IS LIFE EXPECTANCY ENOUGH? ESTIMATING AGE-SPECIFIC MORTALITY RATES BY CONSIDERING THE RELATIONSHIP BETWEEN SILER MODEL PARAMETERS AND LIFE EXPECTANCY AT BIRTH.

Lucia Zanotto

Abstract. Understanding mortality in detail, with respect to age, sex and country, is crucial for comprehending this complex phenomenon. Indirect estimation has emerged as a valid approach to derive analytical values from a synthetic index, allowing us to move beyond the narrow focus on life expectancy at birth that characterises many forecasting models. By exploiting the correlations between the parameters of the Siler model and life expectancy at birth, it becomes possible to reconstruct age-specific mortality rates across the entire lifespan, from age 0 to 100. Preliminary results indicate small estimation errors and a coherent, smooth and easily interpretable fit.

1 Introduction

Life expectancy at birth (usually denoted by e_0) is the most widely used index to summarise and communicate the longevity of a population or its subgroup. It represents the average number of years a newborn is expected to live given the mortality conditions at a specific point in time. However, there are critical issues with the use of e_0. For instance, it does not consider the health status or other characteristics, such as infant and premature mortality levels or inequality at death. Despite these limitations, its value is easily obtained and calculated. This has led to the development of forecasting models focused solely on estimating life expectancy (Pascariu *et al.*, 2018; Raftery *et al.*, 2014; Torri and Vaupel, 2012). These very parsimonious techniques have demonstrated high accuracy in predictions by taking advantage of the linear trend in life expectancy observed over the past 50 to 60 years (Oeppen and Vaupel, 2002; White, 2002). As a result, the need for indirect estimates arose: models that could reconstruct precise life tables based on a given level of e_0. For instance, the U.S. Census Bureau derives period age-specific death rates through projections of life expectancy at birth by sex and race (United States Census Bureau, 2014). Ševčíková *et al.* (2016) demonstrated how estimated projections for e_0 could be converted into age-specific death rates

using the Lee-Carter model. A similar approach was proposed by Pascariu et al. (2020). Alternatively, Nigri *et al.* (2022) suggested employing a deep neural network model to obtain age-specific mortality from observed or predicted life expectancy. While the concept of modeling life tables is not new in the literature (see for example Coale and Demeny, 1966), this paper explores a new, simple, and parsimonious methodology for reconstructing specific mortality rates from life expectancy at birth. The underlying idea builds upon previous studies: an identical average level of mortality (i.e., equal life expectancy at birth) corresponds to the same curve of specific mortality rates, without differentiating between countries and sexes. Given that this curve can be approximated using the Siler model (Siler, 1979, 1983), the same value of e_0 is associated with the same vector of parameters. Consequently, by fitting appropriate functions (not necessarily linear) between e_0 and estimated coefficients, it becomes possible to approximate mortality throughout the entire lifespan. In this paper, we present the first analyses concerning the reconstruction of mortality by exploiting the relationships between Siler model parameters and life expectancy at birth, discussing the advantages and disadvantages of the proposed approach.

2 Method and Data

The model proposed by Siler (Siler, 1979, 1983) to estimate mortality rates at age $x \in (0, \omega)$ consists of three mathematical functions: a negative Gompertz function for infant mortality, a positive constant (known as Makeham's constant (Makeham, 1860)) to account for deaths occurring randomly with respect to age, and a Gomperz function (Gompertz, 1825) for adult/senesent mortality. The specific death rate at age x, denoted as mx, is described by the following formula

$$m_x = a_1 e^{-b_1 x} + a_2 + a_3 e^{b_3 x}, x > 0. \quad (1)$$

In equation 1, parameter $a_1 > 0$ represents the intensity of infant mortality, and $b_1 > 0$ the rate of its decline, $a_3 > 0$ describes the initial size of mortality at approximately age 30 and $b_3 > 0$ expresses the rise of the mortality rate over ages.
Maximum likelihood estimation was used to calculate the model parameters. It assumes that the number of deaths observed in a given age interval follows a Poisson distribution

$$P(D = d; \theta) = f(d; \theta) = \frac{1}{d!}\theta^d e^{-\theta}, \quad (2)$$

where θ is the rate parameter and d is the number of deaths occurred in a given interval of time. For each age interval, we have information about the number of people exposed to the risk of dying, Ex, and the number of deaths that occurred, Dx. Assuming that deaths observed at a given age interval are independent from deaths occurred in other one, the likelihood and log-likelihood functions are as follows

$$L(\theta;\, Ex, Dx) = \prod_{x=0}^{\omega} \frac{1}{D_x!} \theta^{D_x} e^{-E_x \theta};$$
$$l(\theta;\, Ex, Dx) \propto \sum_{x=0}^{x=0} D_x \log(\theta) - E_x \theta. \tag{3}$$

