An age-at-death distribution approach to forecast cohort mortality

Extended abstract for the PAA 2019 conference

Ugofilippo Basellini[1,2], Søren Kjærgaard[2], and Carlo Giovanni Camarda[1]

[1]Institut national d'études démographiques (INED), Paris
[2]Center on Population Dynamics (CPop) and Department of Public Health, University of Southern Denmark, Odense

September 19, 2018

Abstract

Mortality forecasting has received increasing attention during recent decades due to the negative financial effects of longevity improvements on public and private institutions' liabilities. However, little work has been done on forecasting mortality from a cohort perspective. In this article, we introduce a novel methodology to forecast adult age-specific cohort mortality from age-at-death distributions. We show that cohort forecasts can improve our understanding of mortality developments, as well as allow us to complete the mortality experience of partially observed cohorts. We illustrate our methodology on adult female cohort mortality in two high-longevity countries using data from the Human Mortality Database.

1 Introduction

Continuous and widespread gains in life expectancy (Riley, 2001; Oeppen and Vaupel, 2002) are increasingly challenging governments and insurance companies to provide adequate pension products and elderly health care in ageing societies. Mortality forecasting has thus gained relevant prominence during the last decades, as relatively small differences in the expected lifetimes of pensioners can have significant effects on financial institutions' liabilities.

A growing number of models have recently been proposed to forecast human mortality using stochastic methodologies that allow to draw probabilistic statements about the future. For a comprehensive review, see Booth (2006). Unsurprisingly, the vast majority of these techniques are based on *period* mortality: indeed, financial institutions are typically interested in the mortality developments of a group of individuals composed by different birth cohorts. In addition, cohort data might be outdated, unavailable or incomplete, hence period life tables have been developed to analyse a hypothetical cohort if its age-specific death rates pertained throughout its life (Preston et al., 2001).

Cohort forecasts of mortality are nonetheless interesting and worth exploring for two main reasons. First, survival in real birth cohorts is different from survival in the hypothetical situation of unchanged period mortality rates because of: (i) tempo effects, (ii) cohort effects and (iii) selection (for a full discussion, see Borgan and Keilman, 2018, Sect. 2). As such, analysing and forecasting cohort mortality could provide additional insights on mortality developments that could not be discerned from the period perspective. Second, due to the nature of the data, cohort forecasts can be employed to complete the mortality experience of partially observed cohorts.

In this article, we propose a novel methodology to forecast adult cohort mortality that is based on the distribution of deaths. Age-at-death distributions have recently received growing interest in mortality forecasting (Oeppen, 2008; Bergeron-Boucher et al., 2017; Pascariu et al., 2017; Basellini and Camarda, 2018). As we will show, the distribution of deaths has convenient features that makes it particularly suitable to forecast cohort mortality; we thus propose a new extension of a recently proposed methodology to model and forecast adult age-at-death distributions (Basellini and Camarda, 2018), and we employ our model to analyse and forecast cohort mortality.

This paper is organized as follows. In Section 2, we review the methodology proposed in this article, and describe the data used in the analysis. Section 3 shows the results for Swedish and Swiss female adult mortality for the cohorts 1880-1950. In Section 4, we discuss the results and future steps of our research.

2 Methods & Data

2.1 The C-STAD Model

Suppose we have two adult age-at-death distributions defined on the age range $x \geq 40$. Specifically, let $f(x)$ be a "standard", or reference, distribution and $g(x)$ an observed distribution. Let $t(x;\boldsymbol{\theta})$ be a transformation function of the age axis and a vector of parameters $\boldsymbol{\theta}$ such that:

$$g(x) = f\left[t(x;\boldsymbol{\theta})\right] , \tag{1}$$

i.e. the distribution $g(x)$ is derived from a warping transformation of the age axis of $f(x)$. In particular, the term "warping" originates in Functional Data Analysis (Ramsay and Silverman, 2005) and refers to the nonlinear transformation of a time axis to achieve close alignment of functions.

Our goal is to find a parsimonious yet flexible transformation function $t(x;\boldsymbol{\theta})$ that allows us to capture mortality developments rigorously. Let $\boldsymbol{\theta}' = [s, b_L, c_L, d_L, b_U]$ be the vector of our model's parameters, where $s = M^g - M^f$ denotes the difference between the adult modal ages at death of $g(x)$ and $f(x)$. Then, the proposed *Cubic Segmented Transformation Age-at-death Distributions* (C-STAD) model can be written as:

$$t(x;\boldsymbol{\theta}) = \begin{cases} M^f + b_L\,\tilde{x} + c_L\,\tilde{x}^2 + d_L\,\tilde{x}^3 & \text{if } x \leq M^g \\ M^f + b_U\,\tilde{x} & \text{if } x > M^g \end{cases} \tag{2}$$

where $\tilde{x} = x - s - M^f$, and the subscripts $_L$ and $_U$ refer to the *Lower* and *Upper* parts of the segmented transformation (i.e. before and after M^g), respectively.

The warping function $t(x;\boldsymbol{\theta})$ takes the form of a segmented transformation model which breaks at the value of M^g. Below M^g, the transformation function is cubic, while it is linear above M^g. Although acting on $t(x;\boldsymbol{\theta})$, the model's parameters are directly related to different summary measures of the age-at-death distributions. Specifically, while s captures the difference in modal ages between $g(x)$ and $f(x)$, b_L and b_U measure the change in variability before and after the modal ages of the two distributions. For the ages below M^g, c_L and d_L further measure differences in terms of asymmetry and heaviness of the left tail between $g(x)$ and $f(x)$, respectively. In terms of mathematical moments, the parameters b_L and b_U can be related to the variance of the distribution before and after the mode, while c_L and d_L can be related to the skewness and kurtosis of the distribution.

