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

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

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

June 17, 2019

Abstract

Mortality forecasting has received increasing interest during recent decades due to the negative financial effects of continuous longevity improvements on public and private institutions' liabilities. However, little attention has been paid to forecasting mortality from a cohort perspective. In this article, we introduce a novel methodology to forecast adult cohort mortality from age-at-death distributions. We propose a relational model that associates a time-invariant standard to a series of fully and partially observed distributions. Relation is achieved via a transformation of the age-axis. We show that cohort forecasts can improve our understandings of mortality developments by capturing peculiar cohort effects, which might be overlooked by a conventional age-period perspective. Moreover, mortality experience of partially observed cohorts are routinely completed. We illustrate our methodology on adult female mortality for cohorts born between 1835 and 1970 in two high-longevity countries using data from the Human Mortality Database.

Keywords: Mortality forecasting · Mortality modelling · Relational models · Cohort life table · Smoothing

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 produce probabilistic assessments of the future. For comprehensive reviews, see Booth (2006) and Shang et al. (2011). Unsurprisingly, the vast majority of these techniques are based on *period* mortality: financial institutions are typically interested in the mortality developments of groups of individuals that comprise different birth cohorts. In addition, cohort data can 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 can provide different insights on mortality developments than analyses based on 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 increasing attention in mortality forecasting (Oeppen, 2008; Bergeron-Boucher et al., 2017; Basellini and Camarda, 2019; Pascariu et al., 2019). 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, 2019), 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 the data that we employ for the analyses. Section 3 presents two applications of our model to Swedish and Danish female adult mortality for the cohorts 1835-1970. In Section 4, we discuss the results of our methods and conclude.

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 transformation of a time axis to achieve close alignment of functions.

We propose a parsimonious yet flexible transformation function $t(x;\boldsymbol{\theta})$ that captures adult mortality developments across cohorts rigorously. Let $\boldsymbol{\theta}' = [s, b_L, c_L, d_L, b_U]$ be a vector containing the model's parameters, where $s = M^g - M^f$ is the difference between the modal ages at death of the unimodal distributions $g(x)$ and $f(x)$. The proposed *Cohort Segmented Transformation Age-at-death Distributions* (C-STAD) model

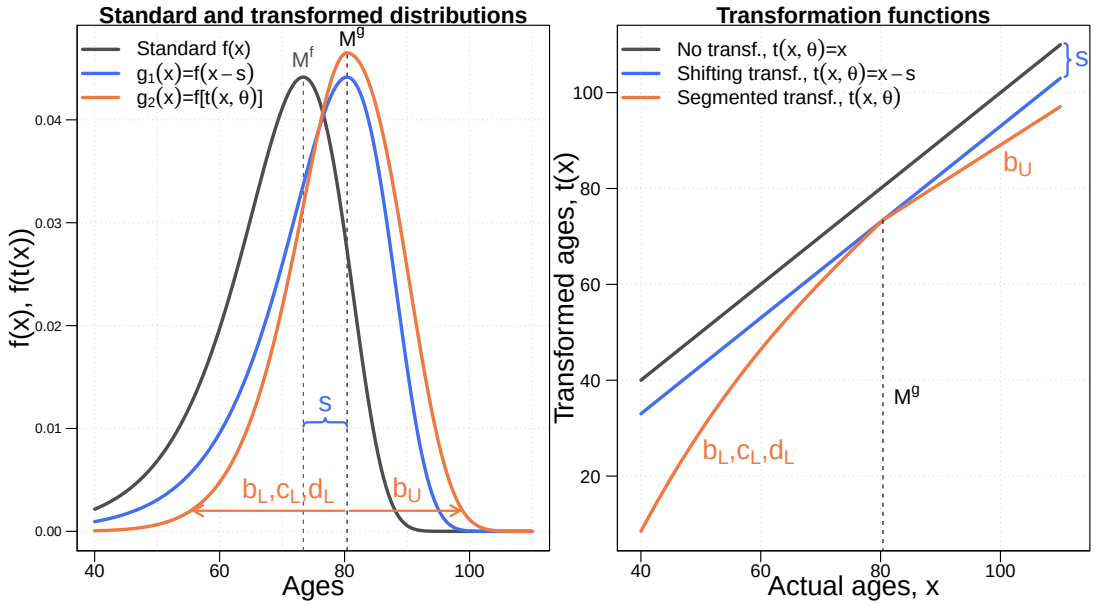

Figure 1. A schematic overview of the *Cohort Segmented Transformation Age-at-death Distributions* model.

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 the 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 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.

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 $g_2(x) = f\left[t(x;\boldsymbol{\theta})\right]$ (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 Denmark. Long-term series of high quality data are available for both countries, even at the very old ages (Vaupel and Lundström, 1994; Wilmoth and Lundström, 1996)[XXX add reference for DNK data], and the two countries display different mortality developments. We therefore test the goodness-of-fit and forecast accuracy of our model with respect to different mortality trajectories. The data are derived from the Human Mortality Database (HMD, 2019), 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).

Our interest in this article is restricted to the senescent component of mortality, hence we start our analyses from age 40. Specifically, we employ two mxn matrices $D_{x,c}$ and $E_{x,c}$ containing observed death counts and exposure-to-risk, respectively, classified by age at death $x = 40, \ldots, 110+$ and birth cohort $c = 1835, \ldots, 1970$. On one hand, we select 1835 as starting cohort of analysis for both populations because it is the first cohort with observed data at all ages in Denmark, and to compare the results for the two countries using the same fitting period. On the other hand, 1970 is the final cohort because it contains enough observed data points in both countries (seven in Sweden, six in Denmark) to estimate the three low parameters accurately (cf. Section 2.3).

Estimation and forecasting of the C-STAD parameters (Section 2.3) is performed on three different groups of cohorts. The first group contains the fully observed cohorts $c_1 = 1835, \ldots, 1905$, i.e. for this group, D_{x,c_1} and E_{x,c_1} have been observed at all ages x. The second group is composed by cohorts $c_2 = 1906, \ldots, \tilde{c}$, where $\tilde{c}$ denotes the last cohort for which two age-groups above the adult modal age at death have 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 for at least two data points above the age $x = M$ for all cohorts in c_2. Also here, the choice of two age groups is dictated by the estimation of the parameter above the mode (cf. Section 2.3). For the two populations analysed here, $\tilde{c}$ is 1925 and 1927 for Sweden and Denmark, respectively. Finally, the third group is composed by the remaining cohorts $c_3 = \tilde{c} + 1, \ldots, 1970$. 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.