The Siler model is estimated by substituting θ with the hazard function of the equation 1.

To illustrate and test the proposed methodology, mortality rates from Sweden and the United States of America in the Human Mortality Database (Human Mortality Database, 2023) were utilized. The HMD contains high quality data that undergone standardised procedures, ensuring the comparability of information across different populations. The two selected countries exhibit notable differences: while both have small infant mortality rates, the USA has a lower life expectancy and experiences higher levels of premature (or mid-life) mortality (Case and Deaton, 2015, 2017). In contrast, Sweden is characterized by its high longevity. The chosen time period spans from 1950 (after the Second World War and the associated economic crises) to 2019 (the year preceding the Covid-19 pandemic). As the HMD often applies various correction procedures and models for old age mortality based on the country (Wilmoth and Horiuchi, 1999), this analysis focuses on ages 0 to 100.

Once the Siler model was estimated, the relationship between its five parameters (on a logarithmic scale) and the life expectancy at birth for the corresponding country and year was examined. Figure 1 displays a clear linear patterns between e_0 and a_1, as well as between e_0 and a_3 (although the values are more dispersed in the first graph, the Pearson correlation coefficient remains high). In contrast, the relationship between life expectancy at birth and a_3 is visibly quadratic. No apparent relationship is observed between e_0 and b_1, particularly for the United States. The value of b_1 for the years considered remains relatively

stable, albeit at two different levels for males and females. In the case of Sweden, a fluctuating trend is observed, potentially due to model identification issues, as highlighted by Mazzuco *et al.* (2018). Finally, both Swedish males and females show decreasing values of life expectancy at birth and Makeham's constant, whereas for the United States, although the trend is downward, the reduction is considerably smaller for both sexes. Looking at the relationship between e_0 and a_2 as a whole, however, there does not seem to be a clear pattern. Considering these observations, the following functions were defined to estimate each parameter of the Siler model, using life expectancy as an explanatory variable:

$$\log(a_{1,t}) = a_{1.1} + a_{1.2}e_{0,t} + \varepsilon_t; \quad (4)$$

$$\log(b_{1,t}) = b_{2.1} + \varepsilon_t; \quad (5)$$

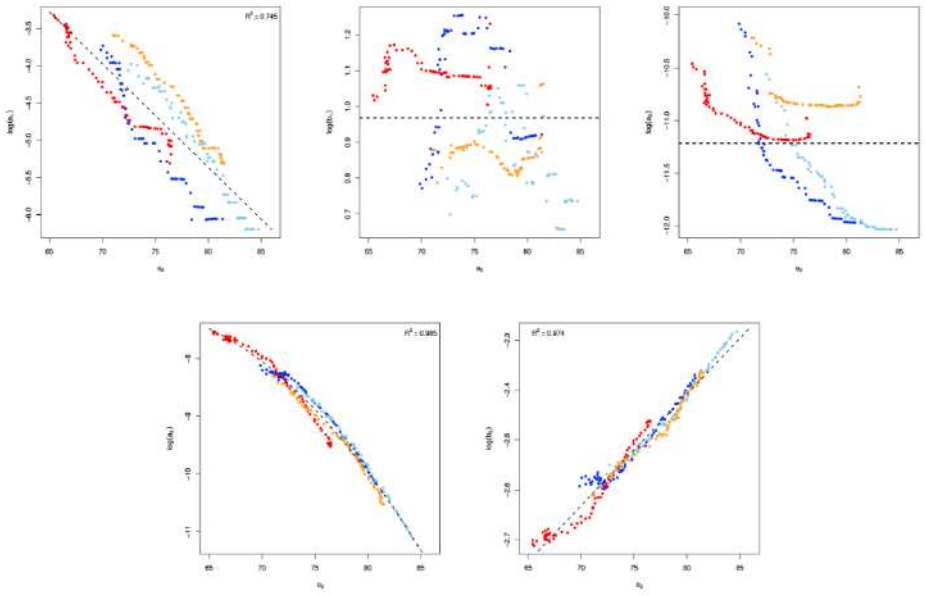
$$\log(a_{2,t}) = a_{2.1} + \varepsilon_t; \quad (6)$$

