Modelling and forecasting adult age-at-death distributions

Ugofilippo Basellini & Carlo Giovanni Camarda

Routledge
Taylor & Francis Group

Check for updates

Modelling and forecasting adult age-at-death distributions

Ugofilippo Basellini [1,2] and Carlo Giovanni Camarda[1]
[1]Institut national d'études démographiques (INED), [2]University of Southern Denmark

Age-at-death distributions provide an informative description of the mortality pattern of a population but have generally been neglected for modelling and forecasting mortality. In this paper, we use the distribution of deaths to model and forecast adult mortality. Specifically, we introduce a relational model that relates a fixed 'standard' to a series of observed distributions by a transformation of the age axis. The proposed Segmented Transformation Age-at-death Distributions (STAD) model is parsimonious and efficient: using only three parameters, it captures and disentangles mortality developments in terms of shifting and compression dynamics. Additionally, mortality forecasts can be derived from parameter extrapolation using time-series models. We illustrate our method and compare it with the Lee–Carter model and variants for females in four high-longevity countries. We show that the STAD fits the observed mortality pattern very well, and that its forecasts are more accurate and optimistic than the Lee–Carter variants.

Introduction

The unprecedented rise in life expectancy over the last two centuries is considered to be one of the most remarkable achievements of modern times. Best-practice life expectancy for females has been rising for 160 years at a steady pace of almost three months per year (Oeppen and Vaupel 2002). In current low-mortality and economically developed countries, life expectancy at birth for both sexes has increased from around 30–45 years in the mid-nineteenth century to about 80 years in recent periods (Meslé and Vallin 2011).

These sharp mortality declines triggered a fast acceleration of population growth starting at the beginning of the nineteenth century (Maddison 2006). Nowadays, an ageing process accompanies the rapid increase in world population: virtually every country is experiencing growth in the number and proportion of older persons (United Nations 2015). In the developed world, the large cohort of baby boomers is moving into post-retirement ages, and the baby bust generations and younger cohorts are relatively small, thus creating funding problems for social security and healthcare provision (Booth 2006).

As the challenges linked to population growth and ageing become manifest, innovative models to portray and forecast mortality accurately are increasingly needed to support public policies, guide individual choices, and further scientific research. Mortality modelling has developed into an established research topic since the first half of the eighteenth century (Tabeau 2001). Actuaries have produced forecasts of mortality since at least 1924, in response to the adverse financial effects of mortality improvements on life annuities and pensions (Pollard 1987). However, it is only in the last 30 years that more sophisticated statistical methods for forecasting mortality have been proposed and used (Booth and Tickle 2008).

Three functions are generally used to study human mortality: the hazard, the survival, and the probability density functions. These functions are

complementary: knowing any one of them allows us to derive the other two uniquely (Klein and Moeschberger 2005). Although they describe the same stochastic phenomenon, most mortality analyses and forecasting models are based on age-specific mortality rates (the discrete-time equivalent of the hazard, e.g., Lee and Carter 1992; Currie et al. 2004; Li and Lee 2005; Cairns et al. 2006; Hyndman and Ullah 2007) or on summary measures of the mortality pattern (such as life expectancy at birth, as in Torri and Vaupel 2012; Raftery et al. 2013).

One of the reasons for the predominant use of mortality rates in modelling and forecasting is that they readily represent the change in the risk of death over age and time. The left panel of Figure 1 shows death rates in logarithmic scale for females in Japan in 1960 and 2010 (data obtained from the Human Mortality Database (HMD 2018) and smoothed for illustrative purposes). The three-component pattern of human mortality, theorized at least one-and-a-half centuries ago (Thiele 1871), is clear from the death rates: (i) a decreasing hazard at juvenile ages, capturing infant and childhood mortality; (ii) a subsequent mortality hump at young adult ages, reflecting accidents and external-cause mortality; and (iii) an exponentially increasing hazard (linearly increasing in logarithmic scale) at older ages, capturing senescence. In addition, the downward shift of the mortality curve shows the general decrease in mortality at all ages over the 50-year period.

However, the investigation of mortality rates does not provide an immediate answer to two key questions in mortality studies: (1) how long does a population live on average? And (2) how variable are ages at death? Life expectancy at birth has been the most widely used measure of longevity, although in current low-mortality countries, the adult modal age at death is claimed to be a valid alternative indicator (Kannisto 2001; Horiuchi et al. 2013). However, stimulated by the 'compression–rectangularization' hypothesis proposed by Fries (1980), the variability in age at death has received considerable attention in recent decades as a natural complement to the average duration of life (Myers and Manton 1984; Rothenberg et al. 1991; Wilmoth and Horiuchi 1999; Kannisto 2000; Shkolnikov et al. 2003). This variability is a form of lifespan inequality intrinsic to a population, with higher variation reflecting greater uncertainty in the expected time of death from an individual standpoint and greater heterogeneity in health from a population perspective (Edwards and Tuljapurkar 2005; Edwards 2013; Sasson 2016).

In addition to providing a very informative description of the mortality experience of a population, the age-at-death distribution yields readily available information on the 'central longevity indicators' (mean, median, and modal ages at death; Cheung et al. 2005; Canudas-Romo 2010) as well as lifespan variability. The right panel of Figure 1 shows the life-table age-at-death distributions corresponding to the mortality rates in the left panel. It clearly shows the significant shift to the right in the modal age at death (age at which most deaths occur) since 1960. Moreover, the spread of the distribution provides an indication of lifespan variability. A decrease in variability over time can be observed directly (the distribution in 2010 being taller and more compressed than the one in 1960) 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). Age-at-death distributions thus provide key insights on longevity and lifespan variability that cannot be grasped directly from either mortality rates or the survival function.

Trends in longevity and lifespan variability across countries and time have been extensively investigated in the demographic literature. While almost universal agreement has been reached on the remarkable and sustained increase in longevity over the last two centuries in more economically developed countries, assessments of trends in lifespan variability depend, at least to some extent, on the measure being used (see, e.g., Shkolnikov et al. 2003, pp. 314–18). A comprehensive review of the findings is beyond the scope of this paper; however, in broad terms, increases in longevity are associated with decreasing variability in age at death in long-lived industrialized human societies (Smits and Monden 2009; Vaupel et al. 2011) and across the full range of human experience, for both males and females (Colchero et al. 2016).

Despite being well suited to portraying mortality patterns and studying longevity and lifespan variability, age-at-death distributions have generally been neglected in modelling and forecasting. Only a few attempts have been made to use them explicitly to forecast mortality. Some efforts have been made using compositional data analysis: Oeppen (2008) and Bergeron-Boucher et al. (2017) suggested a modified version of the Lee–Carter model (Lee and Carter 1992), in which the life-table death distribution (instead of the logarithm of death rates) is forecast using principal component analysis. Furthermore, Pascariu et al. (2017) introduced a vector
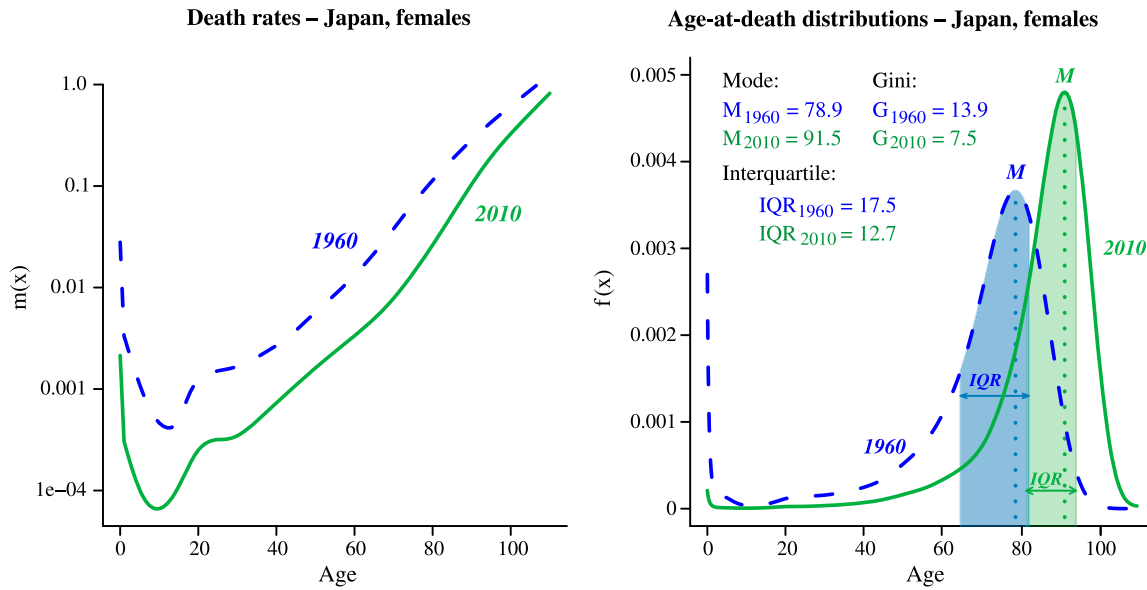
Death rates – Japan, females

Age-at-death distributions – Japan, females

Figure 1 Death rates in logarithmic scale (left panel) and corresponding age-at-death distributions (right panel) for females in Japan, 1960 and 2010

Note: Data have been smoothed for illustrative purposes.

Source: Authors' calculations based on data from the Human Mortality Database (2018).

autoregressive time-series model to forecast a number of statistical moments of the distribution of deaths.

In this paper, we propose a novel method for modelling and forecasting adult mortality that is based on age-at-death distributions. Our model captures mortality dynamics by transforming the age axis of a reference 'standard' distribution that is fixed over time. The transformation function is characterized by three parameters that describe changes in longevity and in lifespan variability, and it takes the form of a segmented linear function. For this reason, we call our model the Segmented Transformation Age-at-death Distributions (STAD) model. The total variability is divided into two components: the variability before (below) and after (above) the modal age at death. The model thus sheds light on the independent contribution of these two age ranges to overall lifespan variability. The shifting and compression dynamics of mortality are captured and disentangled by the three parameters of the model, and mortality forecasts are derived from extrapolation of the parameters using time-series models.

This paper is organized as follows. In the 'Methods' section we review the methods proposed. We start by introducing the STAD model, then show how to choose and compute an appropriate standard distribution. Next, we present the data and smoothing procedure used in our analyses, followed by our methods for estimating and forecasting the STAD parameters. In the 'Results' section, we provide an illustration of

our method by estimating and forecasting the mortality of females in four high-longevity countries, namely Sweden, Japan, France, and Denmark. First, we evaluate the goodness of fit of the STAD model by comparing the estimated mortality pattern with the observed data, both graphically and formally. Second, we assess the accuracy of the STAD forecasts by performing three out-of-sample validation exercises with forecast horizons of 10, 20, and 30 years, respectively. Third, we show the mortality forecasts for the four countries up to 2040. In each section, we use data retrieved from the HMD (2018) and compare our forecasts with those of the Lee–Carter model and its most well-known variants (Shang et al. 2011). Finally, we discuss the results and conclude.