Figure 1 provides a graphical illustration of the C-STAD model. For simplicity, we start from the simple case in which $b_L = b_U = 1$ and $c_L = d_L = 0$. Substitution of these parameters in Eq. (2) yields a unique transformation function $t(x, \boldsymbol{\theta}) = x - s$, and a corresponding distribution $g_1(x) = f(x - s)$ via Eq. (1). In the left panel of Fig. 1, the standard distribution (grey line) is shifted to the right by an amount equal to s, and $g_1(x)$ (blue line) maintains the original shape of $f(x)$. The right panel shows the transformation function related to this simple shifting scenario. Note that a left-shift could be simply obtained with a negative value of s.

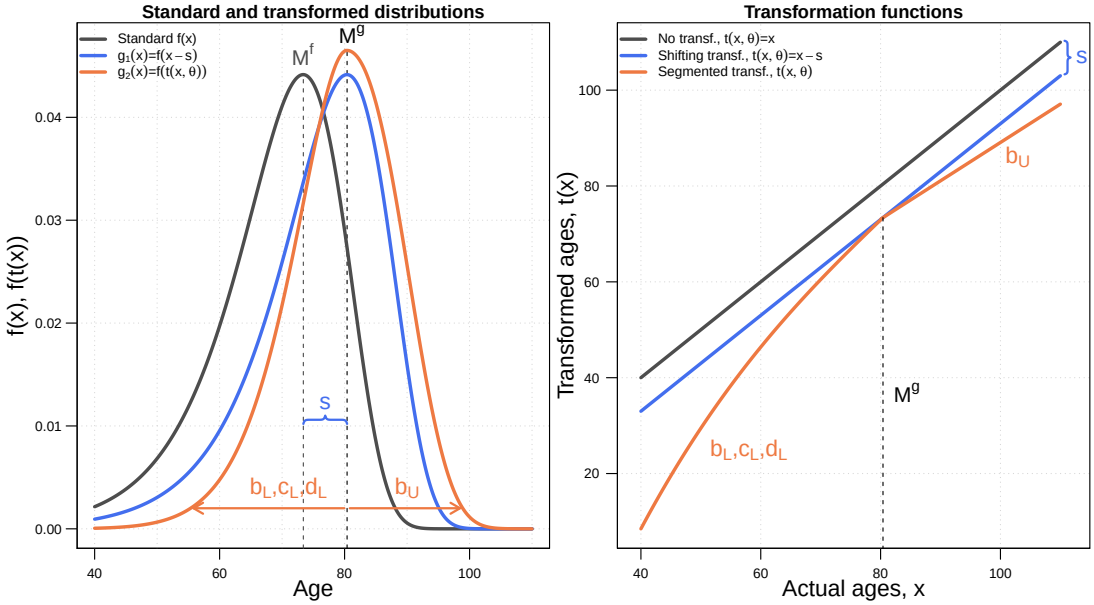

Figure 1. A schematic overview of the *Cubic Segmented Transformation Age-at-death Distributions* model.

Different parameters' values allow to capture broader mortality developments than the simple shifting scenario described above. While b_L and b_U modify the variability of the distribution $g_2(x)$ (orange line, left panel) w.r.t. $f(x)$ before and after the modal age at death, c_L and d_L affect the asymmetry and heaviness of the left tail of $g_2(x)$ as compared to $f(x)$. In the example shown in Fig. 1, $b_L > 1$ reduces the variability of $g_2(x)$ before M^g w.r.t. $f(x)$, while $b_U < 1$ increases the variability of $g_2(x)$ after M^g w.r.t. $f(x)$. The effects of c_L and d_L are difficult to discern from the left panel. However, the right panel shows the warping transformation $t(x, \boldsymbol{\theta})$ applied to $f(x)$ to

derive $g_2(x)$; the transformation (orange line) is composed by a cubic function (due to non-zero values of c_L and d_L) before the cut-off point M^g, and by a linear function above M^g.

2.2 Data

In this article, we aim to estimate and forecast adult cohort mortality for females in two high-longevity countries, namely Sweden and Switzerland. Since our interest is restricted to the senescent component of mortality, we start our analyses from age 40. Specifically, we employ observed death counts $D_{x,c}$ and exposure-to-risk $E_{x,c}$, classified by age at death $x = 40, \ldots, 110+$ for cohorts born in $c = 1880, \ldots, 1950$, collected in two matrices of m rows and n columns.

Estimation and forecasting of the C-STAD model, discussed in the following subsection, is performed on three different groups of cohorts. The first group contains the fully observed cohorts $c_1 = 1880, \ldots, 1905$, i.e. for this group, D_{x,c_1} and E_{x,c_1} have been observed at all ages x. We select 1880 as the first year of analysis to have a sufficiently long time window for fitting the model. The second group is composed by cohorts $c_2 = 1906, \ldots, \tilde{c}$, where $\tilde{c}$ denotes the last cohort for which the adult modal age at death has been observed. In other words, D_{x,c_2} and E_{x,c_2} are incomplete, but D_x and E_x have been observed at least until the age $x = M$ for all cohorts in c_2. For the populations analysed here, $\tilde{c}$ is 1925 and 1924 for Sweden and Switzerland, respectively. Finally, the third group is composed by cohorts $c_3 = \tilde{c} + 1, \ldots, 1950$, for which D_{x,c_3} and E_{x,c_3} have been observed up to some age $x < M$. Figure 2 below provides an illustration of the divisions of cohorts into the three groups by means of a Lexis diagram. Figure 3 shows an example of the observed and missing data for three age-at-death distributions belonging to the different groups of cohorts.

