Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model

Silvia Rizzi, Søren Kjærgaard, Marie-Pier Bergeron-Boucher, Carlo Giovanni Camarda, Rune Lindahl-Jacobsen, James W. Vaupel

ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model

Silvia Rizzi [a,b,1,*], Søren Kjærgaard [a,1], Marie-Pier Bergeron Boucher [a], Carlo Giovanni Camarda [c], Rune Lindahl-Jacobsen [a,b], James W. Vaupel [a]

[a] Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Denmark
[b] Department of Epidemiology Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark, Odense, Denmark
[c] Institut national d' études démographiques (INED), Paris, France

ARTICLE INFO

Keywords:
Cohort life tables
Cohort mortality forecasts
Demographic forecasting
Non-extinct cohorts
Smoothing

ABSTRACT

Mortality forecasting has crucial implications for insurance and pension policies. A large amount of literature has proposed models to forecast mortality using cross-sectional (period) data instead of longitudinal (cohort) data. As a consequence, decisions are generally based on period life tables and summary measures such as period life expectancy, which reflect hypothetical mortality rather than the mortality actually experienced by a cohort. This study introduces a novel method to forecast cohort mortality and the cohort life expectancy of non-extinct cohorts. The intent is to complete the mortality profile of cohorts born up to 1960. The proposed method is based on the penalized composite link model for ungrouping data. The performance of the method is investigated using cohort mortality data retrieved from the Human Mortality Database for England & Wales, Sweden, and Switzerland for male and female populations.

1. Introduction

Mortality forecasts are, in almost all cases, based on period mortality information. These forecasts are calculated using period life tables, constructed under the following assumption: *what if* a hypothetical cohort was subject to the mortality conditions of that specific time period, normally one year (Preston, Heuveline, & Guillot, 2001). This shows the probability of death at a given point in time for people of different ages. A clear limitation of this approach is that people are not living in one time period but instead are members of a birth cohort that is aging one year every year. The period perspective does not reflect any real population and potentially ignores important connections within a birth cohort. For example, it has been argued that conditions in a person's childhood matter for the mortality experienced at more advanced ages (Barker, 2004; Elo & Preston, 1992), and it has been clearly demonstrated that smoking behaviors have an impact on death rates later in life (Doll & Hill, 1950; Doyle, Dawber, Kannel, Heslin, & Kahn, 1962; Janssen & Kunst, 2005; Preston & Wang, 2006). Long-term effects are potentially shared across a specific birth cohort; therefore, cohort life tables are often considered to be more informative than period life tables.

Cohort forecasting has been less common due to the heavy data demand it requires. To have mortality data

* Corresponding author at: Department of Epidemiology Biostatistics and Biodemography, Institute of Public Health, University of Southern Denmark, Odense, Denmark.
E-mail addresses: srizzi@health.sdu.dk (S. Rizzi), skjaergaard@sdu.dk (S. Kjærgaard), mpbergeron@sdu.dk (M.-P. Bergeron Boucher), carlo-giovanni.camarda@ined.fr (C.G. Camarda), rjacobsen@health.sdu.dk (R. Lindahl-Jacobsen), jvaupel@health.sdu.dk (J.W. Vaupel).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ijforecast.2020.03.003

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

on one complete cohort, one must wait for over a century, and the "youngest" cohort with completed mortality might provide outdated information. Because cohort life tables represent the real mortality schedule for a group of individuals, they provide relevant information for life tables users, including life insurance and pension companies. A pension company is better off knowing the actual future life expectancy of its customers than the period life expectancies. Likewise, public institutions that plan health care, public pensions, and so forth, are essentially interested in the real future life time of people living in a society. In order for a cohort life table to be useful to these life table users, forecasting methods that can complete recent cohorts that are not yet extinct are required.

Mortality forecasts are typically obtained by extrapolating period death rates, which are used to calculate life expectancy forecasts. The most widely used forecasting method in countries with reliable vital statistics data is the model by Lee and Carter (1992). It summarizes death rates on a logarithmic scale using principal component techniques and linearly extrapolates the associated time index. The Lee–Carter model has been the inspiration for various modifications and extensions, including Booth, Maindonald, and Smith (2002), Hyndman, Booth, and Yasmeen (2013), Lee and Miller (2001) and Renshaw and Haberman (2006). The model is used by statistical offices to produce official mortality forecasts (Stoeldraijer, van Duin, van Wissen, & Janssen, 2013). Instead of using period death rates, Cairns, Blake, and Dowd (2006) suggested using another period life table measure, the probability of death, which can be linearly extrapolated after a logit transformation. More recently, the life table death distribution has been suggested as another useful measure when forecasting period mortality (Basellini & Camarda, 2019; Bergeron-Boucher, Canudas-Romo, Oeppen, & Vaupel, 2017; Bergeron-Boucher, Simonacci, Oeppen, & Gallo, 2018; Kjærgaard, Ergemen, Kallestrup-Lamb, Oeppen, & Lindahl-Jacobsen, 2019; Pascariu, Lenart, & Canudas-Romo, 2019). Common to all these mortality models is that period life table information is used.

