

Modeling and Forecasting Age at Death Distributions

Ugofilippo Basellini* Carlo Giovanni Camarda†

December 15, 2015

Abstract

Age at death distributions provide an extremely informative description of mortality, yet they are generally neglected in modeling and forecasting. In this article, we use age at death distributions to model the age-specific pattern of mortality and to inform mortality forecasts. In particular, we introduce a segmented linear transformation model based on the modal age at death and the variability of deaths before and after the mode. This approach allows capturing the compression and shifting dynamics of mortality. We illustrate our methodology by estimating the distribution and life expectancy of two high-longevity countries in the last thirty years. We show that the fitted life expectancies are very close to the observed historical values. Furthermore, we forecast distributions and life expectancies fifteen years ahead by using time series models for the parameters of the segmented linear transformation.

1 Introduction

Mortality modeling has a fairly long history in demographic analysis, stretching back at least to the beginning of the nineteenth century (Booth and Tickle, 2008). Mortality models often focus on the age-specific pattern of mortality rates; in particular, parametric distributions that try to capture the force of mortality with few parameters have been very popular (for a comprehensive review, see Tabeau, 2001).

Age at death distributions provide an extremely informative description of the mortality experience of a population (Preston et al., 2001). Yet, this crucial feature of mortality has generally been neglected in modeling and forecasting (with the exception of projections made by compositional data analysis in Oeppen et al., 2008).

The distribution of deaths can be efficiently summarized by its first two moments. The first moment is the mean of the distribution, which is equivalent to life expectancy at birth. The second moment is the variability of the distribution, which measures the dispersion of the distribution around its mean.

Trends of life expectancy and variability of lifespan have received much attention during the last decades (see, for example, Oeppen and Vaupel, 2002; Vaupel et al., 2011). In low-mortality countries, the adult modal age at death has probably become

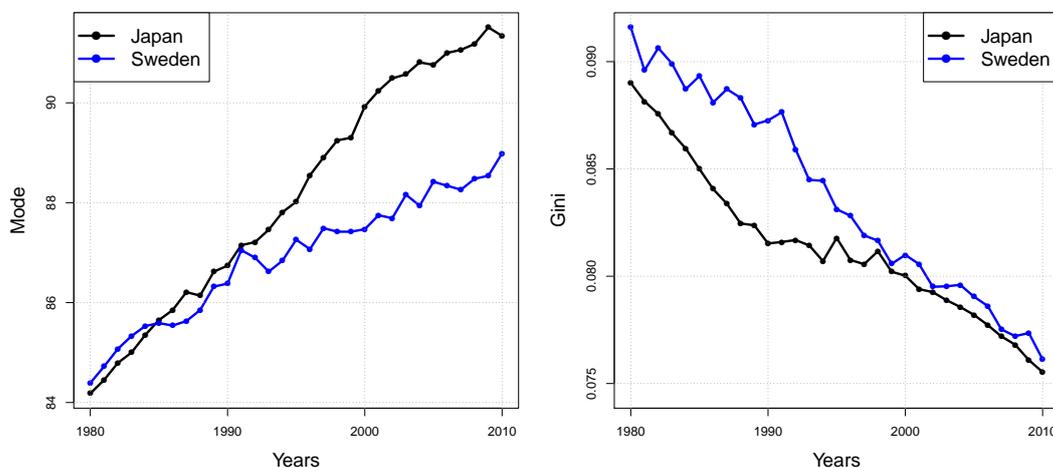
*European Doctoral School of Demography, Sapienza University of Rome and Max Planck Institute for Demographic Research, Rostock.

†Institut National d'Études Démographiques, Paris.

a better indicator of longevity than life expectancy (Horiuchi et al., 2013). The modal age at death is the age where most of the deaths occur, and it corresponds to the age where the density function has a maximum value (Canudas-Romo, 2008).