$$\log(a_{3,t}) = a_{3.1} + a_{3.2}e_{0,t} + a_{3.3}e_{0,t}^2 + \varepsilon_t; \quad (7)$$

$$\log(b_{3,t}) = b_{3.1} + b_{3.2}e_{0,t} + \varepsilon_t; \quad (8)$$

where ε_t are independent and identical random variables normally distributed with mean of zero and variance of σ^2. The logarithmic transformation of the parameters is particularly convenient to avoid issues in their parametric reference space. The linear determination indeces, R^2, for models 4, 7 and 8 are 0.75, 0.99 and 0.97, respectively.

Figure 1 – *The relationship between life expectancy at birth, e_0, and parameters of the Siler model on a logarithmic scale. Warm colors (red for males, orange for females) represent the USA, while cool colors (blue for males, light blue for females) represent Sweden. The dashed lines depict the selected model for estimating the parameters using life expectancy at birth and its coefficient of linear determination, R2, is indicated in the corresponding panels.*



Once the parameters are estimated and the life expectancy at birth for the population of interest is selected, the parameters of the Siler model can be derived, and specific mortality rates for each age interval can be calculated. Although the specification of the functions for b_1 and a_2 may not be entirely convincing (and error assumptions are not fully satisfied), the overall estimations of the Siler model are reasonably accurate, as shown in section 3.

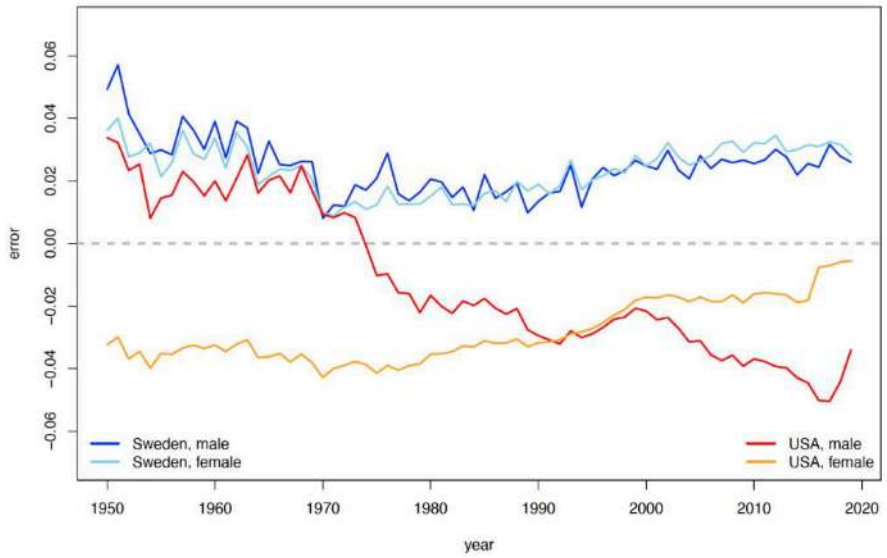
3 Results

In order to compare the classical estimates (obtained year-by-year via maximum likelihood of the equation 3) with those calculated using direct parameter estimations, we analyzed the age-specific death rates for ages 0 to 100 obtained

through both methodologies. The error measure is expressed as a percentage, reflecting the difference between the estimated and original log-death rates.