The data are derived from the Human Mortality Database (HMD, 2018), which provides free access to historical mortality data for 43 different territories and countries. The HMD is a 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).

2.3 Estimation and forecast of the C-STAD parameters

The derivation of the standard distribution $f(x)$ is the first step in the fitting procedure of the C-STAD model. In order to do so, we start by considering only the group of fully observed cohorts c_1 discussed beforehand. Since it is desirable that $f(x)$ contains all the representative features of the fully observed distributions, we adopt a landmark registration procedure to derive $f(x)$ as the mean of the aligned distributions (Basellini and Camarda, 2018).

Specifically, we assume that death counts $D_{x,c}$ at given age x and cohort c follow a Poisson distribution:

$$D_{x,c} \sim \mathcal{P}(E_{x,c}\, \mu_{x,c}) \tag{3}$$

where $E_{x,c}$ is the exposure-to-risk and $\mu_{x,c}$ the hazard or force of mortality (Brillinger, 1986). In order to align the observed distributions to the one of the first cohort (1880 in

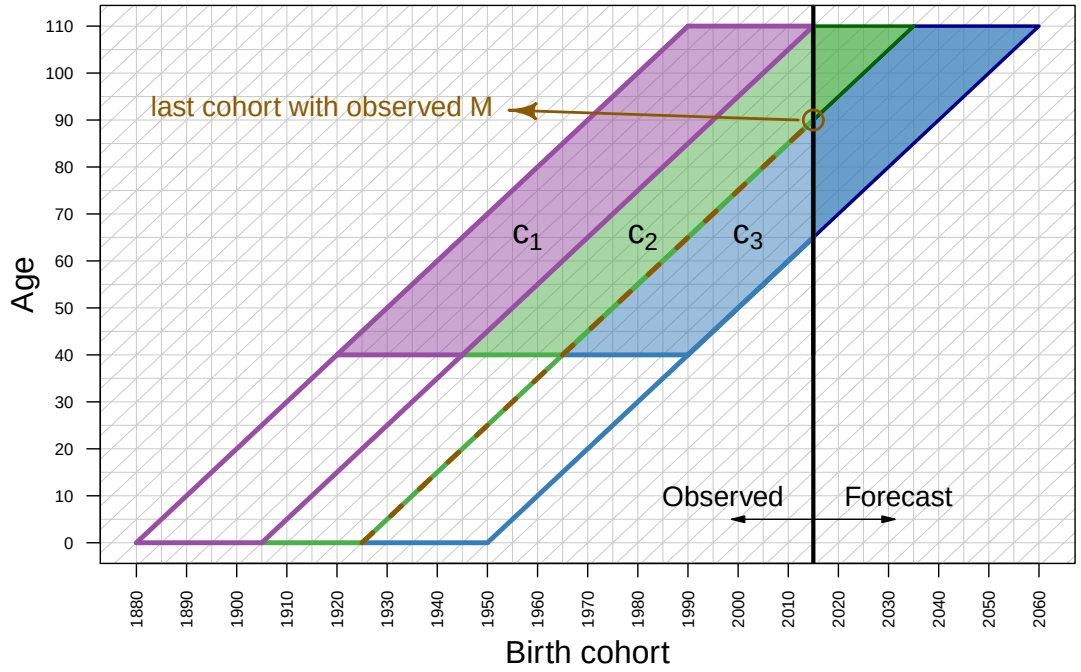

Figure 2. A Lexis diagram illustrating the division of cohorts into three groups. 2015 is the most recent year of data collection, while the last cohort with observed M is $\tilde{c} = 1925$ ($M = 2015 - \tilde{c} = 90$, brown dashed line). The three groups are then $c_1 = 1880, \ldots, 1905$, $c_2 = 1906, \ldots, 1925$ and $c_3 = 1926, \ldots, 1950$.