Figure 1 shows the evolution of the modal age at death and the variability of lifespan (measured with the Gini coefficient) during the last thirty years for females in two high-longevity countries, Sweden and Japan. Both measures have followed a quite stable linear trend during this period; however, it is hard to infer from the summary measures which segments of the age range contributed most towards the observed changes.

Figure 1: Modal age at death (left panel) and variability of lifespan (measured with the Gini coefficient, right panel) in Japanese (black) and Swedish (blue) females in 1980-2010



Source: authors' elaboration on data from the Human Mortality Database (2015).

We wish to study the changes that have occurred in age at death distributions during the last thirty years. In order to do so, we propose a non-parametric segmented linear transformation model that focuses on the modal age at death and the variability of deaths before and after the mode.

This article is organized as follows. In Section 2, we overview the mathematical methods that we will use throughout this article. In Section 3, we first illustrate our methodology by estimating the density and life expectancy of two high-longevity countries during the last thirty years. Then, we forecast age at death distributions for the two countries fifteen years ahead. In Section 4, we discuss the results and conclude.

2 Methods

2.1 Segmented linear transformation model

We seek to find an approach for modeling the age pattern of mortality that incorporates the first two moments of age at death distributions. These two moments are closely related to the concepts of *pace* and *shape* of aging (Baudisch, 2011). Pace is defined as the time scale of mortality, and it measures the characteristic length of life. Shape instead is independent of time, and it refers to the change in the age-pattern of mortality over the life course (Wrycza and Baudisch, 2014; Wrycza et al., 2015).

Whereas the development of the modal age at death can shed light on the changes in pace over time, it is possible to obtain additional insights on the shape of aging by looking at the variability of the distribution over two different segments of the age range.

We propose an approach that captures changes in mortality patterns by a suitable transformation of the age-axis, such that a series of age at death distributions over time can be expressed solely by changes of the modal age at death (pace) and of expansion/compression of the variability of deaths (shape). In particular, we split the total variability of the distribution in changes before and after the mode.

For ease of presentation, here we consider only two age at death distributions: a target distribution $f(x)$, and an observed distribution $g(x)$, where x denotes age. Our aim is to estimate the transformation function $t(x)$ so that the density of the distribution $g(x)$ conforms to the target density $f(x)$ on the warped axis:

$$g(x) = f(t(x))$$

The main idea is that $t(x)$ is a function of: (i) the change in the modal age at death of the two distributions, and (ii) two additional parameters that capture the expansion/shrinkage of the ages before and after the mode which is needed to conform the observed to the target distribution.

Let s be the difference between the modal age at death of $f(x)$ and $g(x)$:

$$s = M^g - M^f$$

The transformation function $t(\cdot)$ can be thus written as follows:

$$t(x; s, b_L, b_U) = \begin{cases} M^f - b_L(s + M^f) + b_L x & \text{if } x \leq M^g \\ M^f - b_U(s + M^f) + b_U x & \text{if } x > M^g \end{cases} \quad (1)$$

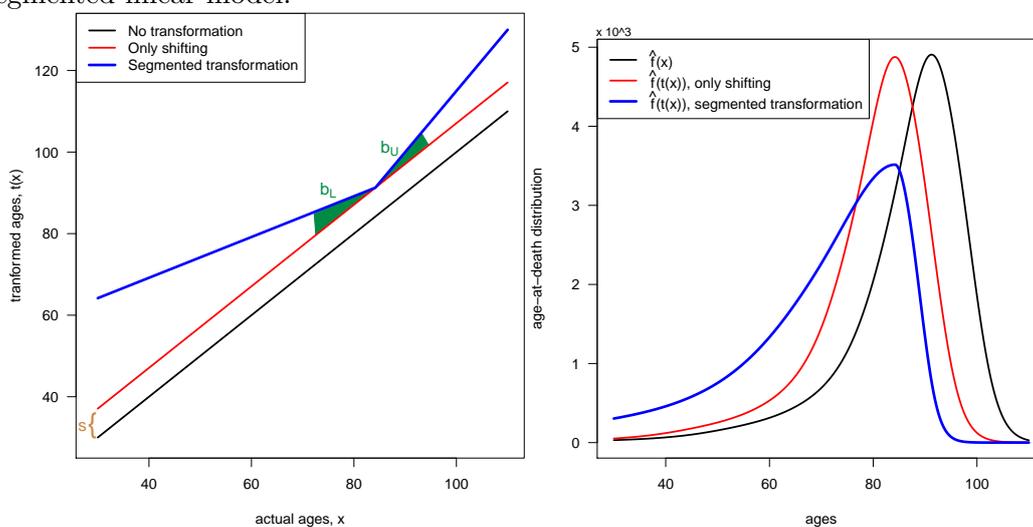
In words, $t(x; s, b_L, b_U)$ takes the form of a segmented linear transformation model which breaks at the value of M^g . The slopes of each linear part b_L and b_U capture the amounts of expansion/shrinkage needed before and after M^g , respectively, in order to fit the observed distribution $g(x)$.