Methods

The STAD model

For ease of presentation, we consider here only two adult age-at-death distributions, defined on the age range 30–110: a standard distribution $f(x)$ and an observed distribution $g(x)$, where x denotes age. The functional form of the distributions has no restrictions, as it can be parametric, non-parametric, or an actual age-at-death distribution. Let $t(x; \boldsymbol{\theta})$ be a transformation function of the age axis and a set of parameters $\boldsymbol{\theta}$, such that the observed distribution

conforms to the standard on the warped axis, that is:

$$g(x) = f[t(x; \boldsymbol{\theta})]. \qquad (1)$$

In other words, the distribution $g(x)$ is derived from altering the age axis of the standard distribution $f(x)$. Our aim is to find a transformation function that describes mortality developments in a parsimonious and rigorous manner. Given this purpose, we propose a transformation function $t(x; \boldsymbol{\theta})$ that depends only on: (i) the changes in modal ages at death between $f(x)$ and $g(x)$; and the differences in the variability of deaths of $f(x)$ and $g(x)$ (ii) before and (iii) after their modal ages.

Formally, let $s = M^g - M^f$ be the difference between the modal ages at death of $g(x)$ and $f(x)$. Then, the transformation function of the proposed STAD model can be expressed as follows:

$$t(x; s, b_L, b_U) = \begin{cases} M^f + b_L(x - s - M^f) & \text{if } x \leq M^g \\ M^f + b_U(x - s - M^f) & \text{if } x > M^g \end{cases}$$
$$(2)$$

where the non-negative coefficients b_L and b_U denote the change in the variability from $f(x)$ to $g(x)$ before and after their modes, respectively. Here, the subscripts L and U stand for lower and upper ages as compared with the mode.

In words, the transformation function $t(x; s, b_L, b_U)$ takes the form of a segmented linear model that breaks at the value of M^g. The slopes of each linear part, b_L and b_U, capture the amount of expansion or reduction in variability of deaths needed before and after M^g, respectively, so that $f(x)$ fits the observed distribution $g(x)$. Changes in the mortality pattern of $g(x)$ with respect to $f(x)$ can thus be concisely described by the three parameters s, b_L, and b_U. These three parameters have a very important demographic interpretation, as they directly capture the shifting and compression dynamics of mortality changes observed during the twentieth century (Fries 1980; Wilmoth and Horiuchi 1999; Bongaarts 2005; Canudas-Romo 2008).

A schematic overview of the STAD model is useful for explaining the segmented linear transformation and the shifting/compression mortality dynamics better. Figure 2 presents the effects of applying the segmented function $t(\cdot)$ to a given standard distribution $f(x)$ (solid line, right panel). When a simple shifted transformation is adopted (dashed line, left panel), the standard distribution is only moved to the right (as shown in the right panel; however, a left shift is also possible), maintaining the same variability (as measured by the variance, if the age range is

shifted too) before and after the modal age at death. This is a special case of equation (2) in which both b_L and b_U are equal to one and the transformation function becomes: $t(x; s) = x - s$. As such, the parameter s regulates the *shifting* dynamic of mortality.

More flexible scenarios can be achieved by modifying the values of b_L and b_U that act jointly with the shifting parameter s. When b_L and b_U are greater than one, variability in ages at death before and after the mode of $g(x)$ decreases with respect to $f(x)$. On the other hand, when b_L and b_U are smaller than one, age-at-death variability increases compared with the standard. These two parameters thus capture the *compression* dynamic of mortality in two age ranges, as they expand or reduce the variability of deaths before and after the mode. In the example in Figure 2, the variability of the segmented distribution (dot-dash line, right panel) is reduced before the mode ($b_L > 1$) and increased above the mode ($b_U < 1$) with respect to the standard.

Interestingly, the STAD model is able to capture two limit scenarios of mortality patterns. On the one hand, the complete rectangularization of the survival curve can be obtained by letting b_L and b_U go to infinity. In this case of maximum compression, the death distribution becomes a probability mass function with probability one at the modal age at death, and lifespan variability is zero, as all individuals die at the same age. On the other hand, the case of maximum expansion is achieved when b_L and b_U are equal to zero. In this case, the death distribution becomes a uniform distribution and lifespan variability is maximized.

Finally, it should be noted that despite introducing a break point at M^g in $t(\cdot)$, there is no disruption at the mode of the transformed distribution. This can be seen in the left panel of Figure 2: although the values of b_L and b_U are different, the segmented linear transformation is continuous at age M^g, which remains the modal age of the segmented distribution.

The standard distribution

The standard distribution of the STAD model plays a very important role: every observed distribution in the time period analysed can be derived from applying the transformation function $t(\cdot)$ with appropriate parameters s, b_L, and b_U to the standard. As such, the standard is a reference distribution that should contain the representative features of the observed distributions. The STAD model can thus be thought of as a relational model (Brass 1971), so choosing a
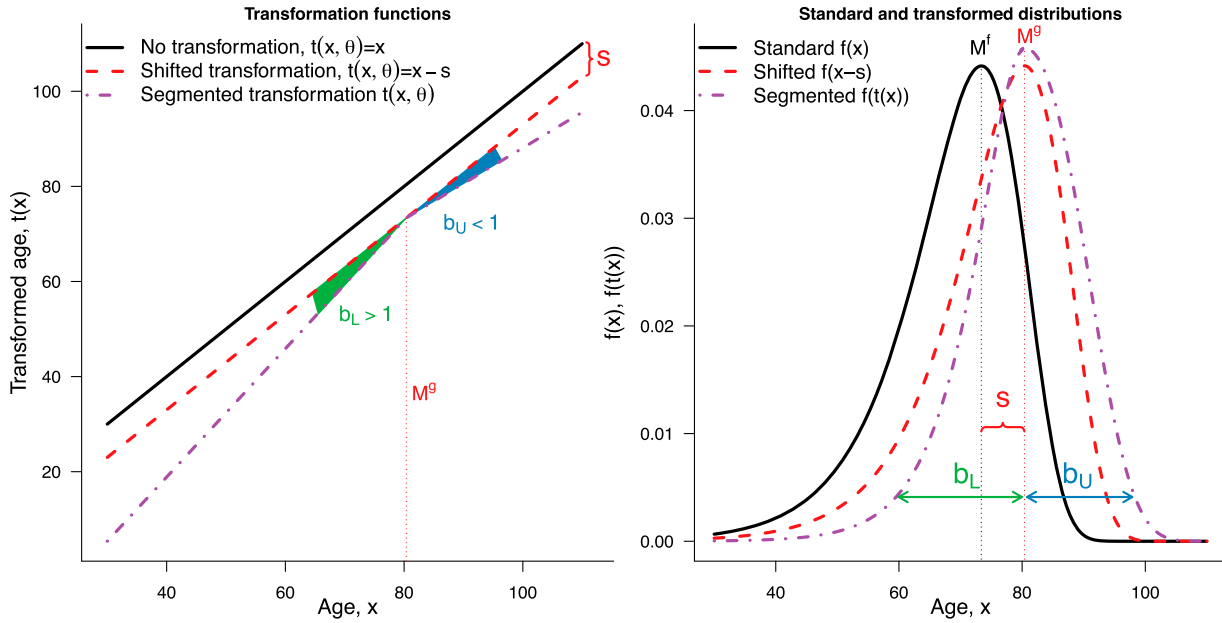
Figure 2   A schematic overview of the effects of transforming the age axis using the STAD model

suitable standard is both desirable and necessary for improving the fit of the model.

A first approach to selecting the standard could be to take a summary measure (e.g., the mean or median) of a series of observed distributions. However, this would lead to a bias, as depicted in the left panel of Figure 3. Let us consider, for simplicity, only two observed distributions: $g_1(x)$ and $g_2(x)$. Taking their simple mean results in a standard $f(x)$ with a fairly unrealistic shape that represents neither the mortality age pattern in $g_1(x)$ nor that in $g_2(x)$. The reason for this behaviour is that $f(x)$ conflates age segments that are positioned before and after the mode in the two observed distributions: the decreasing part of $g_1(x)$ after its mode is averaged with the increasing part of $g_2(x)$ before its mode.

To avoid this problem, we first transform the observed densities to line up their qualitative features. Specifically, we shift the densities so that their modal ages at death are equal to the mode of the first observed density, maintaining their variability unchanged. This procedure is known as 'landmark registration' and is commonly used in functional data analysis (Ramsay and Silverman 2005) with the aim of 'eliminating uninteresting differences in functions so that the remaining functional variation is (more) completely concerned with the differences of interest' (Clarkson et al. 2005, p. 88). Here, by aligning the distributions, we remove the differences in the mode (that are already captured by the parameter s) and focus on the differences in their variability.

Having aligned the observed distributions, the standard distribution is then computed as the mean

of the aligned distributions. In the right panel of Figure 3, $g_2(x)$ is shifted to the left such that its mode aligns with that of $g_1(x)$, while $f(x)$ is computed as the mean of the observed $g_1(x)$ and the aligned $\breve{g}_2(x)$. This registered mean provides a better representation of the two densities: the shape of $f(x)$ is more realistic, its mode is equal to those of $g_1(x)$ and $\breve{g}_2(x)$, and its variability is the average variability of the two distributions.

Finally, for capturing any type of shifting, we need to extend the age support of the standard distribution, $f(x)$. For example, in the simple shift case in Figure 2, the value of the shifted distribution (dashed line) at age 30 is equal to the value of the expanded standard (solid line) evaluated at $x - s$, that is, beyond the original 30–110 age range. We can carry out this extension since we assume that the senescent (older adult) component of mortality is independent from the juvenile and younger adult ones. We achieve this task by means of non-parametric techniques, by extrapolating the mortality age pattern of $f(x)$ without any rigid assumption on its form.

Data and estimation of the standard distribution

We apply the STAD model to the population of females in four high-longevity countries, namely Sweden, Japan, France, and Denmark. Sweden was selected as a benchmark population for its high standard in data quality even at the most advanced ages
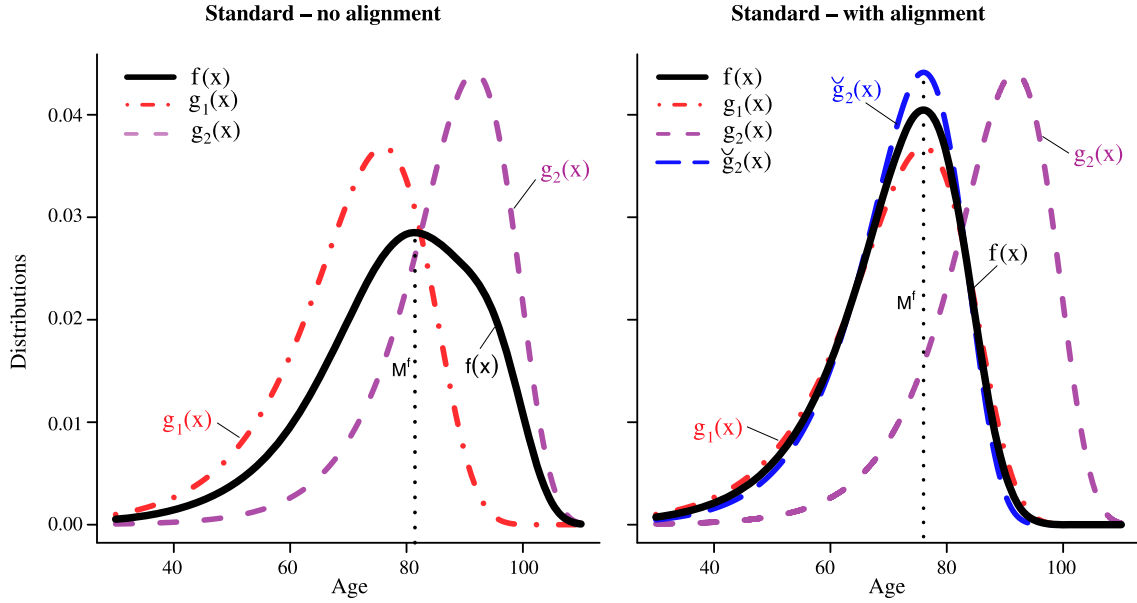
Standard – no alignment   **Standard – with alignment**

Figure 3 Standard distributions $f(x)$ computed as mean of the observed densities $g_1(x)$ and $g_2(x)$ (left panel) and as mean of the observed $g_1(x)$ and aligned $\breve{g}_2(x)$ (right panel)
Note: Black solid lines indicate the standard distribution.