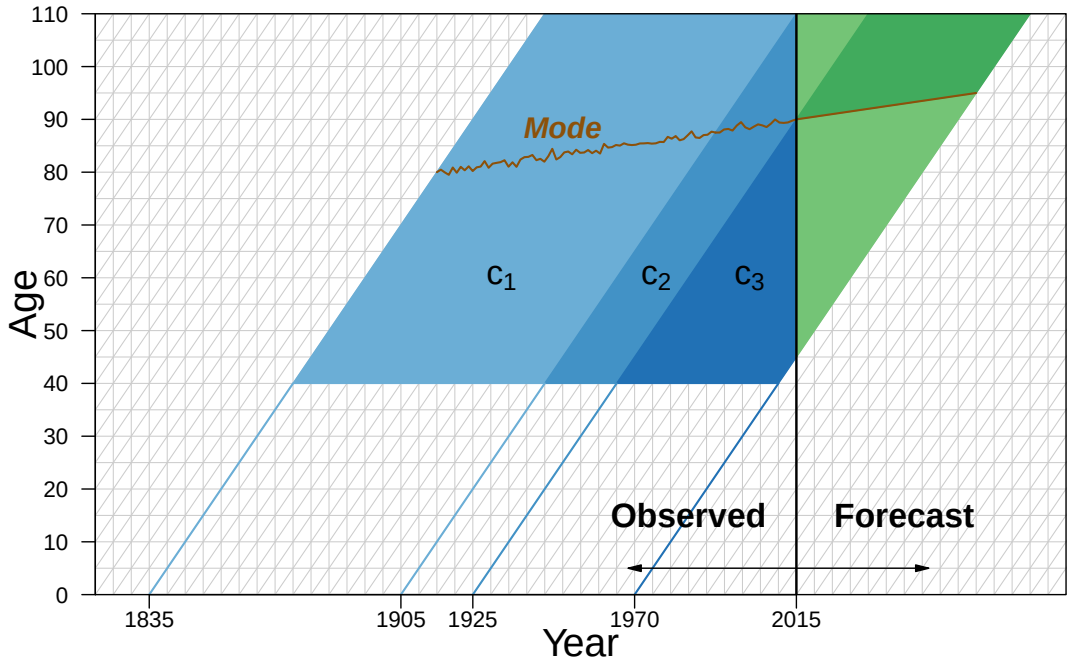

Figure 2. A Lexis diagram illustrating the division of cohorts into three groups. 2015 is the most recent year of data collection, while the last cohort with observed mode M is $\tilde{c} = 1925$ ($M = 2015 - \tilde{c} = 90$, brown line). The three groups are then $c_1 = 1835, \ldots, 1905$, $c_2 = 1906, \ldots, 1925$ and $c_3 = 1926, \ldots, 1970$. The two colours in the forecast years correspond to different parameters' derivations (cf. Section 2.3): estimation with missing data (light green) and forecasting (dark green).

2.3 Estimation and forecast of the C-STAD parameters

The first step needed to fit the C-STAD model is the derivation of the standard distribution $f(x)$. The C-STAD can be interpreted as a relational model (Brass, 1971), hence it is desirable to include the representative features of the observed data for all cohorts in the computation of $f(x)$. To do so, we first derive the age-at-death distribution for each cohort 1835-1970 by adapting the two-dimensional (2D) P-splines smoothing approach of Currie et al. (2004) to cohort mortality. 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). We thus smooth observed death counts over cohorts using a 2D tensor product of B-splines, age- and cohort-specific smoothing parameters, and exposures as an offset (Eilers and Marx, 1996). The smoothing parameters are chosen by BIC minimization, and missing data are estimated by P-spline extrapolation. The `R` package `MortalitySmooth` provides a direct implementation of this procedure (Camarda, 2012).

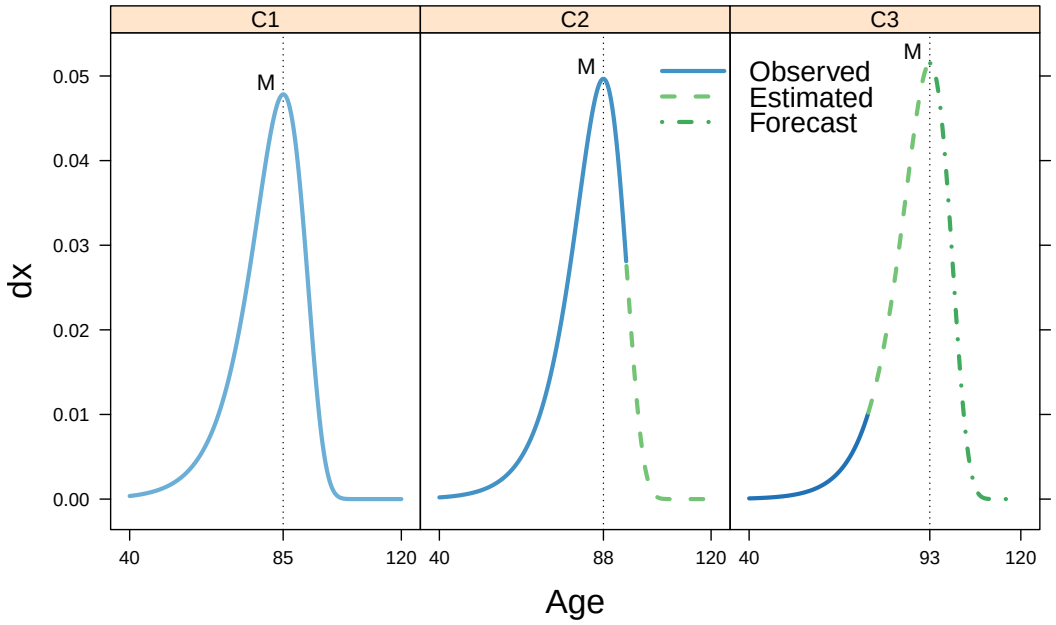

Figure 3. Example of observed, estimated and forecast data for three age-at-death distributions belonging to the groups of cohorts c_1, c_2 and c_3.