Forecasts accounting for a cohort effect have been estimated based on age-period-cohort (APC) models. For example, Renshaw and Haberman (2006) provided an extension of the Lee–Carter model and Cairns et al. (2009) an extension of the Cairns et al. (2006) model to produce APC forecasts of mortality. These models, which have been shown to improve the fit to mortality data, provide period forecasts. They are also subject to limitations such as identifiability problems, which are addressed with parameter constraints, and independence is assumed between the period and the cohort effects even though they can be correlated (Currie, 2012). Cohort (age-cohort) forecasting has rarely been implemented. Such forecasts can be performed by applying parametrization functions (e.g., Heligman-Pollard or Siler models) to incomplete cohort data to estimate mortality after the age of truncation (Booth & Tickle, 2008). However, applying these models to incomplete data can lead to estimation problems. In an extensive literature review of forecasting models, Booth and Tickle (2008) found that only the Continuous Mortality Investigation Bureau (2006) achieved a true cohort forecast, using the P-spline regression method from Currie, Durban, and Eilers (2004) to smooth age-cohort mortality. Later, Chiou and Müller (2009) explored the use of functional data analysis to model cohort lifetables and derive mortality forecasts, and Basellini, Kjærgaard, and Camarda (2019) used an age-at-death distribution model which segments the death distribution. To the knowledge of the authors, these latter three are the only applications of cohort data used for forecasting.

In this article, we suggest a novel method to forecast mortality using cohort information from a cohort life table. We aim to complete the mortality trajectories for non-extinct cohorts with a penalized composite link model (PCLM) (Eilers, 2007). Rizzi, Gampe, and Eilers (2015) previously used the model to estimate detailed age-at-death distributions from coarsely grouped death counts, and the model is further extended in the analysis here. The PCLM has proven useful for ungrouping data aggregated in an open-age interval (e.g., 85+). The number of survivors from a given birth cohort at the age of truncation can be considered as a coarsely grouped death count not yet observed. The remaining deaths can be ungrouped (forecast) by age using the PCLM, and the cohort mortality can be completed. The forecast is based on the death distribution of a cohort life table. This makes it possible to exploit several observable features: the death distribution sums to the radix of the life table, a well-defined local maximum of the distribution can be found at high ages, and the number of deaths after the age of 120 is very close to zero or equal to zero. We exploit all three features in our model to allocate remaining deaths for cohorts that are still alive.

2. Material and methods

2.1. Cohort death distributions

The cohort death distribution is the basis of the proposed method so as to forecast cohort mortality. Standard life table techniques (Preston et al., 2001) are used to calculate death distributions for all complete and non-complete cohorts assuming an initial population size, conventionally set at $N = 100.000$, and known as the radix of the life table. The death distribution is denoted by the matrix $d_{x,c}$, where x stands for ages $x \in (0, 1, \ldots, 110)$ and c for the specific birth cohort.

We divide $d_{x,c}$ into three time windows that are treated differently. First, a time window denoted by T1 consists of all complete cohort death distributions. For observed data, we define complete cohorts by a maximum life length of 110 years following standard conventions similar to, for example, Wilmoth et al. (2017). A small fraction of the population lives past the age of 110, but closing the life table at 110 years is of minor importance for the studied populations and does not affect the results. The T1 time frame ends with the cohort born 1905. The second time frame is denoted by T2 and consists of incomplete cohorts until the birth cohort 1935.[2] Here, the complete death

[2] For Swiss males, T2 ends at the 1933 cohort because the model was unable to find a stable trend for the mode for the 1934 and 1935 cohorts.