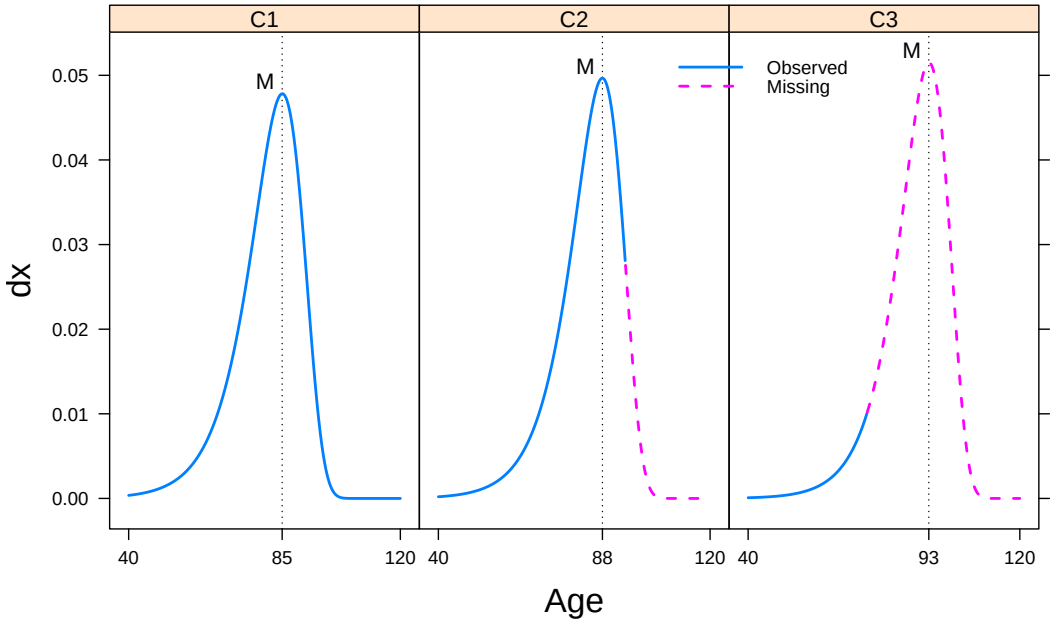

Figure 3. Example of observed and missing data for three age-at-death distributions belonging to the groups of cohorts c_1, c_2 and c_3.

our case), we smooth observed death counts using a P-spline approach and exposures as an offset (Eilers and Marx, 1996; Camarda, 2012). This procedure allows us to: (i) derive the modal age at death M^c for each cohort c (ii) obtain an estimate of the parameter $\hat{s}_c = M^c - M^{1880}$, (iii) align all observed distribution $g_c(x)$ w.r.t. $g_{1880}(x)$, and thus derive the standard $f(x)$ as the registered mean (for additional details, see Basellini and Camarda, 2018). Note that $f(x)$ can be expressed as a linear combination of equally spaced B-spline basis $\boldsymbol{B}(x)$ over ages x and coefficients $\boldsymbol{\beta}_f$ specific to the standard distribution:

$$f(x) = \exp\left[\boldsymbol{B}(x)\,\boldsymbol{\beta}_f\right] . \tag{4}$$

Next, we estimate the parameters $\tilde{\boldsymbol{\theta}}'_c = [b_L, c_L, d_L, b_U]$ for each cohort c in the first group c_1 (we remove the subscript c inside the vector for simplicity). This is achieved by maximising the Poisson log-likelihood:

$$\ln \mathcal{L}\left(\tilde{\boldsymbol{\theta}}_c | \, D_{x,c}, E_{x,c}, s_c, \boldsymbol{\beta}_f, \boldsymbol{w}\right) \propto \sum_x \boldsymbol{w}\left[D_{x,c} \log\left(\mu_{x,c}^{\texttt{C-STAD}}\right) - E_{x,c}\,\mu_{x,c}^{\texttt{C-STAD}}\right] \tag{5}$$

for $c = 1880, \ldots, 1905$, where $\boldsymbol{w}$ is a vector of weights of length m, and $\mu_{x,c}^{\texttt{C-STAD}}$ denotes the estimated hazard of the C-STAD model. Each element of $\boldsymbol{w}$ is equal to one if the corresponding death and exposure data are observed, and zero otherwise. For c_1, all elements of $\boldsymbol{w}$ are equal to one. In words, the optimization procedure looks for a combination of parameters $\hat{\boldsymbol{\theta}}'_c = \left[\hat{s}_c, \hat{\tilde{\boldsymbol{\theta}}}'_c\right]$ for each cohort that produces an age-at-death distribution $\hat{g}_c(x) = f\left[t(x; \hat{\boldsymbol{\theta}}_c)\right]$ whose corresponding hazard maximises the log-likelihood in Eq. (5).

For the second group of cohorts, we employ the same estimation procedure, with only a small modification. On one hand, since the modal age at death has been observed, there are no problems in the estimation of the parameter s_c. On the other hand, $D_{x,c}$ and $E_{x,c}$ are not observed at all ages; Eq. (5) can nonetheless be computed by giving zero weights to the unobserved data. This can be achieved by replacing zeros to the elements of $\boldsymbol{w}$ corresponding to the missing data. Two things are worth noting here. First, the missing data only influence the estimation of b_U, as complete data are observed before the modal age for all cohorts in this group. Second, the estimate $\hat{\boldsymbol{\theta}}_c$ allows us to derive a complete set of age-specific mortality measures, i.e. we can complete the mortality experience of the partially observed cohorts in c_2.

Finally, for the third group of cohorts, we forecast the C-STAD parameters using the estimates of the first two groups. The lack of knowledge about the modal age at death indeed makes it impossible to estimate the parameter s_c from the partially observed data, compute the log-likelihood function in Eq. (5) and estimate the remaining parameters. As such, we specify two different multivariate vector autoregressive time-series models for the C-STAD parameters. The first model is fitted to the parameters s and b_U. From a theoretical perspective, these parameters are related by the fact that mortality changes occurring above the mode could modify its value (see Canudas-Romo, 2010, Appendix B). The second model is specified for b_L, c_L and d_L: these three parameters are theoretically related to each other since they pertain to the same segment of the age-at-death distribution. The results for the countries analysed indeed confirm these hypotheses, showing high correlation between the two

set of parameters and low correlation between the other combinations (see Fig. A.1 in Appendix A).