Figure 2 presents a schematic overview of the effects of $t(x)$ to a given distribution $f(x)$. If a simple shift is adopted (red line of the left panel), the target distribution is only shifted to the left maintaining the same variability before and after the modal

age at death. This is a special case of (1) in which both b_L and b_u are equal to 1 and the transformation function becomes: $t(x) = s + x$.

More realistic scenarios could be obtained by modifying the value of b_L and b_U which act jointly with the shifting parameter s . When b_L and b_U are bigger than 1, the ages before and after the mode of $g(x)$ are shrunk, and therefore variability in age-at-death is decreasing. Vice versa, when b_L and b_U are smaller than 1, we obtain an expansion of the age axis and a consequent increase of the age at death variation before and after the mode, respectively. In the example presented in Figure 2, the ages before the mode are expanded and the ages above are shrunk (see blue lines).

Figure 2: A schematic overview of the effects of transforming the age-axis using a segmented linear model.



Moving from a continuous to a discrete framework, we first smooth all observed age at death distributions. We used a P -splines approach as described in Camarda (2012). Specifically, we describe the age at death distribution over age x for a particular year y as follows:

$$g(x)_y = \exp[\mathbf{B}(x) \boldsymbol{\beta}_y]$$

where $\mathbf{B}(x)$ is a basis of equally spaced B -splines, common for all the years, and $\boldsymbol{\beta}_y$ are year-specific penalized coefficients. By computing $\mathbf{B}(x)$ at finer grid of x , we are able to evaluate each distributions at any finer scale, practically at a continuous level. This allows us to estimate precisely any feature of the age at death distribution, including the mode M (Ouellette and Bourbeau, 2011) as shown in Figure 1.

2.2 The target distribution

As a target density, instead of choosing an actual or a mean age at death distribution, we first transform the observed densities so that their qualitative features line up. Specifically, we register all densities so that their modal age at death are equal to the modal age at death of the first observed density. In functional data analysis, this is called landmark registration (Ramsay and Silverman, 2005).

One can also view this technique as a simple shifting transformation of the age axis where the parameter s is given by the difference between the modal age at death in year y and y_0 , i.e. $s_y = M^y - M^{y_0}$.

In formulas, the aligned age at death distribution in year y can be written as follows:

$$g(w(x))_y = \exp[\mathbf{B}(w(x)) \boldsymbol{\beta}_y]$$

where $w(x) = x + s_y$. In other words, using the already estimated $\boldsymbol{\beta}_y$, it is sufficient to evaluate the B -spline basis at a new shifted axis for aligning the corresponding distribution.

Figure 3 presents the age at death distributions for Sweden and Japan from age 10 and for all years between 1980 and 2011, aligned so that all modes are equal to the modal age at death in 1980. We remove the peak of infant mortality and start our analyses from age 10, as our interest here is limited to the adult mortality pattern.

