

Modeling and Forecasting Age at Death Distributions: A Nonparametric Approach

Ugofilippo Basellini^{1,2}, Carlo Giovanni Camarda¹, and Vladimir Canudas-Romo²

¹Institut National d'Études Démographiques, Paris

²Max-Planck Odense Center on the Biodemography of Aging, Odense

December 15, 2016

Abstract

Age at death distributions have generally been neglected for modeling and forecasting mortality, despite being very well suited to study longevity, lifespan variation and mortality developments. In this article, we propose to use age at death distributions to model the entire pattern of mortality and to inform mortality forecasts. In particular, we introduce a novel methodology that allows us to disentangle, model and forecast the three independent age-specific components of mortality corresponding to young, middle and old ages, respectively. The methodology is based on: (i) a nonparametric decomposition of death counts for the estimation of the three mortality components, and (ii) a segmented linear transformation of the age axis for modeling and forecasting the evolution of each component, which allows capturing compression and shifting dynamics of mortality. We will illustrate our methodology by estimating and forecasting the mortality pattern and life expectancy at birth of females and males in high-longevity countries of the Human Mortality Database.

1 Introduction

Population projections and mortality forecasts have been studied since the beginning of the twentieth century. The seminal works of [Whelpton \(1928, 1936\)](#) and [Lotka \(1939\)](#) on the cohort component method and the stable population contributed significantly to the development and application of the former. The latter have been produced by actuaries at least since 1924 due to the adverse financial effects of mortality improvements on life annuities and pensions ([Pollard, 1987](#)). However, it is only in the last thirty years that more sophisticated statistical methodologies to forecast mortality have been proposed and used ([Booth and Tickle, 2008](#)).

In general, three main functions can be used in the study of human mortality: the hazard, the survival and the probability density function ([Klein and Moeschberger, 2005](#)). Although these three functions describe the same stochastic phenomenon and are uniquely related between each other, the vast majority of forecasting techniques are based on models of the hazard function or of age-specific mortality rates (for example, [Cairns et al., 2006](#); [Currie et al., 2004](#); [Lee and Carter, 1992](#); [Li and Lee, 2005](#)).

Mortality rates have been predominantly used to portray mortality developments because they readily represent the change in the risk of death over age and time ([Camarda, 2008](#)). However, the direct investigation of mortality rates does not provide an immediate answer to two key questions in mortality studies, that is, how long does a population live on average, and how variables are ages at death.

In addition to providing an informative description of the mortality experience of a population, the probability density function (also called age at death distribution, or distribution of ages at death)

yields readily available information on the “central longevity indicators” (mean, median and modal age at death, [Canudas-Romo, 2010](#); [Cheung et al., 2005](#)) as well as on lifespan variability. Nevertheless, this function has generally been neglected in modeling and forecasting, as only few attempts have been made to explicitly use it to forecast mortality (for example, [Basellini and Camarda, 2016](#); [Bergeron-Boucher et al., 2016](#); [Oeppen, 2008](#)).

In this paper, we propose a novel methodology for modeling and forecasting mortality that is based on age at death distribution. This article is organized as follows. In Section 2, we overview the data and mathematical methods that we will use in this article. In Section 3, we summarize our proposed methodology and briefly compare it with other approaches.

2 Data & Methods

2.1 Data

The dataset that will be employed in this paper is the [Human Mortality Database \(HMD, 2016\)](#), which provides free access to historical mortality data for 43 different territories and countries (mostly European, but also covering North America, Russia, Japan and Oceania). The HMD is an important collection of detailed, consistent and high quality human mortality data that were subject to a uniform set of procedures, which allow cross-national comparability of the information ([Barbieri et al., 2015](#)).

