and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research*, 29:323–354.

Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the Danish Life Insurance Company of 1871. *Journal of the Institute of Actuaries and Assurance Magazine*, 16(5):313–329.

Thompson, R. and Baker, R. (1981). Composite Link Functions in Generalized Linear Models. *Applied Statistics*, 30(2):125–131.

Vaupel, J. and Lundström, H. (1994). Longer life expectancy? Evidence from Sweden of reductions in mortality rates at advanced ages. In *Studies in the Economics of Aging*, pages 79–102. University of Chicago Press.

Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, pages 253–270.

Whelpton, P. K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association*, 31(195):457–473.

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

Wilmoth, J. R. and Lundström, H. (1996). Extreme longevity in five countries. *European Journal of Population/Revue Européenne de Démographie*, 12(1):63–93.