(Vaupel and Lundström 1994; Wilmoth and Lundström 1996), while Japan and France represent populations with very high average lifespans (Wilmoth and Lundström 1996; Oeppen and Vaupel 2002; Cheung and Robine 2007; Canudas-Romo 2008, 2010; Horiuchi et al. 2013). Finally, Denmark shows a rather peculiar mortality trajectory: despite being among the highest longevity countries until the 1980s, it experienced a stagnation of life expectancy in the last two decades of the twentieth century (Christensen et al. 2010; Lindahl-Jacobsen, Oeppen, et al. 2016; Lindahl-Jacobsen, Rau, et al. 2016). To test the fitting and forecasting performance of our STAD model, we apply it to these diverse mortality developments.

As already mentioned, our interest is restricted to the senescent component of mortality, thus we start our analyses from age 30. Furthermore, we aim to forecast mortality up to 2040, so choose 1980–2014 as the fitting period. This allows us to study the most recent mortality developments (excluding improvements in the 1970s due to the cardiovascular revolution; Vallin and Meslé 2004) while keeping the length of the fitting period long enough to forecast 26 years ahead.

Data are obtained from the HMD (2018) and consist of observed death counts, $D_{x,y}$, and exposure to risk, $E_{x,y}$, classified by age at death ($x = 30, \ldots, 110+$) and year of death ($y = 1980, \ldots, 2014$). Moreover, in the following, we assume that the number of deaths at a given age x and year y follows a Poisson

distribution (Brillinger 1986):

$$D_{x,y} \sim \mathcal{P}(E_{x,y}\, \mu_{x,y}) \qquad (3)$$

where $\mu_{x,y}$ is the hazard function, whose estimation is the object of any mortality model.

As presented earlier, the STAD model is expressed in continuous notation. To use the model with observed discrete mortality data, a non-parametric smoothing approach is useful, so that mortality can be evaluated at any finer scale of the age axis, practically at a continuous level. In addition, the smoothing procedure avoids imposing rigid parametric mortality structures that could result in misleading outcomes.

The first step in the STAD model consists of extracting the standard distribution as the mean of the aligned observed distributions with respect to the earliest modal age at death (1980 in our case). To perform this task, we need to compute age-at-death distributions and associated modes for each year. Instead of modelling the life-table deaths directly, we make use of the assumption in equation (3) and smooth actual death counts using a P-splines approach and exposure as an offset (Eilers and Marx 1996; Camarda 2012). The advantages of this approach are twofold. First, we use an objective selection of the amount of smoothing based on the Bayesian information criterion (BIC; Schwarz 1978, see Appendix A). This approach accounts for the size of the population analysed due to the Poisson assumption of death counts (Brillinger 1986). In

contrast, smoothing life-table deaths is more subjective, because the optimal degree of smoothing depends on the life-table radix: given the same age-specific mortality pattern, higher starting radices are associated with a lower amount of smoothing. Second, we smooth observed rather than life-table deaths, because the latter are constructed from adjusted death rates, as the HMD uses a logistic function for ages 80 and above (Wilmoth et al. 2007).

Having smoothed mortality, we can derive the corresponding smooth age-at-death distributions: $g(x) = \mu(x)\, l(x)$, where the survival function $l(x)$ can be computed numerically from the smoothed death rates, evaluated at an extremely fine level. As shown in Ouellette and Bourbeau (2011), this procedure further allows us to extract modal ages at death directly for each year, M^y. Given the series of M^y, we already have the estimate of the shifting parameter for each year, that is, the difference in modal ages at death between a given year and the earliest year, $s_y = M^y - M^{1980}$.

Using estimates from the *P*-splines approach, we can express the smooth observed distribution for year y as follows:

$$g_y(x) = \exp[\boldsymbol{B}(x)\,\boldsymbol{\beta}_y] \qquad (4)$$

where $\boldsymbol{B}(x)$ is a basis of equally spaced *B*-splines over ages x and $\boldsymbol{\beta}_y$ are coefficients specific to the distribution $g_y(x)$. The estimated series of s_y allow us to align all distributions by simply re-evaluating the *B*-splines basis in equation (4) on the shifted age axes:

$$\breve{g}_y(x) = \exp[\boldsymbol{B}(x - s_y)\,\boldsymbol{\beta}_y]. \qquad (5)$$

The standard distribution $f(x)$ can then be obtained by averaging the aligned observed distributions. Again, the standard distribution can be expressed as a linear combination of *B*-splines over x and new coefficients $\boldsymbol{\beta}_f$ specific to $f(x)$:

$$f(x) = \exp[\boldsymbol{B}(x)\,\boldsymbol{\beta}_f]. \qquad (6)$$

Note that the coefficients $\boldsymbol{\beta}_f$ contain all the representative age-specific mortality features that we will use for describing mortality developments over time, except shifting and compression/expansion that are captured by the parameters s, b_L, and b_U.

Estimation and forecast of the STAD parameters

Given the estimated standard distribution $f(x)$, the model parameters can then be estimated. For each year y, we seek to find a set of three parameters

$[s_y,\ b_{L,y},\ b_{U,y}]$ that produces a transformation of the age axis defined in equation (2) such that the mortality pattern of the transformed standard $f(t(x; s_y, b_{L,y}, b_{U,y}))$ fits the observed data.

Whereas the shifting parameters, $\hat{s}$, are readily estimated within the alignment procedure, estimation of the compression/expansion parameters is achieved by maximum likelihood. Based on the Poisson assumption in equation (3), the parameters $\hat{b}_L$ and $\hat{b}_U$ are obtained by maximizing the following log-likelihood:

$$\begin{aligned} &\ln \mathcal{L}(b_{L,y},\ b_{U,y} | D_{x,y}, E_{x,y}, \hat{s}_y, \boldsymbol{\beta}_f) \\ &\propto \sum_x \{ D_{x,y} \ln (\mu_{x,y}^{\text{STAD}}) - E_{x,y}\, \mu_{x,y}^{\text{STAD}} \} \end{aligned} \qquad (7)$$

for $y = 1980, \ldots, 2014$, where $\mu_{x,y}^{\text{STAD}}$ denotes the hazard function corresponding to the estimated segmented distribution of the STAD model.

On the one hand, data and log-likelihood are expressed in terms of deaths, exposures, and hazard. On the other hand, the STAD model is described in terms of distributions. This is not an issue because, as already noted in the 'Introduction', hazard and distributions are complementary: knowing one of them allows us to derive the other directly and uniquely. As such, the estimated $\hat{b}_L$ and $\hat{b}_U$ are the parameters that produce a transformed segmented distribution $f[\hat{t}(\cdot)]$ whose corresponding hazard μ^{STAD} maximizes equation (7).

A general-purpose numerical optimizer works well for maximizing equation (7) and estimating $\hat{b}_L$ and $\hat{b}_U$. Routines for estimating the standard distribution and the transformation function for the age axis based on maximum likelihood were implemented in R (R Development Core Team 2018) and are available in the supplementary material.

Having estimated the parameters over the observed years, it is possible to model their time series: mortality forecasts are then obtained by extrapolation of the STAD parameters and the time-fixed standard distribution. In other words, we assume that mortality developments can be approximated well by a standard age pattern and time indices capturing mortality dynamics in shifting and compression/expansion.

Among the numerous approaches available for time-series modelling and forecasting, we combine univariate and multivariate models to describe and forecast the time evolution of the three parameters. Since the shifting and compression/expansion dynamics act independently in the STAD model, that is, changes in s do not affect b_L or b_U, and vice versa (as shown in Figure 2), we treat s separately from the two b parameters. For s, we choose the

standard class of univariate autoregressive integrated moving average models, ARIMA(p, d, q). The three parameters of the ARIMA are chosen from different combinations, using an automatic stepwise algorithm based on minimization of the corrected Akaike Information Criterion (the package default, see Hyndman and Khandakar 2008).

Changes between the two b parameters could instead be correlated: for example, if a mortality change would maintain the same overall lifespan variability but compress the ages before the mode, then b_U would necessarily need to expand (or vice versa). Therefore, we favour a multivariate approach to model b_L and b_U, as this takes into account the interrelationship *within* and *between* the two series. Consequently, for the b parameters we use a vector autoregressive (VAR) model, which is the multivariate generalization of univariate autoregressive time-series models (Hamilton 1994). We perform model selection, estimation, and diagnostic tests using the R package 'vars' (Pfaff 2008a, 2008b). Forecasts of all three parameters are then computed with the 'forecast' package (Hyndman et al. 2018).

Results

Observed vs. fitted data

We fit the STAD model to the populations of females in Sweden, Japan, France, and Denmark using the data described earlier. The estimated standard distribution and parameters $\hat{s}$, $\hat{b}_L$, and $\hat{b}_U$ allow us to derive the mortality pattern over time in each country. To assess the STAD model's goodness of fit, we investigate two summary measures of the mortality pattern that answer the questions outlined in the 'Introduction': remaining life expectancy at age 30 (e_{30}, question 1) and the Gini coefficient at age 30 (G_{30}, question 2). The first measure is computed using standard demographic methods (Preston et al. 2001), while the second is derived from the formula in Shkolnikov et al. (2003).

Figure 4 shows the observed and fitted e_{30} and G_{30} for females in the four countries during 1980–2014. The graphs show that the STAD model performs very well in terms of goodness of fit. In the upper panels, the actual and fitted e_{30} are always very close to each other. Similarly, the goodness of fit for G_{30} (lower panels) is satisfactory, with small discrepancies in the earlier years for Denmark and in the latest years for Japan and France.

To perform a more formal statistical assessment of the goodness of fit of the STAD model beyond this graphical investigation, we compare the fit of the STAD with the standard Lee–Carter (LC) model (Lee and Carter 1992) and its variants, estimated on the same ages and years. We use the BIC (Schwarz 1978) to measure model differences in terms of trade-off between model parsimony and accuracy (see Appendix A for additional details and computational procedure).