2.2 Mortality Functions & Decomposition

Mortality can be effectively summarized by any one of three complementary functions: $l(x)$, the probability of surviving from birth to age x (the survival function); $\mu(x)$, the force of mortality at age x (the hazard function); and $d(x)$, the distribution of ages at death in the life table population at age x (the probability density function). The three functions are uniquely related between each other, and knowing one of them allows to determine the other two. For example, $d(x)$ can be expressed in terms of $\mu(x)$ as:

$$d(x) = l(x)\mu(x) = \left[l(0)e^{-\int_0^x \mu(a)da} \right] \mu(x), \quad (1)$$

where $l(0)$ is the radix of the life table (which may equal one by definition, although it is usually some large, arbitrary number like 100,000) ([Preston et al., 2001](#)).

The supposition that human mortality can be decomposed into three different groups that operates principally, or almost exclusively, upon childhood, middle and old ages, respectively, goes back at least one and a half century ([Thiele, 1871](#)), and it has been extensively used since then for modeling purposes ([Heligman and Pollard, 1980](#); [Siler, 1979](#)). In formal terms, the hypothesis can be expressed as follows:

$$\mu(x) = \mu_I(x) + \mu_A(x) + \mu_S(x), \quad (2)$$

where the force of mortality $\mu(x)$ at age x is additively decomposed into three independent components, $\mu_I(x)$, $\mu_A(x)$, and $\mu_S(x)$, which represent infant mortality, adult mortality and senescent mortality, respectively. The assumption of three independent $\mu_i(x)$, where $i = I, A, S$, can be usefully applied to Equation (1) to decompose the distribution of ages at death:

$$\begin{aligned} d(x) &= l(x)\mu(x) = l(x) [\mu_I(x) + \mu_A(x) + \mu_S(x)] \\ &= l(x)\mu_I(x) + l(x)\mu_A(x) + l(x)\mu_S(x) = d_I(x) + d_A(x) + d_S(x). \end{aligned} \quad (3)$$

Equation (3) shows that $d(x)$ can be decomposed into three independent components $d_i(x)$, each associated with its corresponding $\mu_i(x)$ by the formula $d_i(x) = l(x)\mu_i(x)$, where $i = I, A, S$.

2.3 Estimation

An approach to estimate the three independent age-specific mortality components $\mu_i(x)$ (and consequently $d_i(x)$) is thus required. In demographic and actuarial applications, an assumption often encountered is that the total number of deaths D_x at age x follows a Poisson distribution, that is, $D_x \sim \text{Poisson}(E_x \cdot \mu_x)$, where E_x is the number of person-years exposed to the risk of dying (Brillinger, 1986).

Camarda et al. (2016) have recently proposed a nonparametric approach to decompose complex series of counts that can be successfully applied to human mortality. Given a series of death counts, the Sum of Smooth Exponentials (SSE) model allows to isolate and estimate each mortality components $\mu_i(x)$ without imposing any rigid parametric mortality structures. Among others, one advantage of the SSE methodology is that it adequately blends the transitions between components, without imposing sharp delimitations where one stops and another one continues.

2.4 Modeling & Forecasting

Having estimated the three components $d_i(x)$, the next step is modeling and forecasting the evolution of each one of them. Here, we describe the approach for doing so by focusing on the senescent component $d_S(x)$ (from age 30 onwards, as in Basellini and Camarda, 2016).

The Segmented Transformation Age at Death Distributions (STAD) model relates a fixed “standard” to an observed distribution by a transformation of the age axis uniquely dependent on the modal ages at death (M) and the variability of the two distributions. Let $f(x)$ denote the standard, $g(x)$ the observed distribution, and $t(x)$ the transformation that conforms the density of $g(x)$ with the one of $f(x)$ on the transformed axis, that is, $g(x) = f(t(x))$. Then, the STAD model can be expressed as:

$$t(x; s, b_L, b_U) = \begin{cases} M^f + b_L(x - s - M^f) & \text{if } x \leq M^g \\ M^f + b_U(x - s - M^f) & \text{if } x > M^g \end{cases} \quad (4)$$

where $s = M^g - M^f$ denotes the difference between the mode of $g(x)$ and $f(x)$, and b_L and b_U denote the change in the variability of $g(x)$ with respect to $f(x)$ before and after the mode, respectively. A graphical example is helpful to better explain the segmented linear transformation of the STAD model. In the right panel of Figure 1, the black line corresponds to $f(x)$.