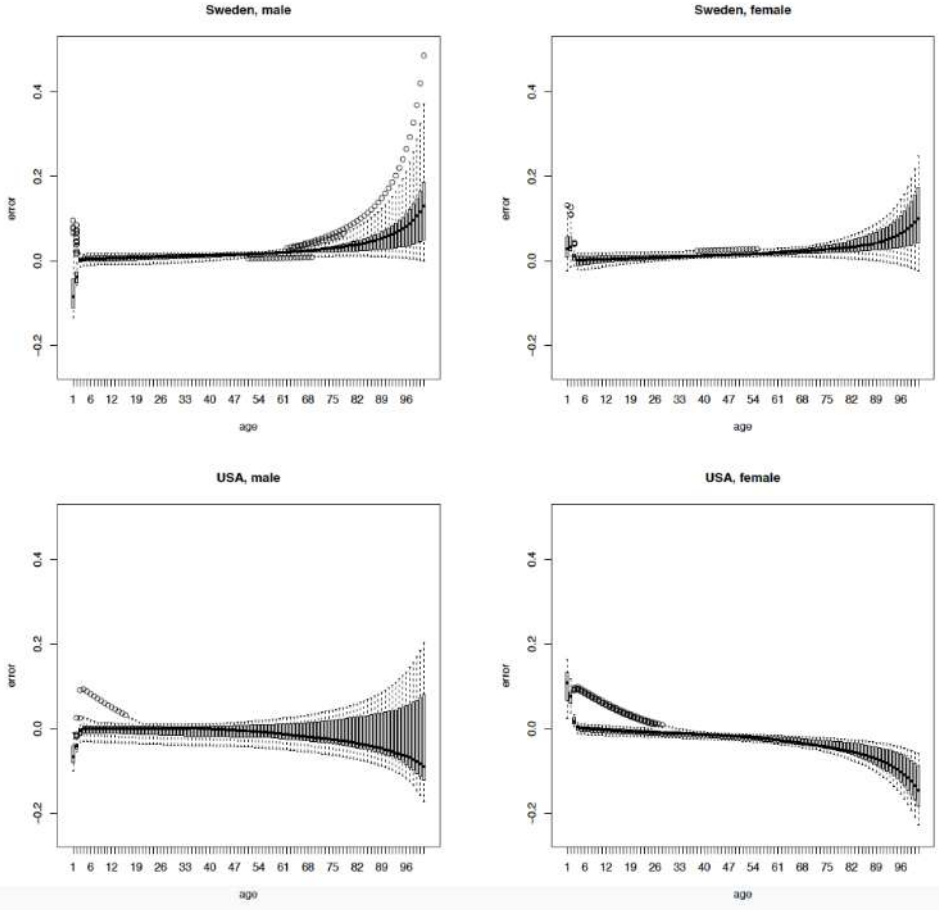
Figure 2 illustrates that the average errors is range between -6% and 6%. Hence, the differences are relatively small for both countries. However, for Sweden (both sexes), the proposed methodology tends to overestimate the death rates. On the contrary, American women consistently exhibit an underestimation, although this tendency decreases in more recent years. In contrast, American men demonstrate the highest differences during the later years of the series.

Figure 2 - *The mean relative errors between the estimated and original log- death rates from 1950 to 2019.*



When examining the errors across the lifespan (see Figure 3), death rates at early and advanced ages are more difficult to estimate. Specifically, Swedish men display an increase in error variability with advancing age. Despite exhibiting similar trends, the discrepancies remain minor for the other scenarios. However, considering the medians, they are very close to t 0 (no errors) for almost all ages. Figure 3 shows that in Sweden, the model tends to overestimate mortality rates beyond the age of 80, whereas in the USA, it tends to underestimate them.

Figure 3 – *Distribution of relative errors between the estimated and original log-death rates across different age groups.*



4. Discussion

By leveraging the relationships between life expectancy at birth and Siler parameter estimates, we were able to reconstruct specific mortality rates for ages 0 to 100, without any additional information about sex and country. Although the link functions between coefficients and e_0 are not always entirely satisfactory, the approximation achieved is generally satisfactory, with small errors observed for both Sweden and the USA.

When comparing this approach to methods proposed by Pascariu et al. (2020) and Nigri et al. (2022), which calculate each mortality rate using specific linear or

non-linear equations, it is worth noting that the estimates obtained using the Siler model yield smoother rate curves but with lower precision. For instance, the Siler model cannot account for the accidental hump often observed for males around the age of 20. However, it is important to consider that the new methodology is much more parsimonious than the approaches described by the aforementioned authors, which generally leads to greater imprecision. Nonetheless, a comparison among the different methodologies is desirable.