The estimated smooth mortality surface allows us to derive smooth densities and corresponding modal ages at death for each cohort analysed. We then align each density to the distribution of the first cohort, and we derive $f(x)$ as the mean of the aligned distributions. This landmark registration procedure has been suggested elsewhere as it enhances the goodness-of fit of the model (for additional details, see Basellini and Camarda, 2019). Importantly, it should be noted that for cohorts in c_2 and c_3, we only use the part of the aligned distribution corresponding to the observed data (for example, for $\tilde{c}$ we only use the part of the distribution corresponding to the ages below the mode). We can then express $f(x)$ 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 can derive the C-STAD parameters $\boldsymbol{\theta}' = [s, b_L, c_L, d_L, b_U]$ for each cohort in 1835-1970. As anticipated in Section 2.2, we use three different approaches to estimate $\boldsymbol{\theta}$, depending on the data available for each cohort; we thus divide cohorts into three groups, as shown in Figure 2.

For the first group of fully observed cohort c_1, we start by estimating the $\boldsymbol{s}$ parameter. Rather than using the 2D smooth estimates obtained from the alignment procedure, we employ a one-dimensional P-spline approach, i.e. we smooth mortality for each cohort independently (as typically done in period analyses, see, for example, Ouellette and Bourbeau, 2011). Similar to the 2D approach, this allows us to extract modal ages at death for each cohort, and hence to estimate the parameter $\hat{s}_c = M_c - M_f$, where M_f denotes the mode of the standard distribution. We favour this over the 2D approach because it allows us to capture cohort-specific mortality

fluctuations, which are typically smoothed over cohorts in the 2D estimates.

Having derived an estimate of the shifting parameter $\hat{\boldsymbol{s}}$, we can estimate the remaining parameters $\tilde{\boldsymbol{\theta}}' = [b_L, c_L, d_L, b_U]$. We take advantage of the Poisson assumption in Eq. (3) and maximise the following log-likelihood function:

$$\ln \mathcal{L}_1 \left(\tilde{\boldsymbol{\theta}}_c | \, D_{x,c}, E_{x,c}, \hat{s}_c, \boldsymbol{\beta}_f \right) \propto \sum_x \left[D_{x,c} \log \left(\mu_{x,c}^{\texttt{C-STAD}} \right) - E_{x,c} \, \mu_{x,c}^{\texttt{C-STAD}} \right] \tag{5}$$

for each cohort in $c = 1835, \ldots, 1905$, where $\mu_{x,c}^{\texttt{C-STAD}}$ denotes the estimated hazard of the C-STAD model. In words, the optimization procedure looks for a combination of parameters $\hat{\tilde{\boldsymbol{\theta}}}_c$ for each cohort that produces an age-at-death distribution $\hat{g}_c(x) = f\left[t\left(x; \left[\hat{s}_c \, , \hat{\tilde{\boldsymbol{\theta}}}_c\right]\right)\right]$ whose corresponding hazard maximises the log-likelihood in Eq. (5). The hazard is derived numerically from the distribution $\hat{g}_c(x)$.

For the second group of partially observed cohorts c_2, we start again from the shifting parameter $\boldsymbol{s}$. We use the same estimation approach used in c_1: data are available until the ages above the mode, therefore the 1D smoothing approach produces an estimate of M and $\hat{s}_c$ for each cohort in c_2. With respect to $\tilde{\boldsymbol{\theta}}$, we modify Eq. (5) to take into account the missing data above the mode. Let $W_{x,c}$ be a matrix of weights at given age x and cohort c, whose elements are equal to one if the corresponding death counts $D_{x,c}$ and exposures $E_{x,c}$ have been observed, and zero otherwise. Then, we estimate $\tilde{\boldsymbol{\theta}}$ by maximizing the log-likelihood:

$$\ln \mathcal{L}_2 \left(\tilde{\boldsymbol{\theta}}_c | \, D_{x,c}, E_{x,c}, W_{x,c}, \hat{s}_c, \boldsymbol{\beta}_f \right) \propto \sum_x W_{x,c} \, \left[D_{x,c} \log \left(\mu_{x,c}^{\texttt{C-STAD}} \right) - E_{x,c} \, \mu_{x,c}^{\texttt{C-STAD}} \right] \tag{6}$$

for each cohort in $c = 1906, \ldots, \tilde{c}$. It should be noted here that the missing data only influence the estimation of b_U, as complete data are observed below the mode for all cohorts in this group.

Finally, for the third group of partially observed cohorts in c_3, we employ a mixture of forecasting and estimation for the C-STAD parameters. The lack of data above the modal age at death makes it impossible to estimate the parameter $\boldsymbol{s}$ and compute the log-likelihoods in Eq. (5) and (6). As such, we start from the time-series of the estimated parameters $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{b}}_U$ in c_1 and c_2. From a theoretical perspective, these two parameters are related by the fact that only mortality changes occurring above the mode can modify its value (cf. Appendix B in Canudas-Romo, 2010). Correlation analyses for the two countries analysed confirm the strong relation between the two series (Pearson correlation of 0.96 and 0.90 for the time-series in first differences for Sweden and Denmark, respectively). As such, we specify a vector autoregressive (VAR) model of order one with constant for the two (differenced) parameters, and we forecast their values for all cohorts in c_3. The `R` package `vars` allows us to perform model selection and estimation, while parameters' forecast are computed via the `forecast` package (Pfaff, 2008a,b; Hyndman et al., 2018)

Then, we take the forecast values of $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{b}}_U$ as given, and we estimate the remaining parameters $\breve{\boldsymbol{\theta}}' = [b_L, c_L, d_L]$ by maximizing the log-likelihood:

$$\begin{aligned} \ln \mathcal{L}_3 \left(\breve{\boldsymbol{\theta}}_c | \, D_{x,c}, E_{x,c}, W_{x,c}, \hat{s}_c, \hat{b}_{U_c}, \boldsymbol{\beta}_f \right) \propto \\ \sum_x W_{x,c} \, \left[D_{x,c} \log \left(\mu_{x,c}^{\texttt{C-STAD}} \right) - E_{x,c} \, \mu_{x,c}^{\texttt{C-STAD}} \right] \end{aligned} \tag{7}$$

for each cohort in $c = \tilde{c} + 1, \ldots, 1970$. In contrast to the estimation procedure in c_2, here the missing data influence the estimation of the *Lower* parameters $\breve{\boldsymbol{\theta}}$.