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

distribution of the cohorts not yet extinct is estimated with the PCLM for ungrouping (Rizzi et al., 2015). The birth cohort 1935 is the final cohort where the PCLM can determine a correct local maximum of the death distribution at adult ages for the studied populations. The local maximum at adult ages is referred to as the number of deaths at the mode, and the corresponding age as the modal age at death (Canudas-Romo, 2010). Any cohort in T2 consists of a part, $d_{0:A,c}$, with observed deaths from age zero to the last observed age (A) and a part, $d_{A+,c}$, which is unobserved and grouped for all ages above A. Because the number of deaths over all ages, by construction, sum to the radix of the life table, it is possible to determine the unobserved part using the equation $d_{A+,c} = N - \sum_{x=0}^{A} d_{x,c}$. The unobserved deaths in that sense remain because the people are still alive but will eventually die. Hence, a forecast of the mortality for each birth cohort can be found by allocating the remaining deaths in a way similar to an ungrouping problem for right censored data, which can be performed accurately for high ages (Rizzi et al., 2015). The third time window, T3, consists of all cohorts born after 1935. In this setting, the complete death distribution of the incomplete young cohorts is estimated with an extended version of the PCLM for ungrouping by augmenting the input data: estimates of the local maximum of the death distribution, the corresponding modal age at death, and the proportion of deaths after the mode are all incorporated into the PCLM. Note that the separation of T2 and T3 is based on the selected populations and might differ for other populations than those analyzed in this study. Fig. 1 illustrates the three time windows and the corresponding analysis strategy.

2.2. PCLM used on T2 cohorts

The aim of the PCLM is to estimate a complete cohort death distribution from an incomplete life table; i.e., from a cohort that is not yet extinct. Rizzi et al. (2015) shows that a PCLM can be formulated so that it relates age-specific death counts from a latent but expected sequence of death counts, $\boldsymbol{\gamma} = (\gamma_j)$, with $j = 0, \ldots, J+1$ to an observed sequence of death counts $\boldsymbol{d} = (d_i)$ with $i = 0, \ldots, I+1$, where $I < J$ gives the number of age intervals; i.e., the length of all the observed single ages plus an open interval, which includes the remaining deaths of a cohort not yet extinct. The sum of the latent sequence $\boldsymbol{\gamma}$ equals the sum of the observed death counts $\boldsymbol{d}$. The observed death counts for each cohort are assumed to be realizations of a random variable X_i that follows a Poisson distribution with the expected value $\mu_i = E(X_i)$ so that μ_i measures the expected number of counts from a realization of X_i. Hence, the age-specific death count is observed with the probability $P(X_i = d_i) = \mu_i^{d_i} e^{-\mu_i} / d_i!$.

Latent death counts in a population of size N equal the age-specific probability of death p_j times the sample size $\gamma_j = N p_j$. In other words, the vector $\boldsymbol{\gamma}$ represents the complete cohort death distribution of expected means so that we aim to estimate from the observed incomplete cohort death distribution $\boldsymbol{d}$. We assume that $\boldsymbol{\gamma}$ and $\boldsymbol{\mu}$ are linearly related by a composition matrix $\boldsymbol{C}$ with a dimension $(I+1) \times (J+1)$ such that $\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{\gamma}$. The composition matrix $\boldsymbol{C}$ is a 0/1 matrix that describes how the latent sequence $\boldsymbol{\gamma}$ was summed before generating the data.

To estimate the complete death distribution of the cohorts not yet extinct in T2, the PCLM for ungrouping redistributes the sum of the remaining deaths on a fine grid of single ages until age 130, under the assumption that $\boldsymbol{\gamma}$ is smooth. An interval of zero death counts at the ages 120–130 is added to provide sufficient flexibility at high ages. Extending with a zero death count interval ensures that the estimated deaths after age 120 in the death distribution forecasts are very close to zero. Results from the PCLM are aggregated for ages 110 to 130. Estimates up to age 110 with the residual category 110+ are used coherently with the input data. Hence, the vector of observed deaths will be equal to $(d_0, \ldots, d_A, d_{A+}, 0)$. The corresponding composition matrix $\boldsymbol{C}$ is equal to the identity matrix for the first A ages in which we observe that deaths since γ_j and μ_i are related one to one. For the unobserved and zero parts, $\boldsymbol{C}$ equals sequences of ones as $\boldsymbol{d}_{A+}$ should be distributed to all ages from $A+1$ to 130, as follows:

$$\boldsymbol{C} = \begin{array}{c} \\ d_0 \\ d_1 \\ d_2 \\ \vdots \\ d_A \\ \\ d_{A+} \\ 0 \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_A & \gamma_{A+1} & \cdots & \gamma_{120} & \gamma_{121} & \cdots & \gamma_{130} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} \end{array}.$$

For further details on the PCLM for ungrouping, see Rizzi et al. (2015).