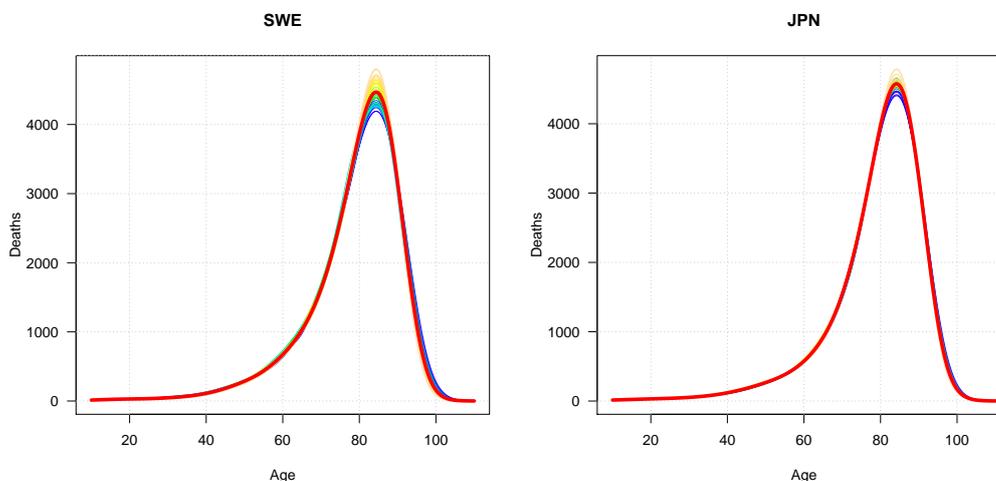
The bold red lines present the mean distributions after registration which we will use as a target distribution $f(x)$. Again the target distribution could be expressed as a linear combination of B -splines and “target” penalized coefficients:

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

It is easy to see that such registered mean is a good representation of all densities because it does not mix up features that in the original axis would have occurred at different distances with respect to the mode.

We decided to use this smoothing functional data approach for the target distribution, because despite the shifting transformation and the segmented linear warping of the ages before and after the mode, all features contained in $f(x)$ are carried in the transformation procedure.

Figure 3: Aligned age at death distributions in the years 1980-2011 for Swedish (left) and Japanese (right panel) females. The alignment is achieved such that modal ages at death in each year line up to the mode in 1980.



2.3 Estimating expansion/shrinkage parameters

Given the target distribution and the shifting parameter s_y estimated as described above, we aim to estimate the expansion/shrinkage parameters $b_{L,y}$ and $b_{U,y}$ for each year. In order to do this, we minimize the sum of squared differences between the observed and the target distribution over the transformed age axis. Let $d_{x,y}$ be the observed deaths derived from period life-tables over age x in year y ; then, we can write the objective function with respect to $b_{L,y}$ and $b_{U,y}$ as follows:

$$\begin{aligned} S(b_{L,y}, b_{U,y} | s_y, d_{x,y}, \beta_f, p_y) &= \sum_x [d_{x,y} - f(t(x, s_y; b_{L,y}, b_{U,y}))]^2 \\ &= \sum_x \{d_{x,y} - p_y \exp[\mathbf{B}(t(x, s_y; b_{L,y}, b_{U,y}) \beta_f)]\}^2 \end{aligned}$$

where p_y is the total number of life-table deaths above age 10 in year y , and the transformation function of the age axis $t(\cdot)$ is given in (1). A general-purpose numerical optimizer works well for minimizing $S(\cdot)$ over the two-dimensional space of the expansion/shrinkage parameters $b_{L,y}$ and $b_{U,y}$. Routines for estimating the target distribution as well as the transformation function for the age axis were implemented in R (R Development Core Team, 2015), and they are available upon request.