The estimate $\hat{\boldsymbol{\theta}}_c$ for each cohort in 1835-1970 allows us to derive a complete set of age-specific mortality measures, i.e. we can complete the mortality experience for the partially observed cohorts of our analysis. In order to derive the C-STAD confidence intervals (CI)[1], we employ a bootstrapping procedure (Efron and Tibshirani, 1994). As suggested by Keilman and Pham (2006), we consider the uncertainty of the C-STAD estimates as well as of the forecast values of the time-series of $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{b}}_U$. On one hand, the first source of uncertainty is accounted for by generating bootstrap death counts from the C-STAD deviance residuals (as in, for example, Koissi et al., 2006; Renshaw and Haberman, 2008; Ouellette et al., 2012). Appendix A provides more details on the computation of deviance residuals and bootstrap death counts. On the other hand, the second source is considered by simulating future values of the VAR model. We employ 25 different matrices of bootstrap death counts, and for each of these, we refit the C-STAD model and simulate 25 future values of $\boldsymbol{s}$ and $\boldsymbol{b}_U$. From the 625 resulting total simulations, we take the lower and higher deciles to construct 80% confidence intervals [XXX increase simulations for final paper].

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 Danish 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 $c_1 = 1835, \ldots, 1905 - \delta$ and we forecast mortality δ years ahead. We then compare the forecast life expectancy at age 40 (e_{40}) and the Gini coefficient at age 40 (g_{40}) with the observed values. Both measures of longevity (the former) and lifespan variability (the latter) are indeed useful to evaluate the accuracy of mortality forecasts (Bohk-Ewald et al., 2017).

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 fit the C-STAD to the fully observed cohorts $c_1 = 1835, \ldots, 1895$, and we forecast 10-year ahead. By doing so, we complete the mortality experience of the partially observed cohorts $1896, \ldots, 1905$, 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

[1] to avoid confusion, we use the general term CI for all cohorts analysed, even when intervals are constructed from the mixture of forecast and estimated parameters (i.e. cohorts c_3)

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 combination of forecasting and estimation of the C-STAD parameters.

In addition to the C-STAD, we perform the same out-of-sample exercises with the 2D P-splines approach of Currie et al. (2004). This is indeed the only model that, up to our knowledge, has been employed to forecast cohort mortality from a cohort perspective (Continuous Mortality Investigation, 2007) and is readily implemented in the `R` software.

Table 1 presents the results of our analysis. The first and second columns contain the cohorts used for fitting and forecasting the C-STAD and 2D P-splines models, respectively. The third column contains the time horizon of the out-of-sample exercise, while the fourth column the measure analysed (e_{40} and g_{40}). Results are shown in the last four columns. We assess the accuracy of the point forecasts by computing the root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{c=1}^{\delta} (\hat{y}_c - y_c)^2}{\delta}} \tag{8}$$

where δ is the forecasting horizon, and $\hat{y}_c$ and y_c 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.

				Sweden		**Denmark**	
Fitting cohorts	Forecast cohorts	Horizon	Measure	C-STAD	2D P-spline	C-STAD	2D P-spline
1835-1895	1896-1905	10y	e_{40}	**0.05**	0.08	**0.05**	0.08
			g_{40}	**0.02**	0.11	**0.02**	0.09
1835-1890	1891-1905	15y	e_{40}	**0.04**	0.09	**0.05**	0.08
			g_{40}	**0.02**	0.10	**0.03**	0.12
1835-1885	1886-1905	20y	e_{40}	**0.03**	0.09	**0.04**	0.08
			g_{40}	**0.04**	0.11	**0.04**	0.13
1835-1880	1881-1905	25y	e_{40}	**0.02**	0.08	**0.04**	0.08
			g_{40}	**0.05**	0.11	**0.08**	0.13
1835-1875	1876-1905	30y	e_{40}	**0.03**	0.10	**0.05**	0.11
			g_{40}	**0.06**	0.14	**0.10**	0.19
1835-1870	1871-1905	35y	e_{40}	0.10	**0.09**	**0.08**	0.15
			g_{40}	**0.10**	0.15	**0.17**	0.21

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

The table shows that the C-STAD forecasts are accurate in completing the mortality experience of partially observed cohorts. The RMSE values of the two indicators are low, and they do not increase significantly with the forecasting horizon. This is probably due to the fact that the completion of the mortality experience takes into account the partially observed data, even for the long term forecasts. Very similar results would be obtained if we employed a different prediction accuracy measure (e.g. MAPE, MAE). [XXXX revise this after analysis]