3 Results

3.1 Out-of-sample validation of the C-STAD model

Before completing the mortality experience of partially observed cohorts, we first assess the accuracy of the C-STAD model by performing six predictive out-of-sample validation exercises on Swedish and Swiss adult females. Specifically, we pretend that the last year of collected data is $2015-\delta$, where $\delta = 10, 15, 20, 25, 30$ and 35 years. We then fit the C-STAD model to the fully observed cohorts $c_1 = (1880, \dots, 1905) - \delta$: we modify the start and the end cohorts of the fitting period to keep its length constant across the six exercises. We then forecast mortality δ years ahead, and we compare the forecast life expectancy at age 40 (e_{40}) and the Gini coefficient at age 40 (G_{40}) with the observed values.

An explicative example of this procedure is useful to clarify the out-of-sample exercises. Let us consider $\delta = 10$: then, the last year of fully observed data is 2005. We thus fit the C-STAD to the fully observed cohorts $c_1 = 1870, \dots, 1895$, and we forecast 10-year ahead. By doing so, we complete the mortality experience of the partially observed cohorts $1896, \dots, 1905$, and for each of these, we compute and compare the estimated e_{40} and G_{40} with the observed ones.

It is worth mentioning at this point that, for the lower values of δ, forecasting is achieved simply by fitting the C-STAD on the partially observed cohorts c_2. In the explicative example above, where the last data collection occurred in 2005, the cohort 1896, for instance, has been observed at all ages except 110. We thus take advantage of the nature of cohort data and consider all possible observations to complete the mortality experience of this partial cohort. Conversely, for higher values of δ, forecasting is achieved by considering also the cohorts c_3, which require the use of the VAR models.

Table 1 presents the results of our analysis. The first and second columns contain the cohorts used for fitting and forecasting the C-STAD, respectively. The third column contains the time horizon of the out-of-sample exercise, while the fourth column the measure analysed (e_{40} or G_{40}). Results are shown in the last two columns. We assess the accuracy of the point forecasts by computing the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{c=1}^{\delta} (\hat{y_c^{\delta}} - y_c^{\delta})^2}{\delta}} \tag{6}$$

where δ is the forecasting horizon, and $\hat{y_c^{\delta}}$ and y_c^{δ} are the forecast and observed values of either e_{40} or G_{40}. For this analysis, we multiplied the values of G_{40} by 100 in order to have comparable magnitude between the two indicators.

The table shows that the C-STAD forecasts are very accurate in completing the mortality experience of partially observed cohorts. The RMSE values of the two indicators are very low, especially for the first four forecast horizons. The accuracy slightly

Fitting cohorts	Forecast cohorts	Horizon	Measure	SWE	CHE
1870-1895	1896-1905	10y	e_{40}	0.08	0.05
			G_{40}	0.04	0.02
1865-1890	1891-1905	15y	e_{40}	0.08	0.05
			G_{40}	0.05	0.02
1860-1885	1886-1905	20y	e_{40}	0.07	0.04
			G_{40}	0.07	0.04
1855-1880	1881-1905	25y	e_{40}	0.05	0.03
			G_{40}	0.12	0.06
1850-1875	1876-1905	30y	e_{40}	0.07	0.56
			G_{40}	0.12	0.55
1845-1870	1871-1905	35y	e_{40}	0.25	0.66
			G_{40}	0.29	0.69

Table 1. Root mean squared errors of the C-STAD forecasts of e_{40} and G_{40} for adult females in Sweden (SWE) and Switzerland (CHE) in six out-of-sample validation exercises: forecast horizon of 10, 15, 20, 25, 30 and 35 years.

decreases in the last exercise (and also in the 30-year scenario for Switzerland), but the errors remain relatively small in magnitude. Very similar results would be obtained if we employed a different prediction accuracy measure (e.g. MAPE, MAE).

It is finally worth mentioning that the results are robust to the choice of the fitting period. Indeed, we performed a sensitivity analysis by modifying the fitting period in each exercise. Specifically, we fixed 1860 as the starting cohort for each exercise, and we fitted the C-STAD on the available fully observed cohorts (e.g. 1860-1895 for the first exercise, 1860-1870 for the last one). The results of this analysis are very similar to those shown in Table 1.

3.2 Completing the mortality experience of partially observed cohorts

Here, we show the results of employing the C-STAD model to estimate and forecast adult female cohort mortality in Sweden and Switzerland for the cohorts 1880-1950. The fitted and forecast parameters for the two countries are reported in Appendix A, together with the correlation analysis of the time-series of the estimated parameters.

Figure 4 shows the observed and fitted remaining life expectancies at age 40 (e_{40}) and Gini coefficient at age 40 (G_{40}) in the two population analysed for the fully observed cohorts 1880-1905. From the graphs, it emerges that the C-STAD estimates are very close to the observed values for both measures in the two populations.

Figure 5 shows the observed (cohorts c_1), completed (c_2) and forecast (c_3) e_{40} and G_{40} for the two population analysed. The derivation of prediction intervals for the C-STAD forecasts is foreseen as our next research step (see Discussion in Section 4).