In the following, we consider all the variants of the LC model that can be readily estimated with available packages in R. These models are referred to here as LM (Lee and Miller 2001), BMS (Booth et al. 2002), BDV (Brouhns et al. 2002), HU (Hyndman and Ullah 2007), HUrob (robust Hyndman and Ullah 2007), and HUw (weighted Hyndman and Ullah 2007); for a concise review of these models, see Shang et al. 2011. The LC variants were fitted using the R packages 'demography' and 'StMoMo' (Hyndman et al. 2017; Villegas et al. 2017). For the HU, HUrob, and HUw models, we chose the 'demography' package default value of six principal components, as suggested by Hyndman and Booth (2008) and used by Shang et al. (2011).

Table 1 shows the BIC values of the STAD, LC, and LC variant models in the four countries (values of the deviance and effective dimension are reported in Appendix B). The STAD is chosen over the other models for Sweden and Denmark. For Japan, the standard HU model is the best fitting, while the Poisson variant of the LC model (BDV) is preferred for France.

Out-of-sample validation

Before projecting mortality into the future, we first assess the accuracy of the STAD forecasts by out-of-sample validation. Specifically, we compare the observed vs. the point and interval forecasts of remaining life expectancy at age 30 (e_{30}), the Gini coefficient at age 30 (G_{30}), and the logged age-specific death rates over all ages and years ($\ln(m_{x,t})$). Lifespan disparity measures such as the Gini index are indeed important indicators for evaluating mortality forecasts (Bohk-Ewald et al. 2017).

The procedure of dividing the observed data set into two parts, the first for fitting the model and the second for checking predictions, is very common in forecasting (Chatfield 2000). Here, we perform three out-of-sample validation exercises: we fit the STAD, LC, and LC variant models to three periods of equal length: 1970–2004, 1960–94, and 1950–84. Then, we forecast mortality up to 2014 in all scenarios, corresponding to three forecast horizons of 10, 20, and 30 years, respectively. We then compare
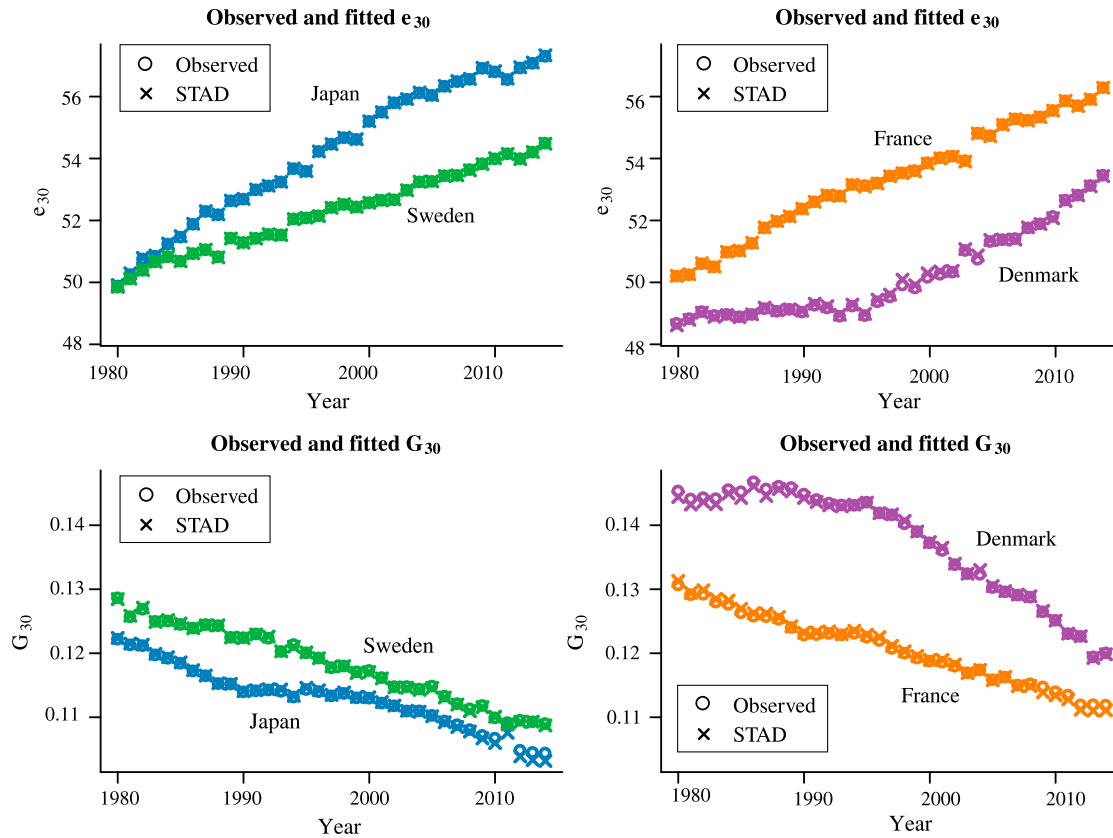
Figure 4　Comparison of observed and STAD-fitted remaining life expectancy at age 30 (e_{30}, upper panels) and Gini coefficient at age 30 (G_{30}, lower panels) for females in Sweden, Japan, France, and Denmark, 1980–2014
Source: As for Figure 1.

the point and interval forecasts of e_{30}, G_{30}, and $\ln(m_{x,t})$ at each forecast year with the observed data.

For the point forecast analysis, we use the mean absolute error (MAE) as our measure of forecast accuracy, following standard time-series practice (Chatfield 2000). In addition, we only compare the forecast accuracy of the STAD model with the LC and first three LC variant models: we do not consider the functional data models (HU, HUrob, or HUw) because their better performance is partially due to the much higher number of parameters used (about four and five times more parameters, respectively, than the LC variants and the STAD; see effective dimension in Appendix B).

Table 2 shows the point forecast accuracy as measured by the MAE of e_{30}, G_{30}, and $\ln(m_{x,t})$ for the STAD, LC, and three LC variant models in the three out-of-sample scenarios for each of the four countries. The results of this analysis show that the STAD forecasts are more accurate overall than the LC ones. Out of 36 indicators, the STAD model is the most accurate almost half of the time (16 indicators). The LM is the second-best performer (nine indicators), followed by the BMS (six indicators),

the BDV (four indicators), and the original LC model (one indicator).

In addition to analysing point forecasts, we further evaluate the accuracy of interval forecasts for e_{30}, G_{30}, and $\ln(m_{x,t})$. Prediction intervals are a valuable tool for assessing the probabilistic uncertainty of point forecasts (Shang et al. 2011), and interval forecasts are very important as they allow, among other things, a more thorough comparison of different forecasting methods (Chatfield 2000).

For the LC and variants, we derived the 80 per cent prediction intervals from the uncertainty of the innovations in the 'random walk with drift' model for the time index κ_t; in addition, we considered the uncertainty of the drift parameter for the BMS method. For the STAD model, we derived the 80 per cent prediction intervals from a bootstrapping procedure (Efron and Tibshirani 1994): we computed 1,000 simulations of the time series of the three parameters, and for each simulation calculated the mortality pattern in each forecast year (age-specific death rates and associated summary measures). From these simulations, we took the median as the central forecast and the

Table 1 BIC values of the STAD, LC, and LC variant models for females aged 30–110+ in four countries, 1980–2014

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	**4,276**	4,699	4,703	4,674	4,611	8,871	9,040	9,018
Japan	11,053	11,475	11,818	11,387	10,790	**10,057**	10,694	10,981
France	9,141	7,822	7,951	7,773	**7,663**	9,402	10,132	9,858
Denmark	**5,332**	5,613	5,604	5,587	5,485	9,020	9,397	9,890

Note: Lower values of the BIC (lowest in bold) correspond to a better model.
Source: Authors' calculations based on data from the Human Mortality Database (2018).

Table 2 Mean absolute error of the STAD, LC, and LC variant forecasts of e_{30}, G_{30}, and In $(m_{x,t})$ for females in four countries and three out-of-sample validation scenarios: forecast horizons of 10 years (fitting period 1970–2004), 20 years (1960–94), and 30 years (1950–84)

Country	Test period	Measure	Model				
			STAD	LC	LM	BMS	BDV
Sweden	10 years	e_{30}	0.11	0.11	0.09	0.08	**0.08**
		G_{30}	0.17	**0.06**	0.06	0.06	0.07
		$\ln(m_{x,t})$	**0.10**	0.12	0.12	0.12	0.11
	20 years	e_{30}	**0.25**	0.41	0.38	0.41	0.38
		G_{30}	0.41	0.15	0.21	**0.15**	0.18
		$\ln(m_{x,t})$	**0.14**	0.16	0.18	0.16	0.16
	30 years	e_{30}	0.87	0.81	0.80	0.80	**0.78**
		G_{30}	0.21	0.16	**0.07**	0.16	0.12
		$\ln(m_{x,t})$	**0.15**	0.19	0.18	0.19	0.18
Japan	10 years	e_{30}	0.94	1.00	0.86	**0.78**	0.98
		G_{30}	0.11	0.55	0.14	**0.07**	0.40
		$\ln(m_{x,t})$	0.12	0.21	0.12	**0.11**	0.15
	20 years	e_{30}	0.86	0.35	0.35	**0.31**	0.33
		G_{30}	**0.74**	1.10	0.78	1.07	1.04
		$\ln(m_{x,t})$	**0.17**	0.25	0.20	0.25	0.22
	30 years	e_{30}	1.18	0.28	**0.23**	0.45	0.32
		G_{30}	**1.06**	1.68	1.25	1.56	1.40
		$\ln(m_{x,t})$	**0.22**	0.46	0.31	0.44	0.40
France	10 years	e_{30}	**0.13**	0.44	0.30	0.35	0.36
		G_{30}	**0.08**	0.42	0.10	0.19	0.35
		$\ln(m_{x,t})$	**0.07**	0.12	0.07	0.10	0.11
	20 years	e_{30}	0.70	0.55	**0.33**	0.47	0.47
		G_{30}	0.37	0.58	**0.17**	0.54	0.51
		$\ln(m_{x,t})$	0.13	0.16	0.13	0.16	**0.12**
	30 years	e_{30}	**0.21**	0.38	0.36	0.38	0.38
		G_{30}	**0.34**	0.45	0.36	0.45	0.45
		$\ln(m_{x,t})$	**0.12**	0.16	0.15	0.16	0.15
Denmark	10 years	e_{30}	**0.75**	1.26	1.01	1.57	1.45
		G_{30}	0.75	1.01	**0.55**	1.12	1.15
		$\ln(m_{x,t})$	**0.16**	0.20	0.18	0.22	0.21
	20 years	e_{30}	1.58	1.06	**1.01**	1.10	1.12
		G_{30}	1.92	1.39	**1.28**	1.39	1.43
		$\ln(m_{x,t})$	0.29	0.26	**0.23**	0.25	0.25
	30 years	e_{30}	0.66	0.72	**0.64**	0.74	0.73
		G_{30}	1.21	0.72	0.82	0.72	**0.71**
		$\ln(m_{x,t})$	0.27	0.26	0.28	**0.24**	0.25

Notes: Lower values (lowest in bold) correspond to greater point forecast accuracy. In cases of equal value, the ranking was decided from the third decimal place.
Source: As for Table 1.

lowest and highest deciles to construct the 80 per cent prediction intervals.