First, consider the case of a simple shift in the distributions (red line of the graphs). The $f(x)$ is shifted to the left, maintaining the same variability before and after the modal age at death. This is a special case of Equation (4), in which $b_L = b_U = 1$ and the transformation function becomes $t_0(x) = x - s$.

Different values of b_L and b_U which act jointly with the shifting parameter s correspond to broader mortality changes. When b_L and b_U are bigger than 1, the ages before and after the mode of $g(x)$ are shrunk, and variability in ages at death is decreasing with respect to $f(x)$. On the other hand, when b_L and b_U are smaller than 1, the ages before and after M^g are expanded, so that age at death variability increases compared to the standard. In Figure 1, the ages before the mode are expanded and the ages above are shrunk (purple line of the graphs). The three parameters of the STAD model thus capture the shifting (s) and compression (b_L and b_U) dynamics of mortality (Bongaarts, 2005; Wilmoth and Horiuchi, 1999).

Moving from the theoretical description of the model to its actual application for modeling and forecasting a series of age at death distributions over time, the first step is the choice of the standard distribution $f(x)$. While several options are available, Basellini and Camarda (2016) suggest to first align the observed distributions, and then to choose the mean as standard. The landmark

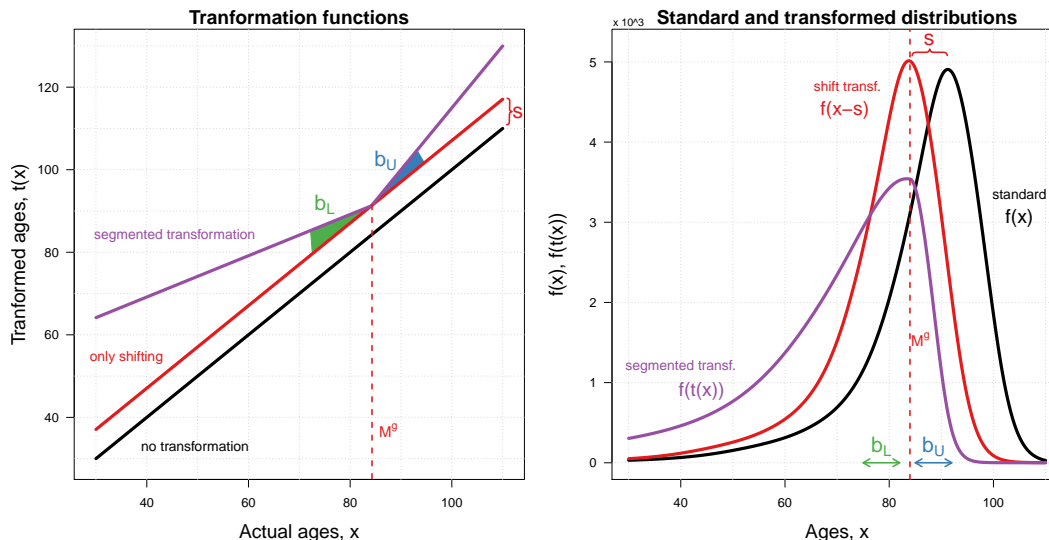


Figure 1: A schematic overview of the effects of transforming the age axis using the STAD model.

registration technique (Ramsay and Silverman, 2005) indeed allows not to mix up features of the observed distributions that occur at different distances with respect to the mode.

Having fixed the standard distribution, the three parameters can be estimated and forecast, and forecast age at death distributions can be thus derived. The estimation procedure is performed by maximum likelihood, following the Poisson assumption described in Section 2.3. Parameters are then modelled and forecast using the standard class of univariate time-series autoregressive integrated moving average (ARIMA) models (Box and Jenkins, 1970).