3 Results

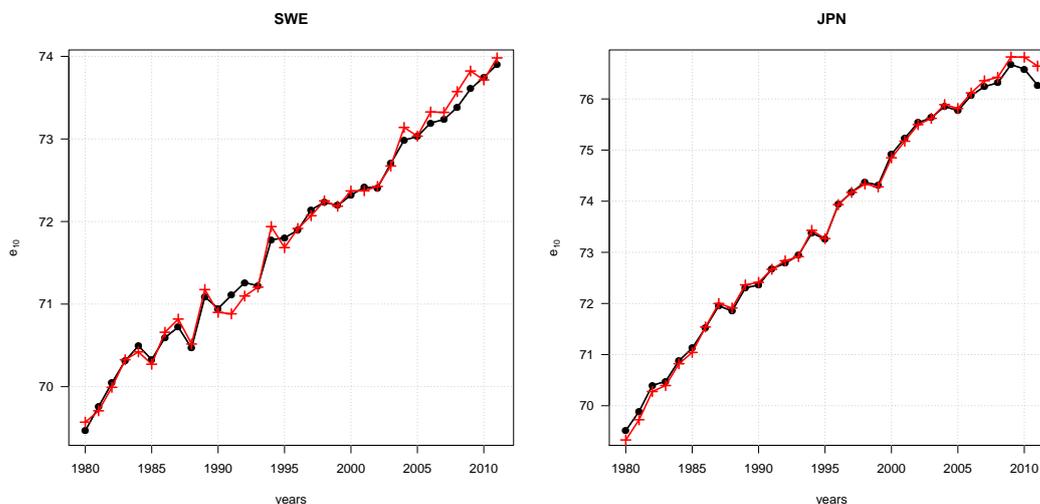
3.1 Observed versus fitted data

In this section, we illustrate our methodology by estimating the distribution and remaining life expectancy of females in two high-longevity countries, Sweden and Japan, during the last thirty years.

We use period life-tables from the Human Mortality Database (2015) to obtain age at death distribution data for the two countries in this period. We fix the target distribution and estimate the parameters s , b_L and b_U as described in Section 2. From these estimates, we can compute the entire distributions over the whole age range (starting from age 10). Moreover, summary measures such as remaining life expectancy can be computed from these densities. Figure 4 shows the observed and fitted remaining life expectancies in Swedish and Japanese females during the years 1980-2011.

The two graphs of Figure 4 show that the observed and fitted remaining life expectancies are very close to each other. In Figure 5, we show the trends of the estimated slopes which have been used to transform the age axis in order to conform the target distribution to the years considered here. In Sweden, both parameters b_L and b_U show an upward trend (upper panels). This is a clear sign of decreasing lifespan variation in both segments of the age range. In the bottom panels, we can observe that Japanese women also experienced a shrinkage in the ages above the mode (right panel), while an erratic behavior is evident in the development before the mode (left panel).

Figure 4: Observed (black) and fitted (red) remaining life expectancies at age 10 in Swedish (left panel) and Japanese (right panel) females in 1980-2011



3.2 Forecasting

The estimated parameters s , b_L and b_U can be additionally used to forecast age at death distributions by projecting the trends of the mode and the compression/expansion of deaths before and after this measure.

We use univariate autoregressive integrated moving average (ARIMA) models to describe and forecast the time series of the three parameters (Box and Jenkins, 1970). Model selection is performed with a step-wise algorithm that minimizes the Akaike Information Criterion (AIC) of different model specifications (Akaike, 1974; Hyndman and Khandakar, 2008).