Specifically, we computed one-step-ahead 80 per cent prediction intervals for each year (and age for $\ln(m_{x,t})$) in the forecasting period of the three out-of-sample scenarios. Following Shang et al. (2011), we calculated the 'coverage probability deviance' as the absolute difference between 0.8 (the nominal coverage probability) and the empirical coverage probability. The latter is given by the actual proportion of out-of-sample data that falls within the calculated prediction intervals. The lower the coverage probability, the more accurate the prediction intervals of a model; furthermore, with a 0.8 nominal coverage probability, the deviance can vary from 0.0 to 0.8.

Table 3 shows the coverage probability deviances of e_{30}, G_{30}, and $\ln(m_{x,t})$ for the STAD, LC, and three LC variant models in the three out-of-sample scenarios for each of the four countries. Here, the STAD model is the third-best performer. The LM outperforms the other models, as it produces more accurate prediction intervals for ten indicators; the BMS, STAD, and LC models follow with six, five, and four indicators, respectively, while the BDV model is most precise for only two. For the remaining nine indicators there was a perfect draw among two or more models.

Forecasts to 2040

We now present the mortality forecasts of the STAD model from 2015 to 2040 using the fitting period 1980–2014, and compare the forecasts with two LC variants that performed best in the previous sections: the HU model (lowest BIC for Japan) and the LM model (most accurate forecasts after the STAD).

Figure 5 shows the estimated and forecast STAD parameters with 80 per cent prediction intervals for females in Sweden, Japan, France, and Denmark during 1980–2040. In all four countries, the observed shifting dynamic captured by the *s* parameter (upper panels) is forecast to continue, with the modal age at death increasing by about 1.3 years per decade in Sweden, 2.1 in Japan, 1.7 in France, and 1.1 in Denmark (increasing values of *s* correspond to increases in the mode). The lifespan variability before and after the modal age at death (lower panels) is forecast to decrease in all countries (increasing values of *b* correspond to greater compression), with b_U increasing at a faster pace than b_L in all countries except Denmark.

Figure 6 shows the observed and forecast remaining life expectancies at age 30 (e_{30}) and Gini coefficients at

age 30 (G_{30}) with 80 per cent prediction intervals in the four countries for 1980–2040, together with the forecasts of the LM and HU models. The graphs show that the STAD forecasts of e_{30} reflect the past linear increase over the years 1980–2014 well and that they are generally more optimistic than the LM and HU forecasts. In all countries, the slope of the future increase in e_{30} of the STAD is larger than those of the LM and HU models. In Japan and France, the initial forecasts of the STAD are lower than the LC ones, but the steeper rate of increase of the STAD results in higher forecasts by 2040. With respect to G_{30}, the STAD forecasts show a slower compression of mortality than the LM and HU ones in Sweden, France, and most noticeably Japan, while a faster compression is forecast in Denmark.

In Figure 7, we compare the STAD, LM, and HU age-specific mortality rate forecasts for 2040 for the four countries. Several differences emerge between the models from this age pattern analysis. Mortality rates from the STAD model are smoother than the LM ones, as they are derived from a smooth standard distribution: this could be particularly advantageous for long-term projections. Discontinuities or jaggedness in the forecast age profile of mortality will be absent in the STAD forecasts, unlike in the LM model (Li et al. 2013). Importantly, the STAD forecasts do not display the fairly unrealistic S-shape that characterizes the LM and HU forecasts for Sweden, Japan, and France. In addition, the HU forecast for Denmark displays an even more undulating pattern.

Finally, Figure 8 shows the actual life-table deaths in 1980 and 2014, alongside the age-at-death distributions in 2040 as computed from the STAD, LM, and HU forecasts. The shifting mortality dynamic is more important and visible in the STAD model: in fact, the STAD age-at-death distribution forecast is more shifted and less compressed than the LM and HU ones in all countries except Denmark (where the three forecasts are more similar).

Discussion and conclusion

Age-at-death distributions provide a very informative description of the mortality pattern of a population and are well suited for studying two important indicators of population health: longevity and lifespan variation. Despite these advantages, they have generally been neglected in modelling and forecasting human mortality, as most techniques are based on age-specific mortality rates.

In this paper, we have proposed a novel method for modelling and forecasting adult mortality based

Table 3 Coverage probability deviances of the STAD, LC, and LC variant forecasts of e_{30}, G_{30}, and In $(m_{x,t})$ for females in four countries and three out-of-sample validation scenarios: forecast horizons of 10 years (fitting period 1970–2004), 20 years (1960–94), and 30 years (1950–84)

Country	Test period	Measure	Model STAD	LC	LM	BMS	BDV
Sweden	10 years	e_{30}	0.20	0.20	0.20	0.20	0.20
		G_{30}	0.70	0.10	0.10	0.10	0.20
		$\ln(m_{x,t})$	0.43	**0.28**	0.41	0.31	0.33
	20 years	e_{30}	0.20	0.20	0.20	0.20	0.20
		G_{30}	0.75	0.20	**0.10**	0.20	0.15
		$\ln(m_{x,t})$	0.41	0.32	0.43	**0.30**	0.37
	30 years	e_{30}	0.73	0.03	0.07	0.13	0.03
		G_{30}	0.43	0.03	0.20	0.03	**0.00**
		$\ln(m_{x,t})$	0.57	0.41	0.45	**0.35**	0.42
Japan	10 years	e_{30}	0.80	0.80	0.50	**0.40**	0.80
		G_{30}	0.30	0.80	0.10	0.10	0.80
		$\ln(m_{x,t})$	0.66	0.67	0.54	**0.40**	0.68
	20 years	e_{30}	0.50	0.10	**0.00**	0.15	0.10
		G_{30}	0.80	0.80	0.80	0.80	0.80
		$\ln(m_{x,t})$	0.61	0.61	**0.51**	0.60	0.60
	30 years	e_{30}	**0.07**	0.20	0.20	0.20	0.20
		G_{30}	0.67	0.80	**0.60**	0.80	0.80
		$\ln(m_{x,t})$	**0.48**	0.65	0.53	0.65	0.65
France	10 years	e_{30}	0.10	0.10	0.10	0.20	0.10
		G_{30}	0.40	0.80	0.20	**0.10**	0.70
		$\ln(m_{x,t})$	0.44	0.35	**0.32**	0.33	0.36
	20 years	e_{30}	0.60	**0.15**	0.20	0.20	0.20
		G_{30}	0.80	0.70	**0.20**	0.30	0.60
		$\ln(m_{x,t})$	0.42	0.29	**0.20**	0.23	0.25
	30 years	e_{30}	**0.00**	0.20	0.20	0.20	0.20
		G_{30}	0.60	0.20	0.20	0.20	0.20
		$\ln(m_{x,t})$	0.51	0.03	0.02	0.07	**0.01**
Denmark	10 years	e_{30}	**0.20**	0.70	0.50	0.80	0.80
		G_{30}	0.80	0.80	0.80	0.80	0.80
		$\ln(m_{x,t})$	0.42	**0.35**	0.38	0.43	0.48
	20 years	e_{30}	0.45	0.40	**0.30**	0.40	0.40
		G_{30}	0.80	0.80	**0.75**	0.80	0.80
		$\ln(m_{x,t})$	0.43	0.42	**0.39**	0.43	0.43
	30 years	e_{30}	0.23	**0.03**	0.13	0.07	0.07
		G_{30}	**0.57**	0.73	0.70	0.73	0.73
		$\ln(m_{x,t})$	0.63	0.43	0.46	**0.38**	0.45

Notes: Lower values (lowest in bold) correspond to greater interval forecast accuracy. In cases of equal value, the ranking was decided from the third decimal place (where possible).
Source: As for Table 1.

on age-at-death distributions. To our knowledge, this is one of the very first attempts in this direction. Specifically, we introduced a segmented linear transformation model that captures mortality dynamics over time from changes in modal ages at death and variability of the distributions with respect to a fixed standard. This approach is both parsimonious and revealing: using only three parameters, it enables us to portray and forecast mortality developments, capture the compression and shifting dynamics of mortality, and construct life-table functions. Given its features, we called our suggested

model the Segmented Transformation Age-at-death Distributions (STAD) model.

The results indicate that the STAD model performs very well in terms of goodness of fit: estimated remaining life expectancies (e_{30}) and Gini coefficients (G_{30}) for females in four high-longevity countries (Sweden, Japan, France, and Denmark) are very close to their observed historical values. Moreover, the STAD forecasts of e_{30}, G_{30}, and the logged age-specific death rates are more accurate than the LC model (Lee and Carter 1992) and its variants in three out-of-sample validation scenarios with forecast horizons of 10, 20,
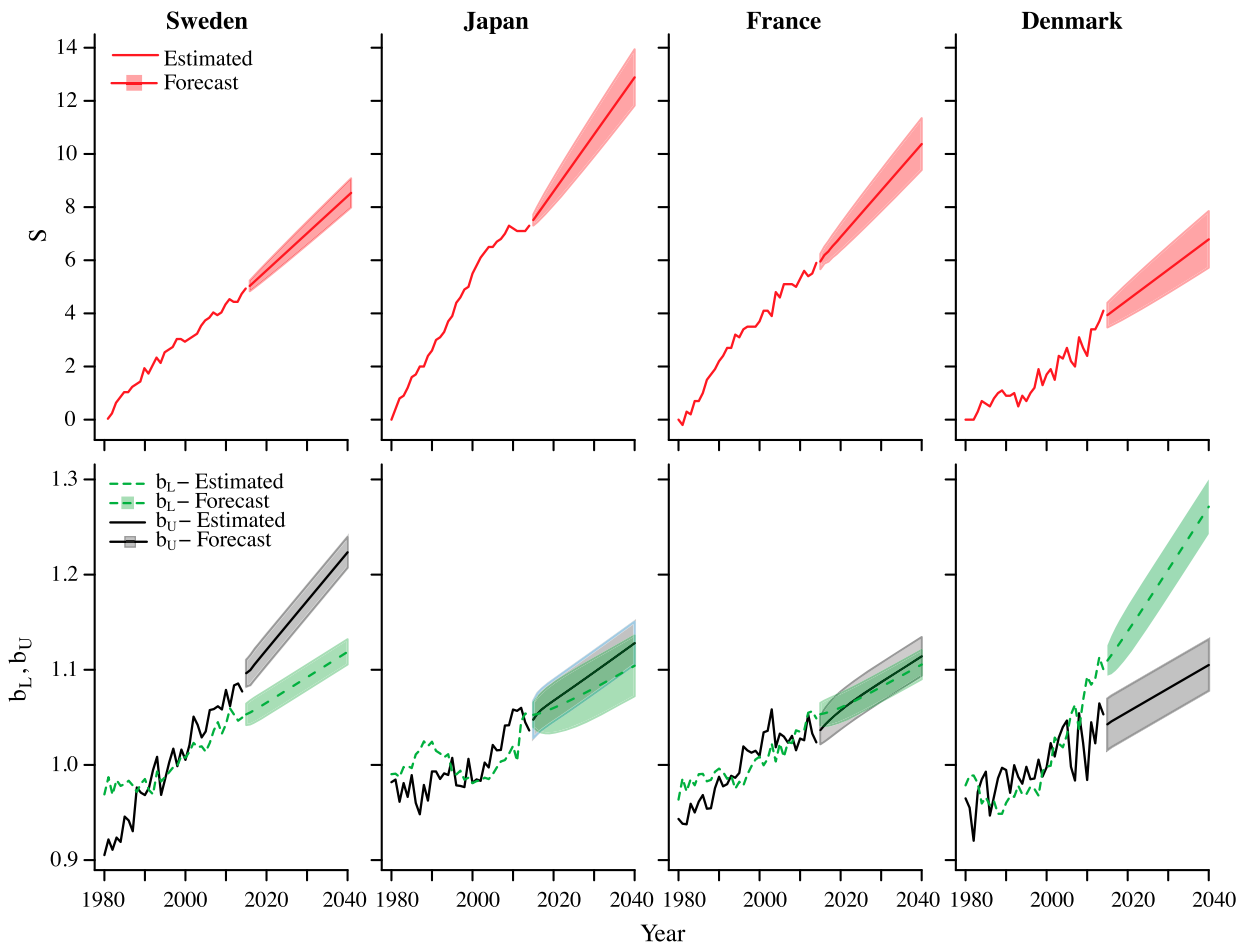
Figure 5 Estimated and forecast STAD parameters: s (upper panels), b_L, and b_U (lower panels) with 80 per cent prediction intervals for females in Sweden, Japan, France, and Denmark, 1980–2040
Source: As for Figure 1.