The accuracy of the suggested method relies on the quality of the year-by- year estimates obtained through maximum likelihood, as well as the precision of the link functions. In the latter case, exploring additional populations is crucial. Analyzing countries that have experienced a sharp decline in infant mortality since the Second World War (such as Italy and Spain) and nations where it is still significant (e.g., Russia) would provide valuable insights. Moreover, estimating other datasets could further enhance our ability to model (parametrically or not) the relationships between parameters and e_0, particularly for b_1 and a_2.

Conducting, additional tests is therefore essential to validate this approach, but the initial results are certainly promising. In response to the question "is life expectancy enough" we can confidently answer in the affirmative.

Acknowledgements

This work is funded by the Piano Nazionale di Ripresa e Resilienza (PNRR) -- Progetto PE8 -- AGE-IT ``A novel public-private alliance to generate socioeconomic, biomedical and technological solutions for an inclusive italian ageing society".

References

CASE A. AND DEATON A. 2015. Rising morbidity and mortality in midlife among white non-hispanic americans in the 21st century. *Proceedings of the National Academy of Sciences,* 112 (49), 15078–15083.

CASE A. AND DEATON. A. 2017. Mortality and morbidity in the 21st century. *Brookings papers on economic activity* 397.

COALE A. AND DEMENY P. 1966. *Regional Model Life Tables and Stable Populations*. Princeton University Press.

GOMPERTZ B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London* (115), 513–583.

HUMAN MORTALITY DATABASE. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). *https://www.mortality.org/* (accessed on May 2023).

MAKEHAM W. M. 1860. On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries,* Vol. 8, No. 6, pp. 301–310.

MAZZUCO S., SCARPA B. ZANOTTO L. 2018. A mortality model based on a mixture distribution function, *Population Studies, Vol. 72, No.* 2, pp. 191–200.

NIGRI A., LEVANTESI S., ABURTO J.M. 2022. Leveraging deep neural networks to estimate age-specific mortality from life expectancy at birth, *Demographic Research,* Vol. 47, pp. 199–232.

OEPPEN J. AND VAUPEL J. W. 2002. Broken limits to life expectancy, *Science,* Vol. 296, No. 5570, pp. 1029–1031.

PASCARIU M. D., BASELLINI U, ABURTO J.M. AND CANUDAS-ROMO V. 2020. The linear link: Deriving age-specific death rates from life expectancy, *Risks,* Vol. 8, No. 4, pp. 109.

PASCARIU M. D., CANUDAS-ROMO V. AND VAUPEL J. W. 2018. The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics,* Vol. 78, pp. 339–350.

RAFTERY A. E., LALIC N. AND GERLAND P. 2014. Joint probabilistic projection of female and male life expectancy. *Demographic research,* Vol. 30, 795.

ŠEVČÍKOVÁ H., KANTOROVÁ V., GERLAND P., AND RAFTERY A. E. 2016. Age- specific mortality and fertility rates for probabilistic population projections, *Dynamic demographic analysis*, pp. 285–310.

SILER W. 1979. A competing-risk model for animal mortality, *Ecology,* Vol. 60, No. 4, 750–757.

SILER W. 1983. Parameters of mortality in human populations with widely varying life spans, *Statistics in medicine,* Vol. 2, No. 3, pp. 373–380.

TORRI T. AND VAUPEL J. W. 2012. Forecasting life expectancy in an international context. *International Journal of Forecasting,* Vol. 28, No. 2, pp. 519–531.

UNITED STATES CENSUS BUREAU. 2014. Methodology, assumptions, and inputs for the 2014 national projections. *Washington, DC: U.S. Census Bureau*.

WHITE K. M. 2002. Longevity advances in high-income countries, 1955–96, *Population and Development Review, Vol. 28, No.* 1, pp. 59–76.

WILMOTH J. R. AND HORIUCHI S.. 1999. Rectangularization revisited: Variability of age at death within human populations. *Demography,* Vol. 36 No. 4, pp. 475–495.

Lucia ZANOTTO, Department of Statistics "Paolo Fortunati", University of Bologna, lucia.zanotto@unibo.it