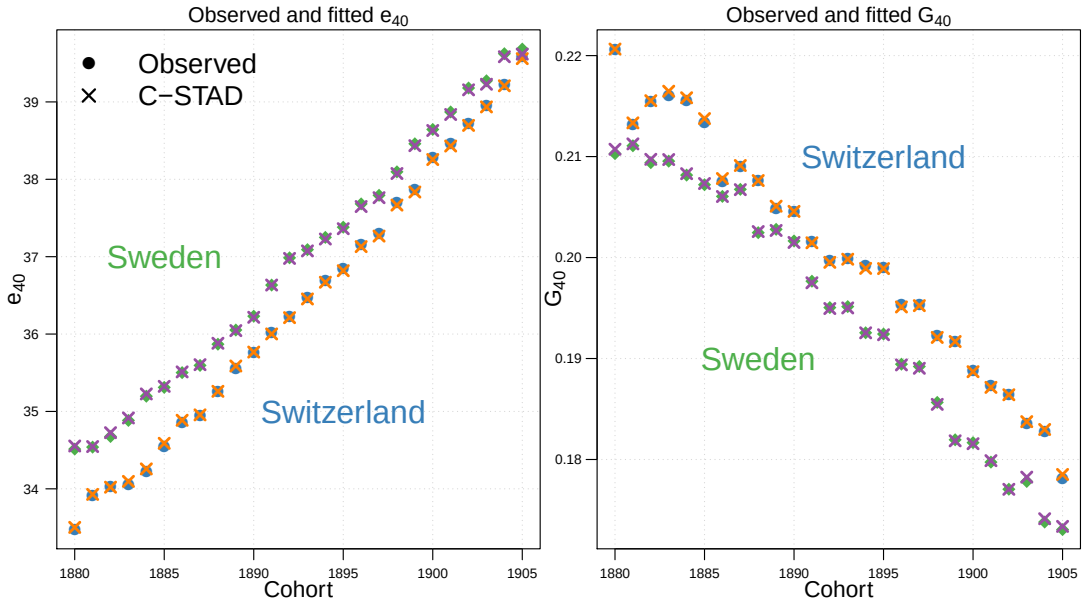

Figure 4. Observed and C-STAD estimated remaining life expectancies at age 40 (e_{40}, left panel) and Gini coefficient at age 40 (G_{40}, right panel) for adult females in Sweden and Switzerland for the fully observed cohorts 1880-1905.

4 Discussion

Mortality forecasting has drawn considerable interest in recent decades among academics and financial sector practitioners due to the increasing challenges posed by population ageing. Advances in the field have almost exclusively been made on period mortality, as the most recent and innovative techniques are based on modelling and forecasting different functions of period life tables.

In this article, we take an alternative perspective and introduce a new methodology to model and forecast adult cohort mortality. Our approach focuses on cohort age-at-death distributions: specifically, we propose a parametric warping transformation of a standard distribution to describe and forecast mortality developments. The warping function takes the form of a cubic segmented transformation, hence called *Cubic Segmented Transformation Age-at-death Distributions* (C-STAD) model.

We have shown the results of fitting and forecasting mortality with the C-STAD model for Swedish and Swiss adult females for the cohorts 1880-1950. Our methodology is accurate from a point forecast perspective: for each population, we performed six out-of-sample validation exercises of different forecast horizons. The resulting point forecast errors as measured by the RMSE are generally small, even for the longer forecast horizons.

Models to forecast cohort mortality are relatively few because of heavy data demand; however, this problem is reduced when only adult mortality is considered (Booth, 2006). Our interest in this article is restricted to adult mortality, so this issue does not affect us to a great extent. In addition, we believe that forecasts of cohort mortality are interesting beyond period forecasts in two regards: (i) cohort mortality

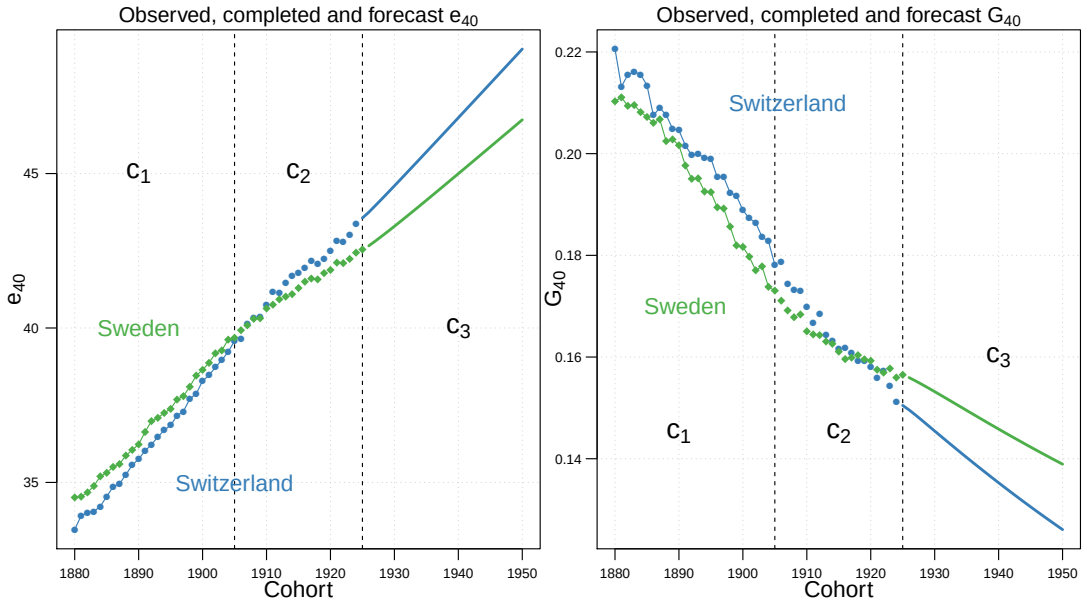

Figure 5. Observed (cohorts c_1), C-STAD completed (c_2) and forecast (c_3) remaining life expectancies at age 40 (e_{40}, left panel) and Gini coefficient at age 40 (G_{40}, right panel) for adult females in Sweden and Switzerland for the cohorts 1880-1905.

developments are *actually* observed, and thus they differ from the hypothetical ones assumed in period life tables, and (ii) cohort forecast can be exploited to complete the mortality experience of partially observed cohorts.