and 30 years. The STAD forecasts of e_{30} for the four countries up to 2040 reflect the past linear increase observed between 1980 and 2014 well, and are more optimistic than the LC variant forecasts, which have often underpredicted future gains in life expectancy (Lee and Miller 2001). Finally, the STAD forecasts are characterized by greater shifting and smaller compression than the LC ones.

Theoretically, the STAD model can be viewed as a relational model (Brass 1971). Here, transformations of a standard age-at-death distribution $f(x)$ describe mortality developments *over time* rather than across countries. From this perspective, an analogy can be made between the STAD and LC models. The series of α_x in the LC model can be viewed as a standard mortality age profile, whose development over time is captured by the index κ_t modulated by a fixed rate of mortality improvement at age x, β_x. As such, the LC model can also be interpreted as a relational model.

Our method is also inspired by two other models: one proposed by Camarda et al. (2008) for analysing

mortality developments from age-at-death distributions by transforming the age axis with a smooth non-linear warping function. Second, Cadena and Denuit (2016) recently introduced an accelerated hazard relational model based on a smooth transformation of the age scale free of monotonic constraints and used it to estimate and project mortality in Belgium. The STAD model has elements that are common to both approaches but is also significantly different. The first model was developed for mortality analyses and the smooth transformation function is not well suited for forecasting. The second model is based on the hazard function and is subject to the same characteristics as the models based on mortality rates discussed in the 'Introduction'.

As in most relational models, the choice of standard is important: all the features contained in $f(x)$ are carried in the transformation function $t(\cdot)$. To maximize the representativeness of the standard with respect to the observed densities, we chose $f(x)$ as the mean of the aligned distributions, using a landmark registration approach inspired from functional data
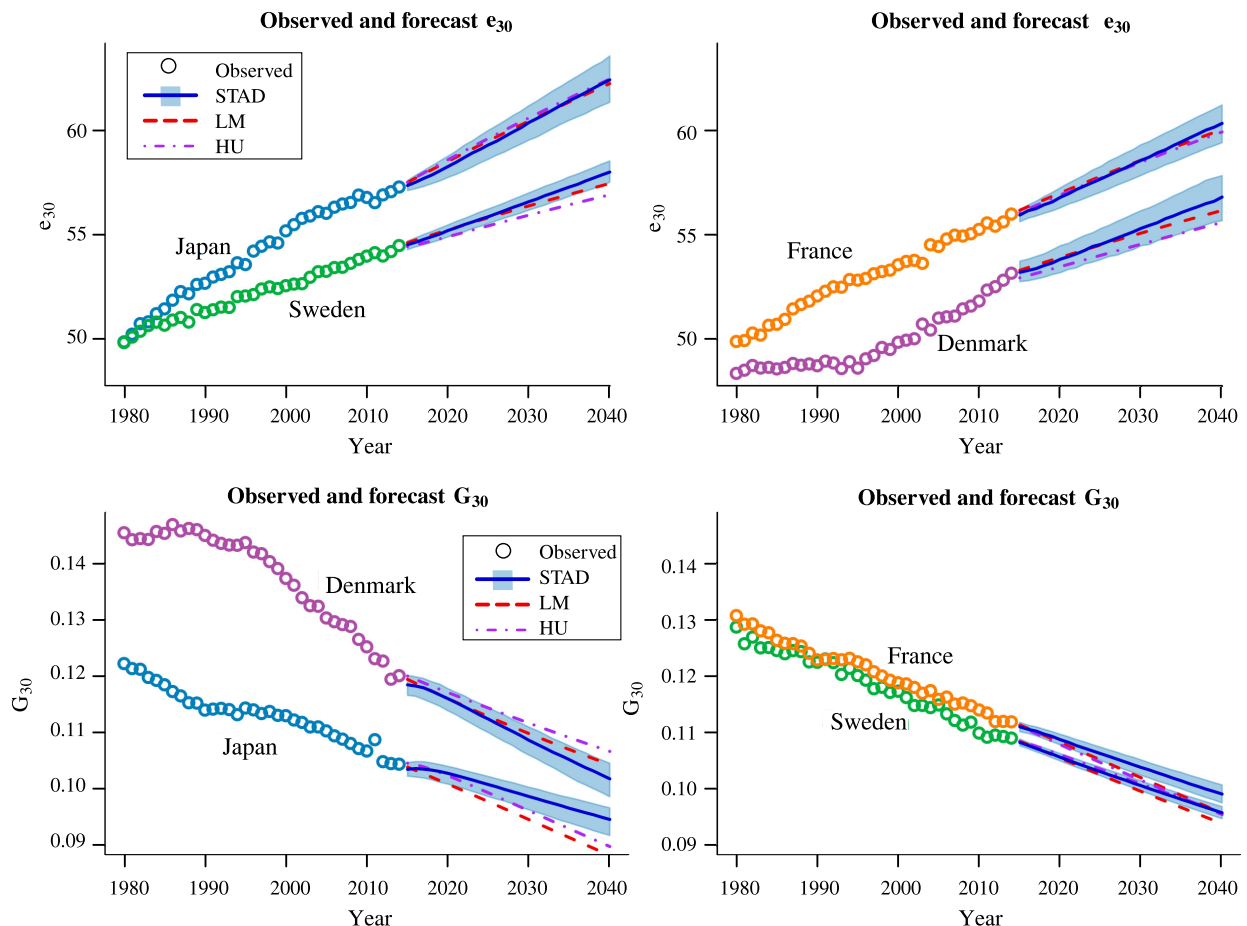
Figure 6   Observed and forecast remaining life expectancies at age 30 (e_{30}, upper panels) and Gini coefficients at age 30 (G_{30}, lower panels) from the STAD, LM, and HU models for females in Sweden, Japan, France, and Denmark, 1980–2040
Note: Shading indicates 80 per cent prediction intervals for the STAD model.
Source: As for Figure 1.

analysis (Ramsay and Silverman 2005). With this technique, the ages occurring before and after the modal age are no longer conflated: differences in the mode (already accounted for by the parameter *s*) are removed and the focus is only on differences in variability of the distributions. The use of functional data techniques in modelling and forecasting mortality already has other precedents in the literature (Hyndman and Ullah 2007; Hyndman and Booth 2008; Hyndman et al. 2013).

The shifting and compression dynamics of mortality change have been investigated extensively in recent decades (e.g., Fries 1980; Kannisto 2000; Bongaarts and Feeney 2002; Bongaarts 2005; Canudas-Romo 2008; Thatcher et al. 2010; Bergeron-Boucher et al. 2015; de Beer and Janssen 2016). In most developed countries, a compression of mortality was observed during the first half of the twentieth century. In the second half of the century, this dynamic was mainly replaced by the shifting of the mortality schedule,

with lifespan variability remaining practically constant. In addition to capturing longevity and lifespan variability changes, the three parameters of the STAD model provide information on these two dynamics. The parameter *s* regulates the shifting dynamic, as it measures the change in the modal age at death over time. The parameters b_L and b_U regulate the compression dynamic, as they expand or reduce the variability of deaths before and after the mode with respect to the standard distribution. These two parameters disentangle the contributions of two groups—those younger and older than the modal age—to the compression dynamic. This decomposition brings new evidence on the compression dynamic, adding a novel perspective to this stream of research (Fries 1980; Myers and Manton 1984; Wilmoth 1997; Wilmoth and Horiuchi 1999; Lynch and Brown 2001).

The STAD model can thus be used to study the developments of the shifting and compression dynamics over a specified time interval and this
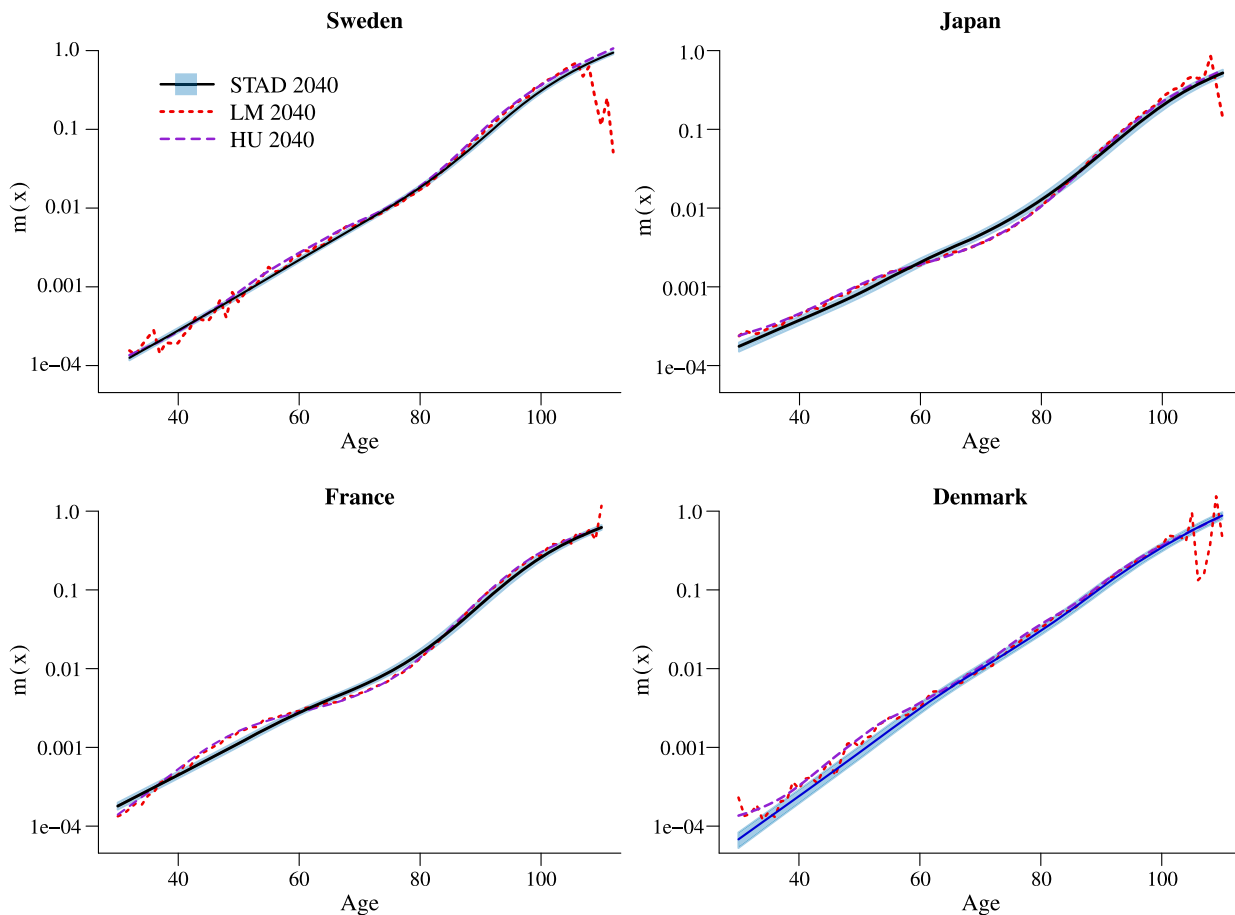
Figure 7   Forecast mortality rates by age in 2040 with 80 per cent prediction intervals (logarithmic scale) from the STAD, LM, and HU models for females in Sweden, Japan, France, and Denmark
Note: Shading indicates 80 per cent prediction intervals for the STAD model.
Source: As for Figure 1.