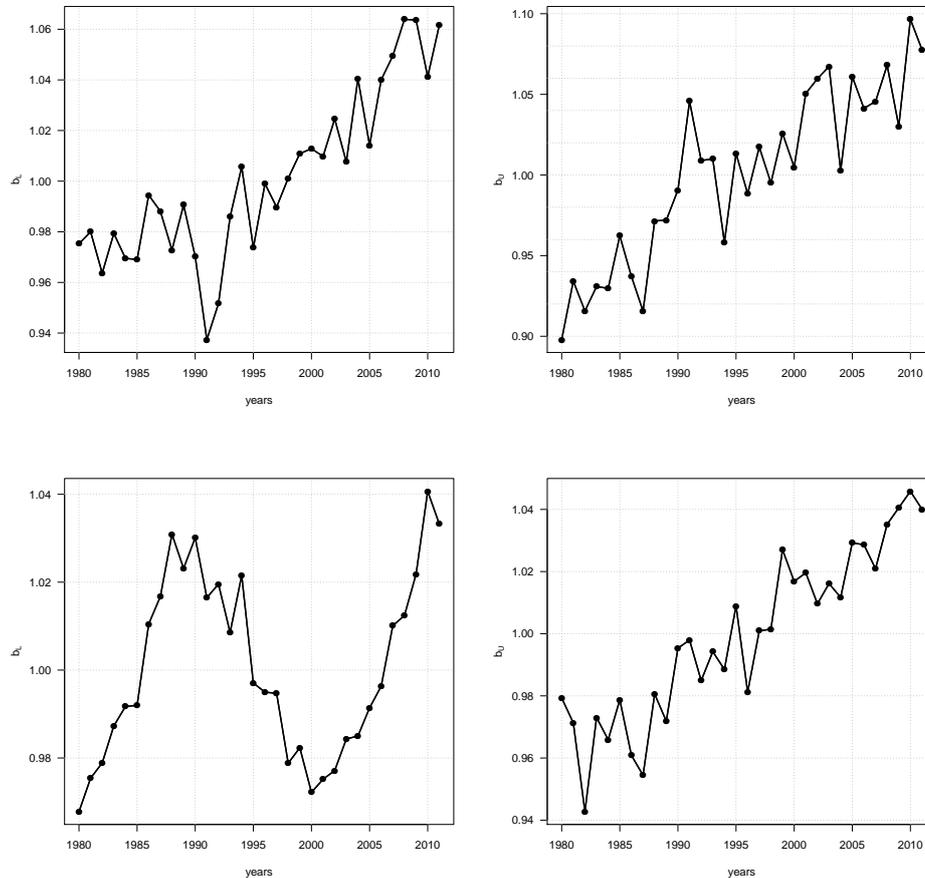
In order to bring coherence to the development of mortality, we started by hypothesizing a single multivariate specification for the evolution of the three parameters. Multivariate models indeed allow to take into account (and estimate) the dynamic interrelation between variables having a common data generating process.

However, we did not find empirical support for such multivariate models: the multivariate vector autoregressive (VAR) specification of the time series did not have significant coefficients for the interrelated parameters, and we did not find evidence of co-integrated time series.

The reasons for these findings might be several. First of all, the time series considered here might be too short: multivariate models have more parameters than univariate ones, and the statistical power might not be enough to estimate the additional parameters. Furthermore, it is difficult to generalize nonlinear procedures to the multivariate case. Finally, outliers can have a more serious effect on multivariate than one univariate models.

As a result, we decided to restrict our attention to univariate models for now. However, we will further investigate the possibility of using multivariate models (see Section 4

Figure 5: Estimated slopes of the segmented transformations b_L and b_U for Swedish (upper panels) and Japanese (bottom panels) females in 1980-2011



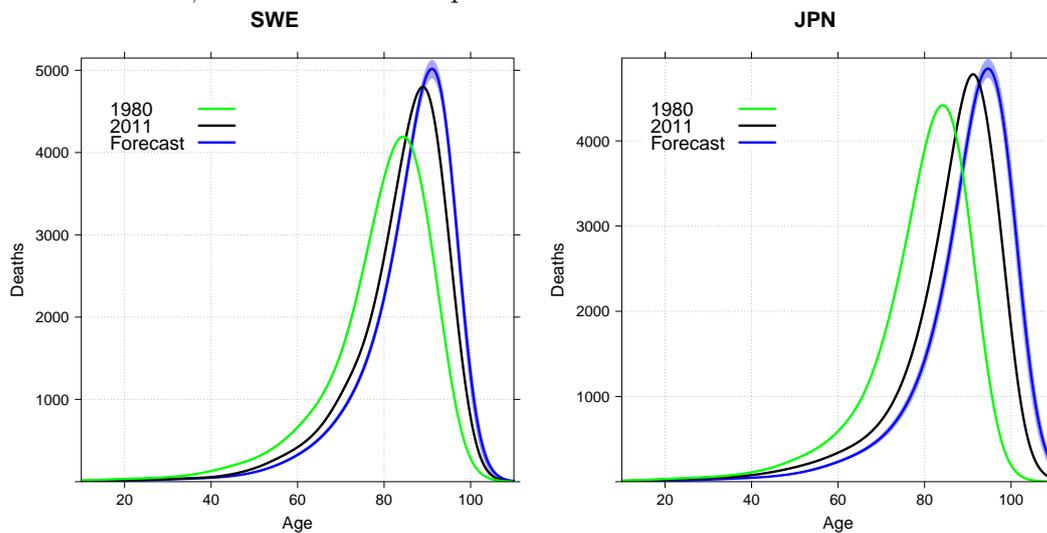
for further discussion).