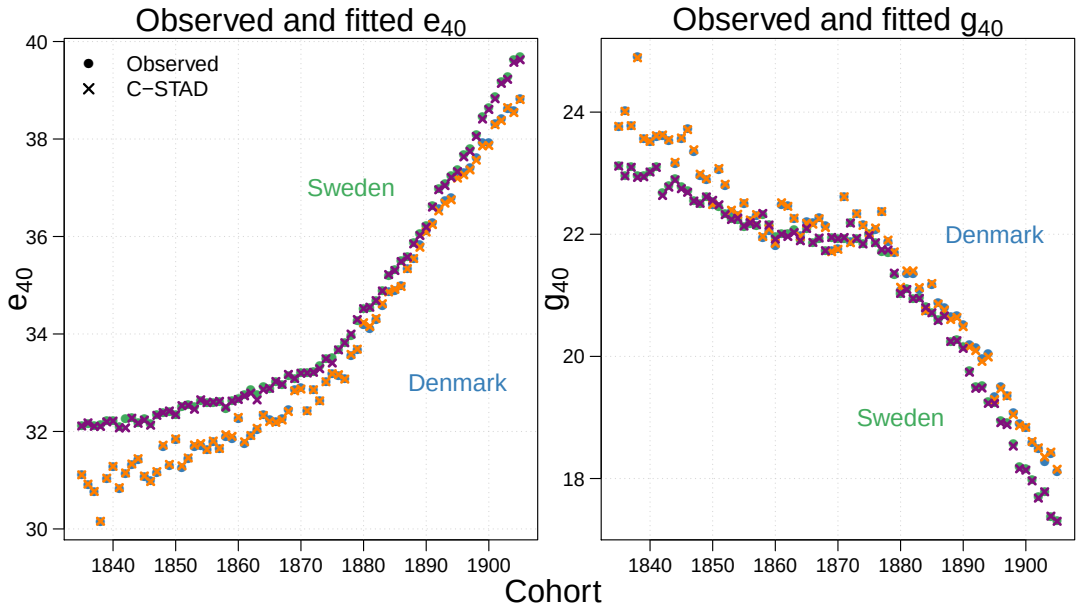

Figure 4. Observed and C-STAD estimated remaining life expectancies at age 40 (e_{40}, left panel) and Gini coefficient at age 40 (G_{40}, right panel) for adult females in Sweden and Switzerland for the fully observed cohorts 1835-1905.

XXXXXX ?? INTERVAL FORECAST ACCURACY ?? XXXXXX

3.2 Mortality developments for Swedish and Danish females, cohorts 1835-1970

Here, we show the results of employing the C-STAD model to estimate and forecast adult female cohort mortality in Sweden and Denmark for the cohorts 1835-1970. The estimated and forecast parameters are shown and discussed in Appendix B. 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 c_1 (1835-1905). The two graphs provide evidence on the goodness-of-fit of the C-STAD model, whose estimates are very close to the observed values for both measures in the two populations. Inspection of the deviance residuals (shown in Appendix B) provides additional evidence on the adequacy of the C-STAD model.

Figure 5 shows the observed (cohorts c_1) and completed (c_2 and c_3) e_{40} and g_{40} computed with the C-STAD (with 80% confidence intervals) and 2D P-spline model for the two population analysed. Despite sharing similar country trends in the fully observed cohorts c_1, it is interesting to observe the different mortality developments in the partially observed cohorts c_2: while Swedish adult females show continuous improvements in longevity and lifespan equality, Danish ones display a stagnation of e_{40} and an increase in lifespan inequality. The trends of the mortality measures for the partially observed cohorts are similar across the models, with the exception of Danish e_{40}: the 2D P-spline forecast is much faster then the C-STAD one, resulting in a crossover among the two populations. With respect to C-STAD forecasts, it is moreover interesting to observe how confidence intervals are rather narrow for both

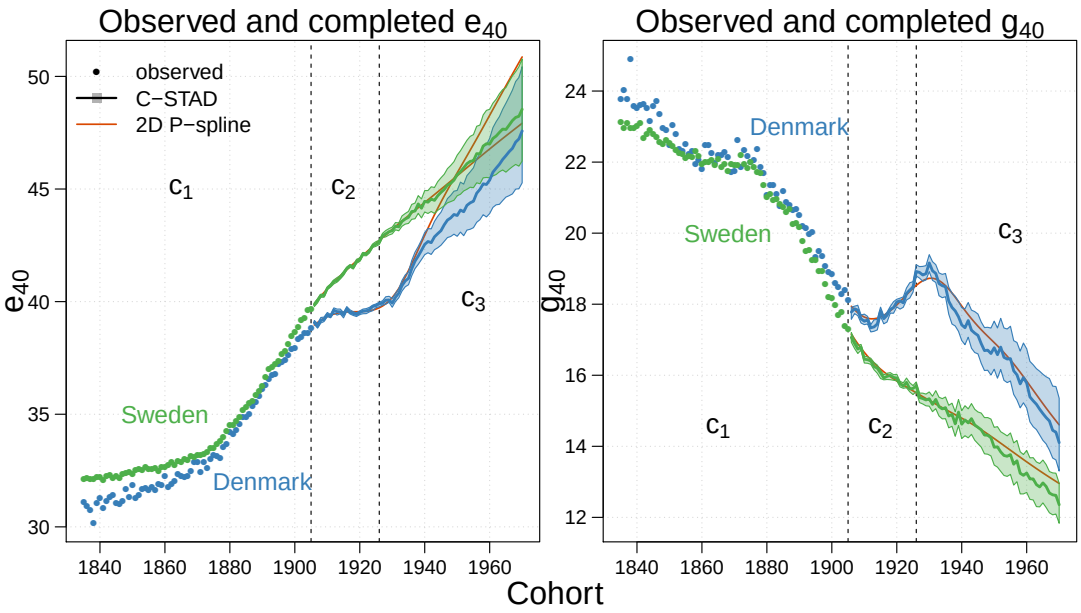

Figure 5. Observed (cohorts c_1) and completed (c_2 and c_3) remaining life expectancies at age 40 (e_{40}, left panel) and Gini coefficient at age 40 (g_{40}, right panel) for the C-STAD (with 80% confidence intervals) and 2D P-spline models for adult females in Sweden and Denmark for the cohorts 1835-1970.

countries in c_2 (as the great majority of data is observed for these cohorts), while they increase in the cohorts c_3 proportionally to the amount of missing data.

The age-specific mortality rates analysis shown in Figure 6 offers additional insights on cohort mortality developments of the two populations. In the top panels, observed, fitted and forecast mortality rates over cohorts are shown for some selected ages. In addition to the goodness-of-fit of the C-STAD model, the graphs highlight diverse age-specific developments in the two countries: for example, mortality at ages 40 and 60 of Danish cohorts born at the beginning of the twentieth century did not improve, resulting in the atypical trends of the summary measures shown in Figure 5 (stagnation of e_{40} and increase in lifespan inequality). In the bottom panels, mortality rates over all ages are shown for some selected cohorts. This second perspective shows how the shape of the mortality curve, appropriately captured by the C-STAD model, changed over time: for example, for the 1835 cohort, maternal mortality was still high in both countries, with age-specific rates showing small increases in the age range 40-50. The subsequent reduction of maternal mortality over cohorts clearly emerges in the top panels (at age 40) and in the other cohorts of the bottom panels. An interesting observation that emerges from Figure 6 is that the confidence intervals of the C-STAD widens as expected: for example, variability increases with age for the completed cohorts, as fewer age-specific data have been observed at higher ages. Finally, Figure 7 shows the observed and C-STAD age-at-death distributions for the three cohorts analysed in the previous panels.