analysis can, in turn, usefully inform mortality projections. Appropriately accounting for these different dynamics in projections is a desirable property for forecasting mortality. For example, our analyses on females in four high-longevity countries between 1980 and 2014 suggest that the shifting dynamic has played a very important role in mortality patterns (as observed elsewhere, e.g., Canudas-Romo 2008; Bergeron-Boucher et al. 2015; de Beer and Janssen 2016). Compared with the LC model, the age-specific pattern of the STAD forecasts is characterized by a greater shift in all countries, as reflected in the forecast age-at-death distribution. Similar results have been shown for the Netherlands using a new parametric model that considers both shifting and compression (Janssen and de Beer 2016).

From a survival analysis perspective, the STAD model can be linked to the class of accelerated failure time (AFT) models. In survival analysis, AFT models have been introduced as an alternative to proportional hazard regression models to describe the effects of

covariates in accelerating or decelerating the ageing process (Wei 1992; Kalbfleisch and Prentice 2002). The STAD model can be interpreted as a particular type of AFT model, in which the ageing process is not uniformly modified with respect to the reference (the standard distribution) but is rather modified in different ways before (b_L) and after (b_U) the modal age at death. For example, the ageing process could be accelerated before and decelerated after the mode with respect to the standard, or vice versa, or accelerated/decelerated at different rates in the two age segments.

From this perspective, one limitation of the STAD model lies in the difficulty of interpreting its result from the individual's standpoint. It is quite unrealistic to contend that all individuals in a population age at two different rates, one before and one after the overall modal age at death. However, our aim here is to introduce a novel approach for modelling and forecasting mortality at the population level. The assumption of different rates of ageing can be more easily justified from this aggregate perspective, as the two
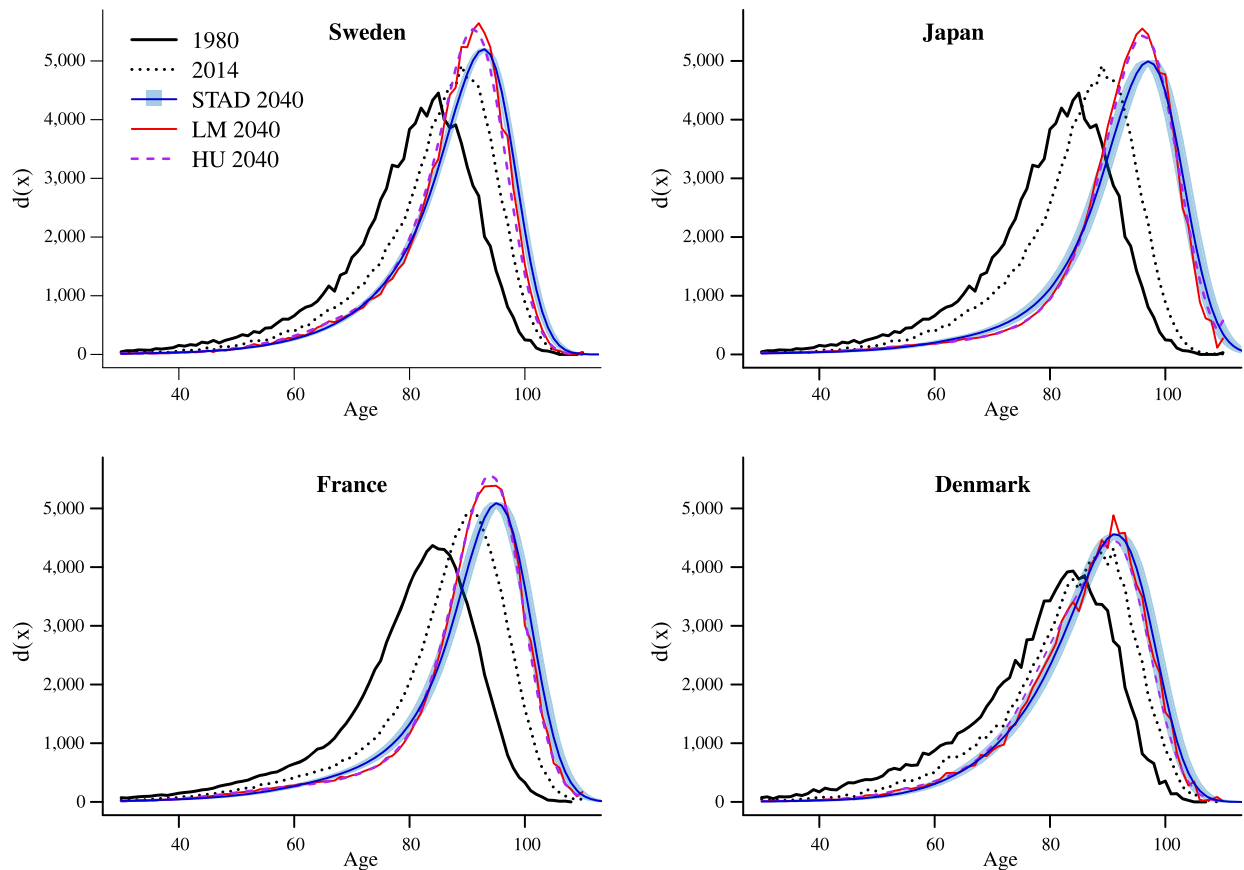
Figure 8   Observed life-table deaths in 1980 and 2014 and age-at-death distributions forecast for 2040 by the STAD, LM, and HU models for females in Sweden, Japan, France, and Denmark
Note: Shading indicates 80 per cent prediction intervals for the STAD model.
Source: As for Figure 1.

parameters capture the *average* ageing rate in these two age segments of the population. Nevertheless, while no experiments have been performed with medical and epidemiological data sets, the STAD might be unsuitable as a regression model for the analysis of survival data; a completely smooth transformation function, as in Camarda et al. (2008), might be preferable. Furthermore, an additional limitation of the STAD model, as of all other models used here, is that cohort effects are neither considered nor modelled. Future work is likely to include analysis of cohort data.

Our interest in this paper is limited to the senescent mortality pattern and our analyses start from age 30. Applying the STAD model to the entire age range would reduce fitting accuracy. The reason for this loss of accuracy lies in the shape of the human mortality pattern. Since Thiele (1871), demographers have tended to decompose the mortality age profile into three different components, operating principally on juvenile, younger adult, and older adult ages. Applying the model from age 30 produces satisfactory results for the populations of females analysed in this paper because the first two components are negligible with

respect to the senescent one in the age range that we study. For males, the STAD model produces satisfactory results too (analyses not shown), although the goodness of fit is slightly reduced compared with females. This is because the assumption that the younger adult mortality component (the 'accident hump') is negligible at age 30 is probably too strong for males. More accurate results for males could be obtained by extending this work to model and forecast the entire mortality pattern by: (i) disentangling and estimating the three independent age-specific mortality components; and (ii) applying the STAD model, or a modified version of it, to each component. The 'sum of smooth exponentials' model (Camarda et al. 2016) can be thought of as a starting solution for the first issue.

We conclude with one last remark on our STAD model. We chose the modal age at death as the break point of the segmented linear transformation because several demographic studies mentioned earlier had investigated the variability of deaths after the modal age at death. In this respect, we add to this literature by studying the variability before and after the mode *simultaneously*. However, any other age in the

distribution can be chosen as a break point, for example, another central tendency measure (mean or median age). In practice, this might be more appropriate for studying mortality in countries with data deficiencies or for the analysis of causes of death; this approach will be explored in future work.

Notes and acknowledgements

ORCID

Ugofilippo Basellini http://orcid.org/0000-0003-0292-1404

References

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

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

Bohk-Ewald, C., M. Ebeling, and R. Rau. 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.

Bongaarts, J. and G. Feeney. 2002. How long do we live? *Population and Development Review* 28(1):13–29.

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

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

Booth, H., J. Maindonald, and L. Smith. 2002. Applying Lee–Carter under conditions of variable mortality decline, *Population Studies* 56(3): 325–336.

Brass, W. 1971. On the scale of mortality, in W. Brass (ed.), *Biological Aspect of Demography*. London: Taylor & Francis, pp. 69–110.

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

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

Cadena, M. and M. Denuit. 2016. Semi-parametric accelerated hazard relational models with applications to mortality projections, *Insurance: Mathematics and Economics* 68: 1–16.

Cairns, A. J., D. Blake, and K. Dowd. 2006. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration, *Journal of Risk and Insurance* 73(4): 687–718.

Camarda, C. G. 2012. Mortalitysmooth: An R package for smoothing Poisson counts with P-splines, *Journal of Statistical Software* 50: 1–24. Available: www.jstatsoft.org/v50/i01.

Camarda, C. G., P. H. Eilers, and J. Gampe. 2008. A warped failure time model for human mortality, *Proceedings of the 23rd international workshop of statistical modelling*, pp. 149–154.

Camarda, C. G., P. H. Eilers, and J. Gampe. 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.

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

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

Cheung, S. L. K. and J.-M. Robine. 2007. Increase in common longevity and the compression of mortality: The case of Japan, *Population Studies* 61(1): 85–97.