The models for the stationary differenced series s are a MA(1) model with drift for Sweden, and a random walk model with drift for Japan. For b_L , a MA(1) model with drift and a random walk model are used for the two countries. Finally, an AR(1) model with and without drift are estimated for b_U in Sweden and Japan respectively.

From the projections of the three parameters, it is possible to forecast the entire age at death distribution by extrapolating the total number of deaths occurring after the first age in the analyses (age 10 here). We fitted a logistic function to the total number of deaths (not shown here) to consider the upper bound of life table deaths. Figure 6 shows fifteen years ahead forecasts of age at death distributions, with 80% confidence intervals, for Swedish and Japanese females. Confidence intervals are constructed from a bootstrapping procedure.

It is interesting to observe the importance played by the variability of deaths. Indeed, while the shift of the mode in the two distributions is comparable (the parameter s increases linearly for both countries, albeit at a faster rate in Japan), the projected compression of deaths in both segments of the age range in Sweden results in a shrink-

Figure 6: Fifteen years ahead forecasts of age at death distributions, with 80% confidence intervals, for Swedish and Japanese females.



age of the forecasted density. This is not observed in the case of Japan, where the forecasted density is only characterized by a shift from the target distribution (due to constant forecasts of both b_L and b_U).

Finally, Figure 7 shows the forecasts of remaining life expectancy at age 10, with 80% confidence intervals, computed from the forecasted distributions in the two countries.

4 Discussion

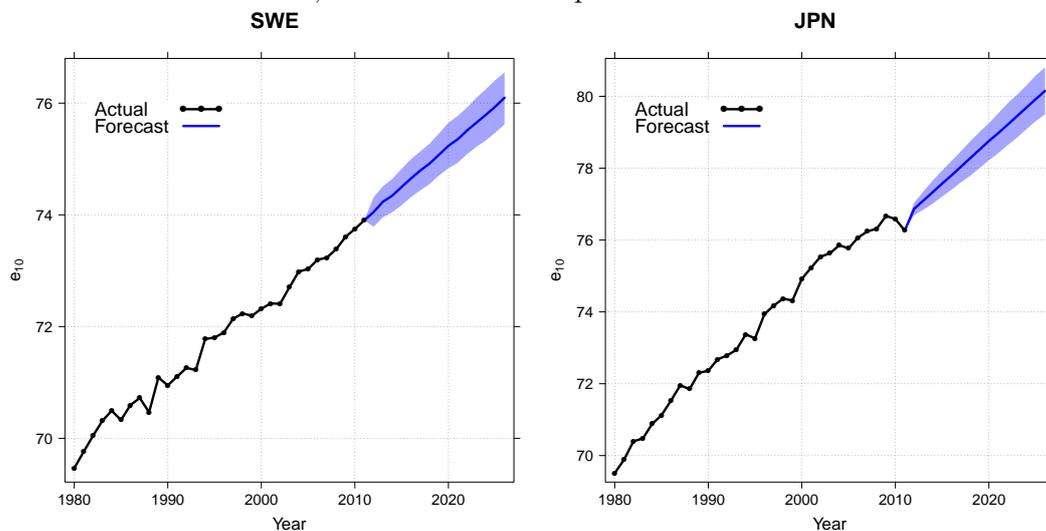
Trends in longevity and in the variability of lifespan have been investigated extensively during the last decades. In most developed countries, a compression of mortality has been observed during the first half of the 20th century. In the second half of the century, the compression dynamic has been mainly replaced by the shifting of the mortality schedule, with lifespan variability remaining nearly constant (Bergeron-Boucher et al., 2015).