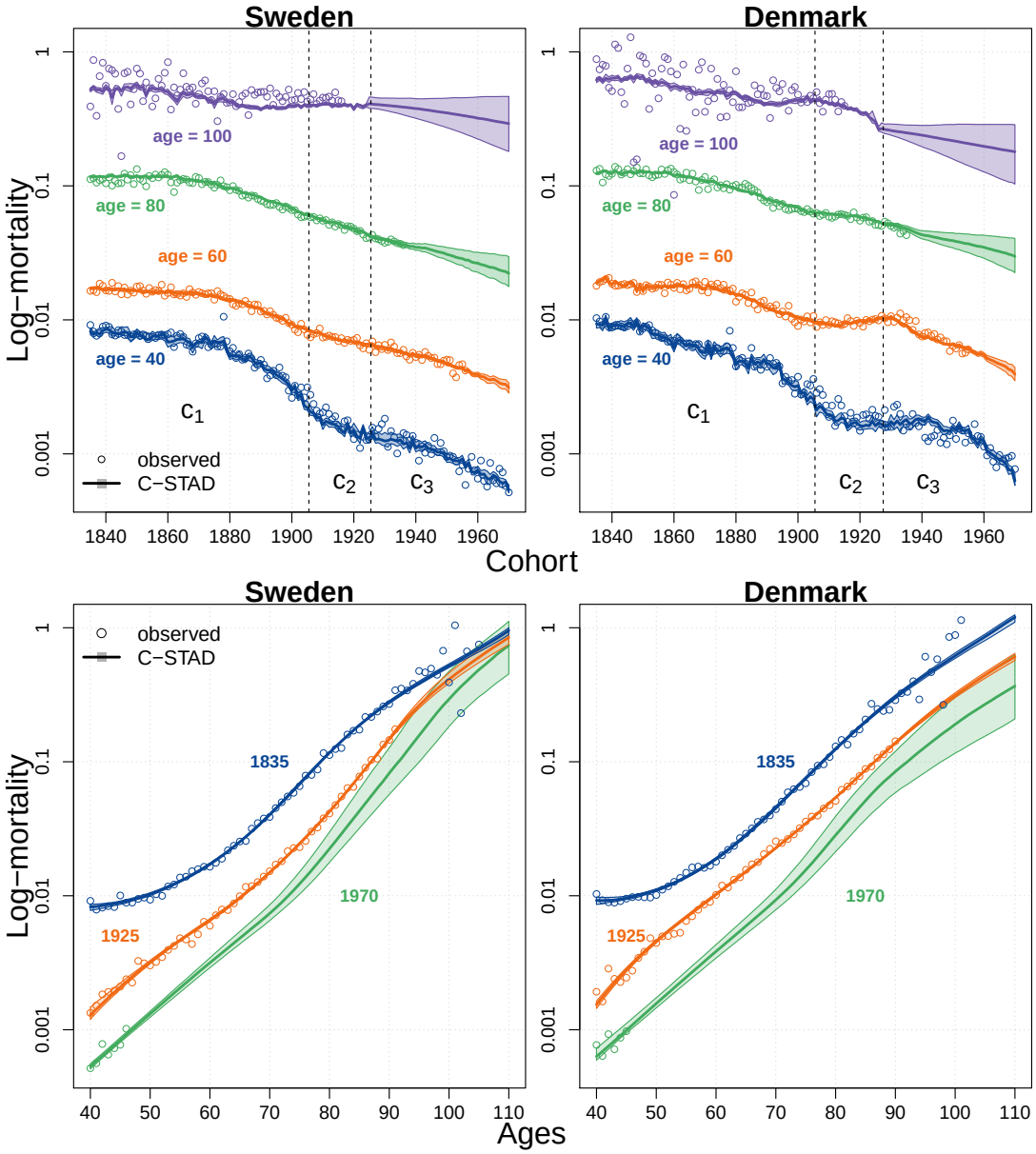

Figure 6. Observed, fitted and forecast age-specific mortality rates for selected ages (top panels) and for selected cohorts (bottom panels) with 80% confidence intervals for females in Sweden and Denmark aged 40-110+ for the cohorts 1835-1970.

4 Discussion [XXX to work on in June]

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-

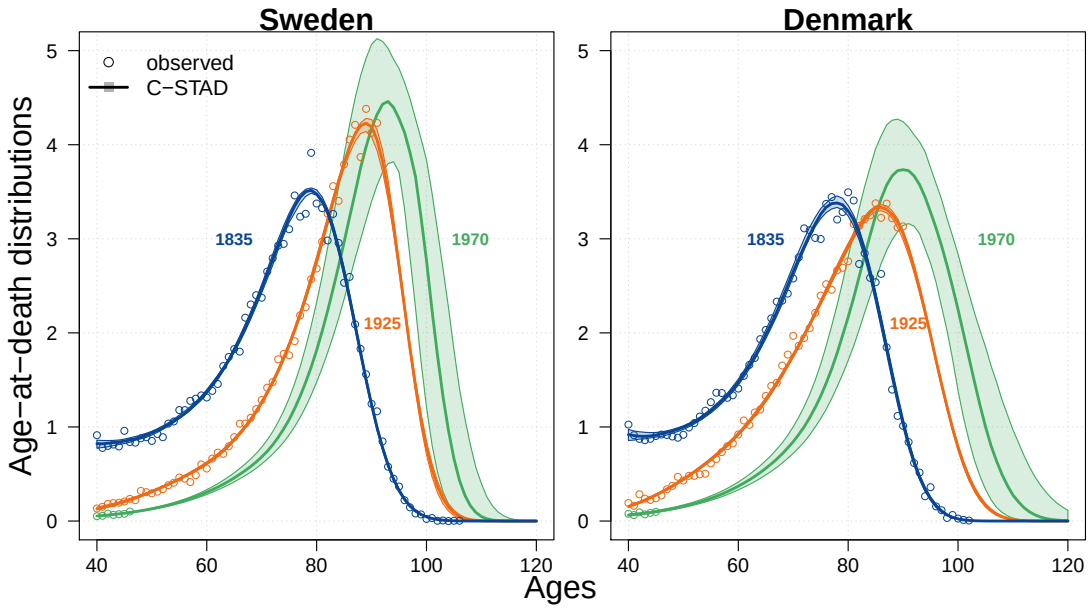

Figure 7. Observed, fitted and forecast age-at-death distributions for selected cohorts (bottom panels) with 80% confidence intervals for females in Sweden and Denmark aged 40-110+.

death distributions: specifically, we propose a cubic segmented transformation of the age-axis of a standard distribution to describe and forecast mortality developments. By looking at the cohort perspective, we denote our methodology *Cohort 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 Danish adult females for the cohorts 1835-1970. 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 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 (2019) 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. 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.