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

Christensen, K., M. Davidsen, K. Juel, L. Mortensen, R. Rau, and J. W. Vaupel. 2010. The divergent life-expectancy trends in Denmark and Sweden - and some potential explanations, in E. M. Crimmins, S. H. Preston, and B. Cohen (eds), *International Differences in Mortality at Older Ages: Dimensions and Sources*. Washington, DC: National Academies Press, pp. 385–407.

Clarkson, D. B., C. Fraley, C. C. Gu, and J. O. Ramsay. 2005. *S+Functional Data Analysis: User's Manual for Windows*. New York: Springer.

Colchero, F., R. Rau, O. R. Jones, J. A. Barthold, D. A. Conde, A. Lenart, L. Nemeth, A. Scheuerlein, J. Schoeley, C. Torres, V. Zarulli, J. Altmann, D. K. Brockman, A. M. Bronikowski, L. M. Fedigan, A. E. Pusey, T. S. Stoinski, K. B. Strier, A. Baudisch, S. C. Alberts, and J. W. Vaupel. 2016. The emergence of long-evous populations, *Proceedings of the National Academy of Sciences* 113(48): E7681–E7690.

Currie, I. D., M. Durban and P. H. Eilers. 2004. Smoothing and forecasting mortality rates, *Statistical Modelling* 4 (4): 279–298.

de Beer, J. and F. Janssen. 2016. A new parametric model to assess delay and compression of mortality, *Population Health Metrics* 14(1): 46. doi:10.1186/s12963-016-0113-1.

Edwards, R. D. 2013. The cost of uncertain life span, *Journal of Population Economics* 26(4): 1485–1522.

Edwards, R. D. and S. Tuljapurkar. 2005. Inequality in life spans and a new perspective on mortality convergence across industrialized countries, *Population and Development Review* 31(4): 645–674.

Efron, B. and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. London: Chapman and Hall.

Eilers, P. H. and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties, *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.

Hamilton, J. D. 1994. *Time Series Analysis, Volume 2*. Princeton, NJ: Princeton University Press.

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

Human Mortality Database (HMD). 2018. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available: www.mortality.org or www.humanmortality.de (data downloaded on 22 March 2018).

Hyndman, R. J. and H. Booth. 2008. Stochastic population forecasts using functional data models for mortality, fertility and migration, *International Journal of Forecasting* 24(3): 323–342.

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

Hyndman, R. J., C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. 2018. forecast: forecasting functions for time series and linear models. *R Package Version* 8.3.

Hyndman, R. J., H. Booth, L. Tickle, and J. Maindonald. (2017). demography: Forecasting Mortality, Fertility, Migration and Population Data. R package version 1.20.

Hyndman, R. J. and M. S. Ullah. 2007. Robust forecasting of mortality and fertility rates: a functional data

approach, *Computational Statistics & Data Analysis* 51 (10): 4942–4956.

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

Janssen, F. and J. de Beer. 2016. Projecting future mortality in the Netherlands taking into account mortality delay and smoking, in Joint Eurostat/UNECE Work Session on Demographic Projections.

Kalbfleisch, J. D. and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. Second edition. New York: John Wiley & Sons.

Kannisto, V. 2000. Measuring the compression of mortality, *Demographic Research* 3(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 M. L. Moeschberger. 2005. *Survival Analysis: Techniques for Censored and Truncated Data*. Second edition. New York: Springer Science & Business Media.

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

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

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

Li, N., R. D. Lee, and P. Gerland. 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.

Lindahl-Jacobsen, R., J. Oeppen, S. Rizzi, S. Möller, V. Zarulli, K. Christensen, and J. W. Vaupel. 2016. Why did Danish women's life expectancy stagnate? The influence of interwar generations' smoking behaviour, *European Journal of Epidemiology* 31(12): 1207–1211.

Lindahl-Jacobsen, R., R. Rau, B. Jeune, V. Canudas-Romo, A. Lenart, K. Christensen, and J. W. Vaupel. 2016. Rise, stagnation, and rise of Danish women's life expectancy, *Proceedings of the National Academy of Sciences* 113 (15): 4015–4020.

Lynch, S. M. and J. S. Brown. 2001. Reconsidering mortality compression and deceleration: An alternative model of mortality rates, *Demography* 38(1): 79–95.

Maddison, A. 2006. The contours of world development, in *The World Economy*. Volume 1: A Millennial Perspective and Volume 2: Historical Statistics. Paris: OECD Publishing, pp. 29–50.

Meslé, F. and J. Vallin. 2011. Historical trends in mortality, in G. R. Rogers and M. E. Crimmins (eds), *International Handbook of Adult Mortality*. New York: Springer, pp. 9–47.

Myers, G. C. and K. G. Manton. 1984. Compression of mortality: myth or reality? *The Gerontologist* 24(4): 346–353.

Oeppen, J. 2008. Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data, *Compositional data analysis conference*.

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

Ouellette, N. and R. Bourbeau. 2011. Changes in the age-at-death distribution in four low mortality countries: A non-parametric approach, *Demographic Research* 25: 595–628.

Pascariu, M., A. Lenart, and V. Canudas-Romo. 2017. Forecasting mortality by using statistical moments, *13th international longevity risk and capital markets solutions conference*.

Pfaff, B. 2008a. *Analysis of Integrated and Cointegrated Time Series with R*. New York: 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., P. Heuveline, and M. Guillot. 2001. *Demography. Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

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

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

Ramsay, J. O. and B. W. Silverman. 2005. *Functional Data Analysis*. Second edition. New York: Springer-Verlag.

Rothenberg, R., H. R. Lentzner, and R. A. Parker. 1991. Population aging patterns: the expansion of mortality, *Journal of Gerontology* 46(2): S66–S70.

Sasson, I. 2016. Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010, *Demography* 53(2): 269–293.

Schwarz, G. 1978. Estimating the dimension of a model, *The Annals of Statistics* 6(2): 461–464.

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

Shkolnikov, V. M., E. E. Andreev, and A. Z. Begun. 2003. Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples, *Demographic Research* 8(11): 305–358.

Smits, J. and C. Monden. 2009. Length of life inequality around the globe, *Social Science & Medicine* 68(6): 1114–1123.

Tabeau, E. (2001). A review of demographic forecasting models for mortality, in *Forecasting Mortality in Developed Countries*. Dordrecht: Kluwer Academic Publishers, pp. 1–32.

Thatcher, A. R., S. L. K. Cheung, S. Horiuchi, and J.-M. Robine. 2010. The compression of deaths above the mode, *Demographic Research* 22: 505–538

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.

Torri, T. and J. W. Vaupel. 2012. Forecasting life expectancy in an international context, *International Journal of Forecasting* 28(2): 519–531.

United Nations. 2015. *World Population Ageing 2015 (ST/ ESA/SER.A/390)*. New York, NY: Department of Economic and Social Affairs, Population Division.

Vallin, J. and F. Meslé. 2004. Convergences and divergences in mortality: A new approach of health transition, *Demographic Research* 2: 11–44.

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

Vaupel, J. W. and H. Lundström. 1994. Longer life expectancy? evidence from Sweden of reductions in mortality rates at advanced ages, in *Studies in the Economics of Aging*. Chicago: University of Chicago Press, pp. 79–102.

Vaupel, J. W., Z. Zhang, and A. A. van Raalte. 2011. Life expectancy and disparity: an international comparison of life table data, *BMJ Open* 1(1): e000128. doi:10. 1136/bmjopen-2011-000128.

Villegas, A. M., P. Millossovich, and V. K. Kaishev. 2017. StMoMo: An R Package for Stochastic Mortality Modelling. R package version 0.4.0.

Wei, L.-J. 1992. The accelerated failure time model: A useful alternative to the cox regression model in survival analysis, *Statistics in Medicine* 11(14-15): 1871–1879.

Wilmoth, J. R. 1997. In search of limits, in K. W. Wachter and C. E. Finch (eds), *Between Zeus and the Salmon: The Biodemography of Longevity*. Washington, DC: National Academies Press (US), pp. 38–64.

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

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

Wilmoth, J. R., K. Andreev, D. Jdanov, D. A. Glei, C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, and P. Vachon. 2007. Methods Protocol for the Human Mortality Database. Last Revised: May 31, 2007 (Version 5).

Appendix A

Deviance, effective dimension, and BIC

The deviance is often used as a measure of discrepancy between observed and fitted data. Within a Poisson framework, it is defined as:

$$\text{Dev} = 2 \sum_y \sum_x \left[D_{x,y} \ln\left(\frac{D_{x,y}}{\hat{D}_{x,y}}\right) - (D_{x,y} - \hat{D}_{x,y}) \right] \tag{A1}$$

where $D_{x,y}$ and $\hat{D}_{x,y}$ denote the observed and fitted numbers of deaths at age x and year y, respectively. Higher values of the deviance correspond to inferior models in terms of goodness of fit.

The BIC (Schwarz 1978) is frequently used to assess model differences in terms of trade-off between model parsimony and accuracy. In a two-dimensional age and time setting, it can be computed as:

$$\text{BIC} = \text{Dev} + \ln(mn)\text{ED} \tag{A2}$$

where m and n are the dimensions (length) of age and time, respectively. ED denotes the effective dimension, or total number of parameters, of a model. Lower BIC values are associated with better models and the trade-off between accuracy and parsimony is accounted for by the two components of the BIC.

Appendix B

Additional results for 'Observed vs. fitted data' subsection

Here, we present the deviance and ED of the STAD, LC, and LC variant models corresponding to the analysis in the 'Observed vs. fitted data' subsection. Table B1 shows the deviance values of the STAD and LC variants in the four countries, while the ED of each model is reported in Table B2.

In terms of the deviance, the HU model is the best fitting, as its deviance is always smaller than those of other models. One explanation for the improved fit of the HU model compared with the STAD and other LC variants is that it uses a very high number of parameters: almost four times more than the LC (765 vs. 195) and over five times more than the STAD (765 vs. 138). This high parameterization is penalized by the BIC measure, making the HU model the optimal choice in only one instance (Japan, see Table 1). In terms of parsimony, the STAD model is the best performer, as it uses fewer parameters than the other models.

Table B1   Deviance values of the STAD, LC, and LC variant models for females aged 30–110+ in four countries, 1980–2014

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	3,179	3,149	3,153	3,123	3,060	**2,789**	2,958	2,936
Japan	9,956	9,925	10,268	9,837	9,240	**3,975**	4,612	4,900
France	8,043	6,272	6,400	6,223	6,113	**3,320**	4,050	3,776
Denmark	4,235	4,063	4,054	4,036	3,934	**2,938**	3,316	3,808

Note: Lower values of the deviance (lowest in bold) correspond to a better fit to the data.
Source: As for Table 1.

Table B2   Effective dimension (total number of parameters) of the STAD, LC, and LC variant models for females aged 30–110+ in four countries, 1980–2014

Country	Model							
	STAD	LC	LM	BMS	BDV	HU	HUrob	HUw
Sweden	**138**	195	195	195	195	765	765	765
Japan	**138**	195	195	195	195	765	765	765
France	**138**	195	195	195	195	765	765	765
Denmark	**138**	195	195	195	195	765	765	765

Note: Lower ED (lowest in bold) corresponds to a more parsimonious model.
Source: As for Table 1.