Our approach is inspired by the Segmented Transformation Age-at-death Distributions (STAD) model recently proposed by Basellini and Camarda (2018) to forecast adult age-at-death distributions. In addition to shifting the focus from period to cohort mortality, our methodology extends the STAD to a cubic transformation before the modal age at death. The additional parameters c_L and d_L are indeed necessary to adequately describe cohort mortality developments at young adult ages. A possible explanation for this could be the significant improvements in maternal mortality across the cohorts that we analyse. Non-linear functions above the mode were tested too, but they did not provide a better fit compared to a linear function. We plan to investigate in more details the reasons for the need of the additional parameters at lower ages.

Future work is currently foreseen along different directions. First, we will derive prediction intervals for the C-STAD forecasts. A residual bootstrapping approach is under development for this purpose. The derivation of prediction intervals will allow us to complement the point forecast accuracy analysis in Section 3.1 with the assessment of interval forecast accuracy. Second, we will compare the C-STAD forecasts with those derived from other approaches. A possible candidate for this comparison is a cohort adaptation of the two-dimensional smoothing and forecasting P-splines method proposed by Currie et al. (2004). Third, we will include cohort mortality rates as a third indicator to assess forecasts' accuracy in the out-of-sample exercises.

In addition to this, we are currently working on different approaches to estimate b_L, c_L and d_L using the partially observed data in c_3, taking the forecast s and b_U as

given inputs. It would be indeed desirable to use the observed data to estimate rather than forecast b_L, c_L and d_L in c_3. Conversely, the current VAR model for b_L, c_L and d_L does not consider the observed data in c_3 (see Figure 2, light blue area). Finally, we plan to: (i) perform sensitivity analyses on the choice of the cut-off cohort $\tilde{c}$ between c_2 and c_3; (ii) compute the C-STAD forecast summary measures of *period* mortality implied by the cohort forecasts, and compare them with those of other approaches, such as the Lee and Carter (1992) model; and (iii) extend these analyses to other populations.

References

Barbieri, M., Wilmoth, J. R., Shkolnikov, V. M., Glei, D., Jasilionis, D., Jdanov, D., Boe, C., Riffe, T., Grigoriev, P., and Winant, C. (2015). Data Resource Profile: The Human Mortality Database (HMD). *International Journal of Epidemiology*, 44(5):1549–1556.

Basellini, U. and Camarda, C. G. (2018). Modelling and forecasting adult age-at-death distributions. *Population Studies*, Forthcoming.

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

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

Borgan, Ø. and Keilman, N. (2018). Do Japanese and Italian Women Live Longer than Women in Scandinavia? *European Journal of Population*.

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

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. (2010). Three measures of longevity: Time trends and record values. *Demography*, 47(2):299–312.

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

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, pages 89–102.

Human Mortality Database (2018). 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 31 July 2018).

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

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

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

Pascariu, M., Lenart, A., and Canudas-Romo, V. (2017). Forecasting mortality by using statistical moments. In *13th International Longevity Risk and Capital Markets Solutions Conference*.

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

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

Riley, J. C. (2001). *Rising life expectancy: a global history*. Cambridge University Press.

A Section 3.2: additional results

Figure A.1 shows the absolute correlation of the estimated C-STAD parameters in first differences for Swedish and Swiss adult females for cohorts 1880-$\tilde{c}$, where $\tilde{c}$ is 1925 and 1924 for Sweden and Switzerland, respectively. The graph shows that the parameters of the two VAR models, s and b_U on one hand, and b_L, c_L and d_L on the other, are highly related to each other (light blue colours). Conversely, the correlation between the other combinations of parameters is rather low (dark blue colours).