- remember to add Rune's references on the increase of lifespan variability using APC models
- discuss the advantage of reconstructing cohort mortality forecast, ie can reconstruct the trends in f5
- cite JASA paper (x2)
- cite Lucia EM paper (thesis)

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. (2019). Modelling and forecasting adult age-at-death distributions. *Population Studies*, 73(1):119–138.

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.

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan Disparity as an Additional Indicator for Evaluating Mortality Forecasts. *Demography*, 54(4):1559–1577.

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*.

Brass, W. (1971). On the Scale of Mortality. In Brass, W., editor, *Biological Aspect of Demography*. Taylor & Francis, London.

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.

Continuous Mortality Investigation (2007). Stochastic projection methodologies: Further progress and P-Spline model features, example results and implications. Revised Working Paper 20 (November 2007).

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

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

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

Human Mortality Database (2019). 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 8 May 2019).

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., and Yasmeen, F. (2018). *forecast: Forecasting functions for time series and linear models*. R package version 8.4.

Keilman, N. and Pham, D. (2006). Prediction intervals for LeeCarter-based mortalityforecasts. In *European Population Conference 2006, Liverpool, June 21–24, 2006*.

Koissi, M.-C., Shapiro, A. F., and Hgns, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1):1–20.

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

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. London: Chapman and Hall.

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.

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

Ouellette, N., Bourbeau, R., and Camarda, C. G. (2012). Regional disparities in Canadian adult and old-age mortality: A comparative study based on smoothed mortality ratio surfaces and age at death distributions. *Canadian Studies in Population*, 39(3-4):79–106.

Pascariu, M. D., Lenart, A., and Canudas-Romo, V. (2019). The maximum entropy mortality model: forecasting mortality using statistical moments. *Scandinavian Actuarial Journal*, pages 1–25.

Pfaff, B. (2008a). *Analysis of integrated and cointegrated time series with R*. 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.

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.

Renshaw, A. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance: Mathematics and Economics*, 42(2):797–816.

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

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

Vaupel, J. and Lundström, H. (1994). Longer life expectancy? Evidence from Sweden of reductions in mortality rates at advanced ages. In *Studies in the Economics of Aging*, pages 79–102. University of Chicago Press.

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

A Deviance residuals and bootstrap death counts

Model residuals are routinely analysed to explore the goodness-of-fit of a model as well as the adequacy of assumptions about error terms. Within a GLM setting (such as the Poisson considered here), the deviance is often used to measure discrepancy between fitted and actual data. Deviance residuals for the Poisson distribution are given by:

$$r_{\mathrm{D}} = \mathrm{sign}(D_{x,c} - \hat{D}_{x,c})\,\sqrt{2}\,\left[D_{x,c}\ln\left(\frac{D_{x,c}}{\hat{D}_{x,c}}\right) - \left(D_{x,c} - \hat{D}_{x,c}\right)\right]^{1/2} \tag{9}$$

where $D_{x,c}$ and $\hat{D}_{x,c}$ denote the observed and fitted death counts at age x and for cohort c, respectively (McCullagh and Nelder, 1989).

Deviance residuals can be further employed to take into account the uncertainty related to the estimation of a model parameters as suggested by Koissi et al. (2006). Specifically, bootstrap death counts can be computed by resampling deviance residuals with replacement and mapping them to corresponding death counts. We refer the interested reader to Renshaw and Haberman (2008) for details on the inverse formulas, which are based on the seminal work of Efron and Tibshirani (1994).

B Section 3.2: additional results

Figure B.1 shows the fitted and forecast C-STAD parameters with 80% confidence intervals for Swedish and Danish adult females for cohorts 1835-1970.

We performed diagnostic checks on the fitted C-STAD model for the two populations analysed in this paper by using Equation (9). Poisson deviance residuals for the

two populations are shown in Figure B.2. No clear patterns (except for the Spanish flu) emerge from this graphical analysis.