As concluding remark, it should be noted that a modification of the STAD model will be necessary for the analysis of the first two components $d_I(x)$ and $d_A(x)$, in which the shifting mortality dynamic may be absent or only very minor (and fixed within a small age range).

3 Discussion

Age at death distributions have generally been neglected for modeling and forecasting mortality, despite being well suited to study longevity, lifespan variation and mortality developments. In this article, we propose to model and forecast mortality by studying changes that have occurred in age at death distributions. In particular, we aim to provide a comprehensive methodology that captures the entirety of the mortality pattern. In order to do so, two steps are required and currently envisioned: (i) the disentanglement and estimation of the three independent age-specific mortality components, and (ii) the modeling and forecast of each component. Recently proposed methodologies for both issues will be used, modified as needed and adopted in combination. Analyses will be performed on freely available data from the Human Mortality Database.

The decomposition of the distribution of deaths has already precedents in the literature. For example, Pearson (1948) developed a method for expressing the “curve of death” by a combination of three independent Normal distributions covering different periods of life. More recently, Zanotto et al. (2016) proposed a mixture three-component parametric model with response variable the life table distribution at deaths to study changes of premature mortality over time. However, the methodology proposed here is the first attempt, according to our knowledge, to decompose the distribution of deaths to model and forecast mortality developments. In addition, our approach differentiates from the other two as it is completely nonparametric, thus lacking any assumption on strict parametric mortality structures.

References

- Barbieri, M., J. R. Wilmoth, V. M. Shkolnikov, D. Gleij, D. Jasilionis, D. Jdanov, C. Boe, T. Riffe, P. Grigoriev, and C. Winant (2015). Data Resource Profile: The Human Mortality Database (HMD). *International Journal of Epidemiology*, dyv105.
- Basellini, U. and C. G. Camarda (2016). Modeling and forecasting age at death distributions. In *European Population Conference*.
- Bergeron-Boucher, M.-P., V. Simonacci, M. Gallo, and J. Oeppen (2016). Coherent modeling and forecasting of mortality patterns for subpopulations using multi-way analysis of compositions: An application to Canadian provinces. In *Annual Meeting of the Population Association of America*.
- Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography* 42(1), 23–49.
- 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).
- Brillinger, D. R. (1986). A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. *Biometrics*, 693–734.
- 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. (2008). *Smoothing methods for the analysis of mortality development*. Ph. D. thesis, Universidad Carlos III de Madrid, Department of Statistics.
- Camarda, C. G., P. H. Eilers, and J. Gampe (2016). Sums of smooth exponentials to decompose complex series of counts. *Statistical Modelling*, 1471082X16641796.
- Canudas-Romo, V. (2010). Three measures of longevity: Time trends and record values. *Demography* 47(2), 299–312.
- 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.
- Currie, I. D., M. Durban, and P. H. Eilers (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4(4), 279–298.
- Heligman, L. and J. H. Pollard (1980). The age pattern of mortality. *Journal of the Institute of Actuaries* 107(01), 49–80.
- Human Mortality Database (2016). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.
- Klein, J. P. and M. L. Moeschberger (2005). *Survival Analysis: Techniques for Censored and Truncated Data* (Second ed.). 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.
- Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42(3), 575–594.
- Lotka, A. J. (1939). Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particuliere alespece humaine. *Actualités Scientifiques et Industrielles* (780).
- Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. In *Compositional Data Analysis Conference*.
- Pearson, K. (1948). *Karl Pearson's early statistical papers*. Cambridge University Press.
- 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*. Blackwell.
- Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). Springer-Verlag.

- Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 750–757.
- 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.
- Whelpton, P. K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, 253–270.
- Whelpton, P. K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association* 31(195), 457–473.
- Wilmoth, J. R. and S. Horiuchi (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography* 36(4), 475–495.
- Zanotto, L., V. Canudas-Romo, and S. Mazzucco (2016). Evolution of premature mortality. In *European Population Conference*.