In this article, we present a new methodological approach that allows capturing the compression and shifting dynamics of mortality by modeling age at death distributions. In particular, we introduce a segmented linear transformation model based on the modal age at death and the variability of deaths before and after the mode.

In addition to modeling the age pattern of mortality, our approach can be used to inform mortality forecasts by projecting trends in the shift and compression of mortality. In Section 3, we present the results of forecasting mortality for females of two high-longevity countries, Sweden and Japan.

The forecasts that we derive for Japan are in line with the observed trend of a predominance of the shifting mortality dynamic during the second part of the 20th century.

Figure 7: Fifteen years ahead forecasts of remaining life expectancy at age 10, with 80% confidence intervals, for Swedish and Japanese females.



Interestingly, the forecasts for Sweden are instead characterized by a mixture of shifting and compression of mortality, due to the consistent decrease of lifespan variation observed during the last thirty years in this country. Nevertheless, our forecasts provide evidence of a likely continuation of the shifting mortality dynamic, which is due to the consistent upward trend of the shifting parameter s .

As we stated in Section 3, we have used univariate models to describe and forecast the parameters of the segmented linear transformation. Indeed, we did not find empirical support for multivariate models. However, we will further investigate this possibility by analysing more countries and more years. Furthermore, another option that we will explore is that of simulated mortality data, which should bring additional insights on the behaviour of the model.

Finally, we will also investigate the possibility of having a break-point for the segmented linear transformation different from the modal age at death. For example, we will analyse whether breaking the transformation at the median or mean of the distribution will improve the fit of the model. This might be an important avenue of future research, for example in the analysis of different causes of death. Indeed, in such context, it may be more interesting to study the variability of deaths before and after a certain age x smaller than the mode.

References

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19(6), 716–723.
- Baudisch, A. (2011). The pace and shape of ageing. *Methods in Ecology and Evolution* 2(4), 375–382.
- Bergeron-Boucher, M.-P., M. Ebeling, and V. Canudas-Romo (2015). Decomposing

- changes in life expectancy: Compression versus shifting mortality. *Demographic Research* 33, 391–424.
- Booth, H. and L. Tickle (2008). Mortality modelling and forecasting: A review of methods. *Annals of actuarial science* 3(1-2), 3–43.
- Box, G. E. and G. M. Jenkins (1970). *Time-series Analysis, Forecasting and Control*. San Francisco: Holden-Day (revised edn., 1976).
- 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.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19, 1179.
- 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 (2015). 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 December 5, 2015).
- Hyndman, R. and Y. Khandakar (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software* 27(3), 1–22.
- Oeppen, J. et al. (2008). Coherent forecasting of multiple-decrement life tables: a test using japanese cause of death data. In *Compositional Data Analysis Conference*. Universitat de Girona. Departament d’Informàtica i Matemàtica Aplicada.
- 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 nonparametric approach. *Demographic Research* 25, 595–628.
- Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.
- R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). Springer-Verlag.
- Tableau, E. (2001). A review of demographic forecasting models for mortality. In *Forecasting Mortality in Developed Countries*, pp. 1–32. Springer.
- 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.
- Wrycza, T. F. and A. Baudisch (2014). The pace of aging: Intrinsic time scales in demography. *Demographic Research* 30, 1571.

Wrycza, T. F., T. I. Missov, and A. Baudisch (2015). Quantifying the shape of aging.
PloS one 10(3), e0119163.