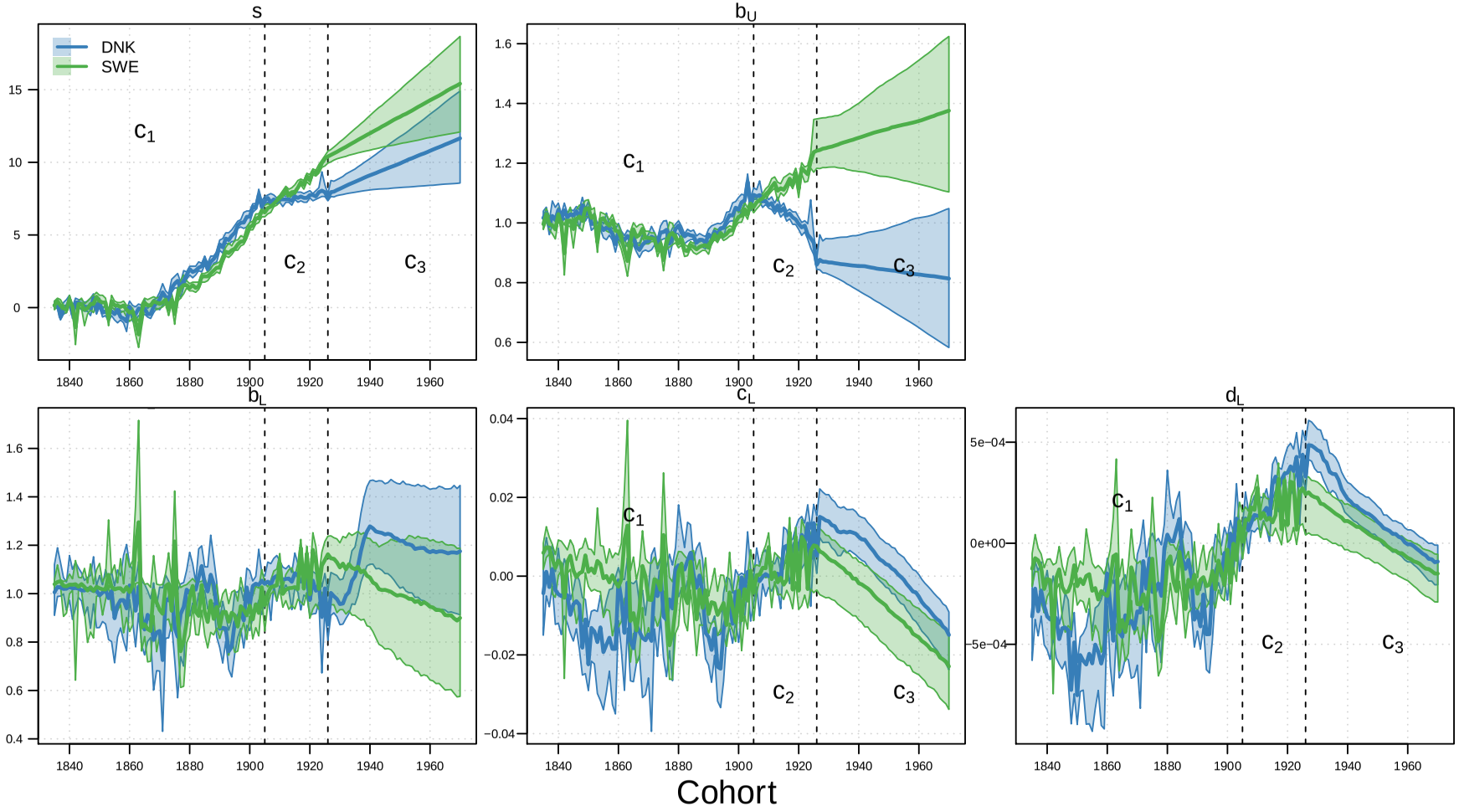

Figure B.1. Estimated and forecast C-STAD parameters for adult females in Sweden and Denmark for the cohorts 1835-1970.

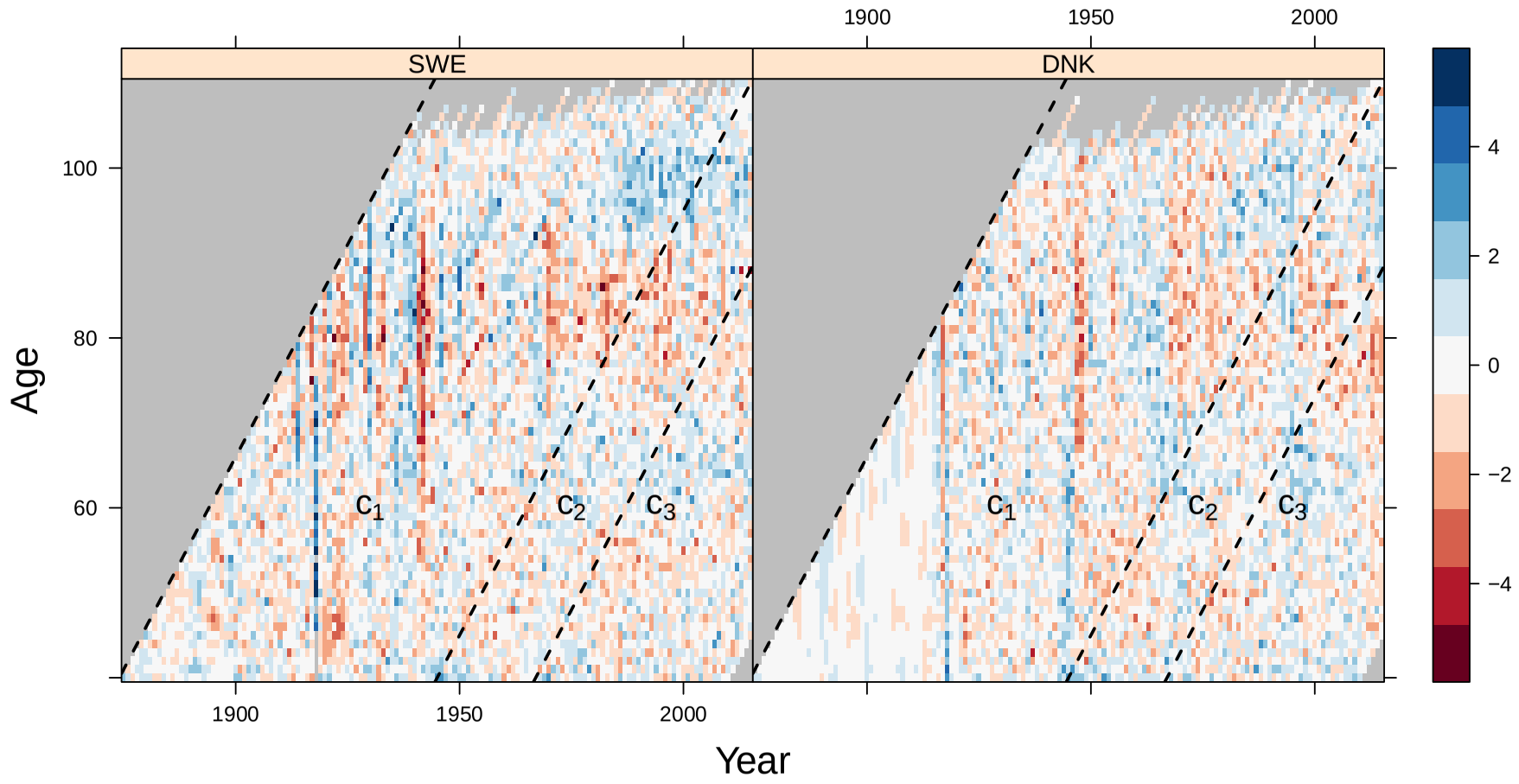

Figure B.2. Poisson deviance residuals of the C-STAD model for adult females in Sweden and Denmark for the cohorts 1835-1970.