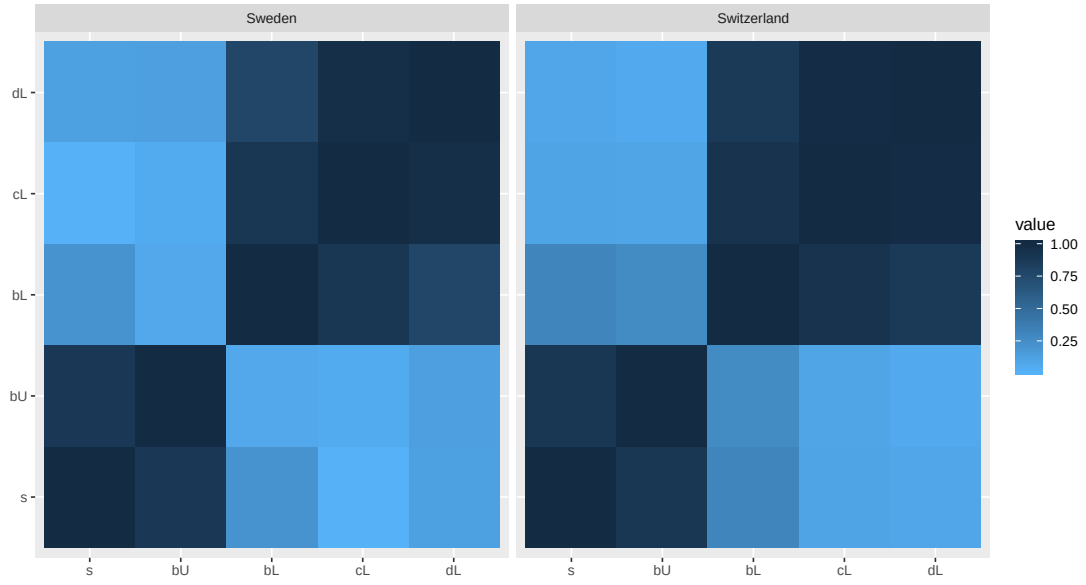

Figure A.1. Absolute correlation of the C-STAD parameters estimates (in first differences) for Swedish and Swiss adult females for cohorts 1880-$\tilde{c}$.

Figure A.2 shows the fitted and forecast C-STAD parameters for Swedish and Swiss adult females for cohorts 1880-1950.

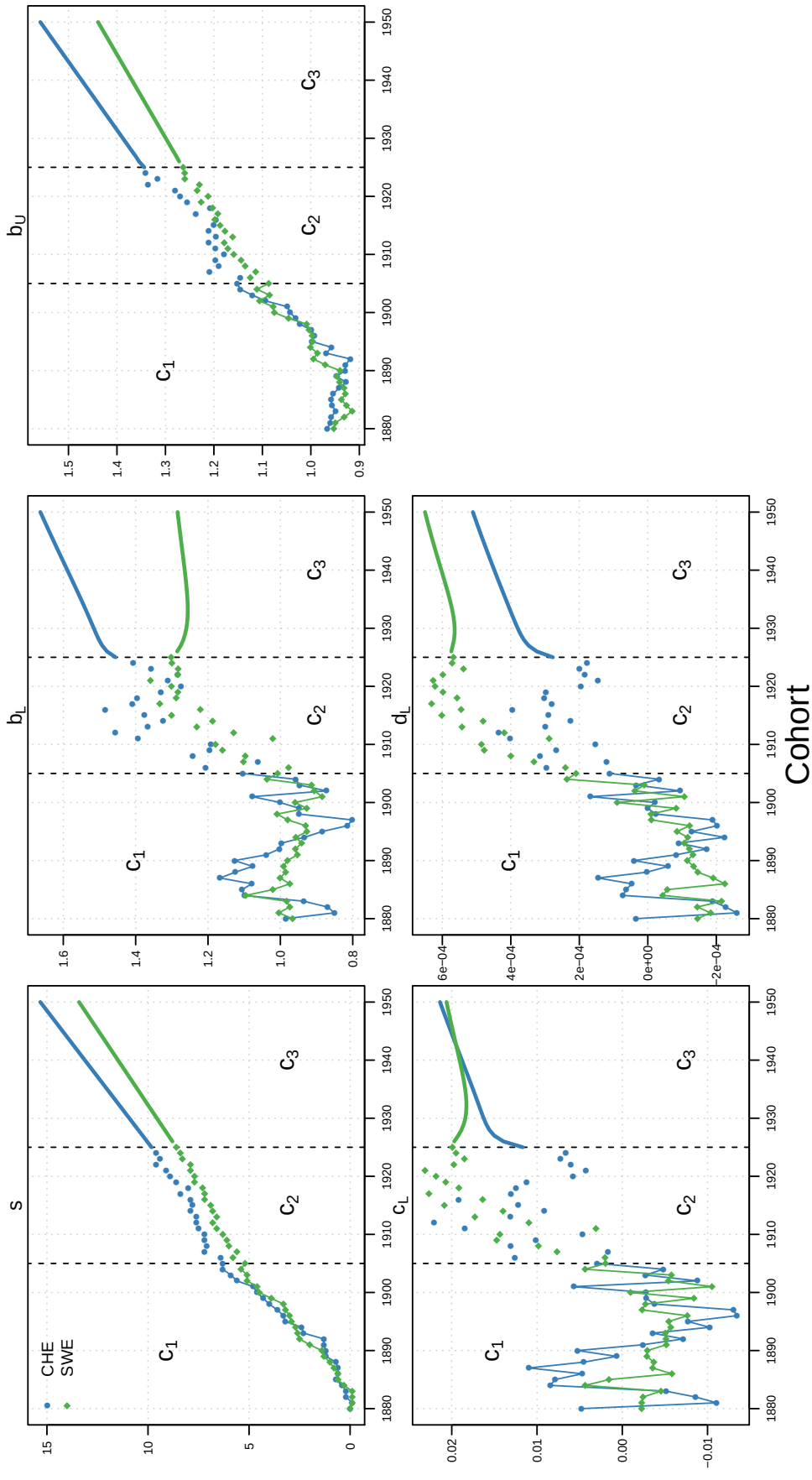

Figure A.2. Fitted (points with line for cohorts c_1, points for cohorts c_2) and forecast (lines) C-STAD parameters for adult females in Sweden and Switzerland for the cohorts 1880-1950.