2.3. PCLM used on T3 cohorts

For cohorts younger than the birth cohort of 1935 the PCLM is not able to determine a reliable forecast of the death distribution without further assumptions: in particular, the procedure fails to estimate the modal age at death. The PCLM relies on the modest assumptions that death counts are Poisson distributed and that the resulting complete death distribution is smooth, which makes it less suitable for redistributing deaths over a high number of ages without further restrictions. Hence, we augment the matrix of observed deaths $d_{x,c}$ with forecasts of deaths at the mode $d_{M,c}$ of the modal age at death M and of the number of deaths after the mode $d_{M+,c}$. Information on the structure of the death distribution is thereby given to the PCLM, making it possible to efficiently allocate deaths over a long age range. Augmenting the data by $d_{M,c}$ and M leaves two intervals of deaths to be distributed: one with deaths before M ($d_{M-,c}$) and one after ($d_{M+,c}$). Estimates of these are found by forecasting the proportion of deaths after the mode and, from that, by calculating $d_{M^+,c}$ and $d_{M^-,c}$, again using the property that all deaths sum to the radix. The corresponding compositional matrix is given in Box I:

For the populations considered in this study, $d_{M,c}$ and M have shown a highly linearly increasing pattern over

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

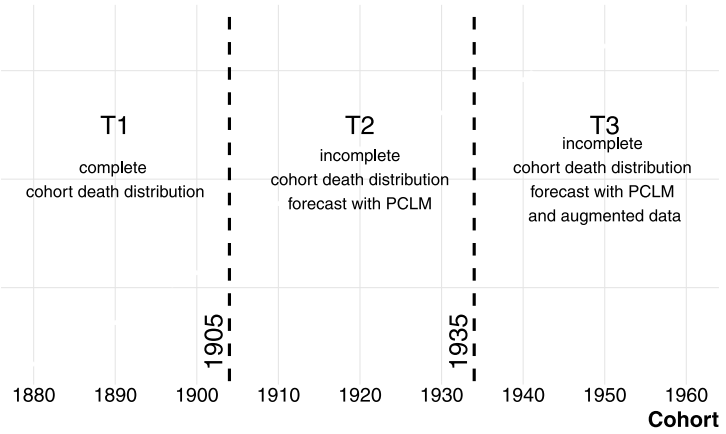

Fig. 1. Analysis scheme. Summary of the time windows T1, T2, T3, in which the cohort death distributions are divided and follow different modeling strategies.

$$\boldsymbol{C} = \begin{array}{c} \\ d_0 \\ d_1 \\ d_2 \\ \vdots \\ d_A \\ d_{M-} \\ d_M \\ d_{M+} \\ 0 \end{array} \begin{array}{c} \begin{array}{ccccccccccccccccc} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_A & \gamma_{A+1} & \cdots & \gamma_{M-1} & \gamma_M & \gamma_{M+1} & \cdots & \gamma_{120} & \gamma_{121} & \cdots & \gamma_{130} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & 1 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} \end{array}.$$

Box I.

many years - see Fig. 2 panels a) and b). The proportion of deaths after the mode does also show an increasing pattern, but is less pronounced compared to the two other time series.

The three time series are forecast using a state space model; more precisely, a structural time series model; i.e., the local linear trend model (LLT) (Durbin & Koopman, 2001). The LLT model allows an explicit estimation of both trend and local components for long- and short-term fluctuations and it is in general more flexible compared with the ARIMA models that are traditionally used (Durbin & Koopman, 2001). The model allows for a stochastic trend that can model an increasing and trending pattern. A detailed description of the LLT model is provided in the Supplementary Material. Residuals of the models are shown in the Supplementary Material and show that the models provide a reasonably good estimate with stationary residuals.

An increasing pattern is most pronounced for the modal age and the number of deaths at the modal age, but an increasing pattern can also be found for the proportion of deaths after the mode for some of the considered populations. For the populations with a less increasing pattern, the slope component (modeling the trending behavior) will be close to zero and not affect the forecast. The proportion of deaths after the mode cannot trend upwards in a long horizon because this will collapse the distribution. Thus, we recommend that the model is re-estimated when new data are available so that new data patterns can be modeled. Furthermore, very long-term forecasts should be interpreted with caution.

Fig. 2 shows both the observed time series, fitted (smoothed with the Kalman smoother), and forecast values using the LLT model for Swedish females. The number of deaths at the mode shows little high-frequency variation compared to the long-term trending behaviour, and thus the fitted values are close to the observed values. The LLT model fits the time series well without any notable signs of autocorrelation in the residuals. See the Supplementary Material for further details and plots of the residuals.

2.4. Estimation of the PCLM model

For both time windows T2 and T3, the PCLM can be estimated by maximizing a penalized log-likelihood function. To avoid negative deaths counts, $\boldsymbol{\gamma}$ is assumed to

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

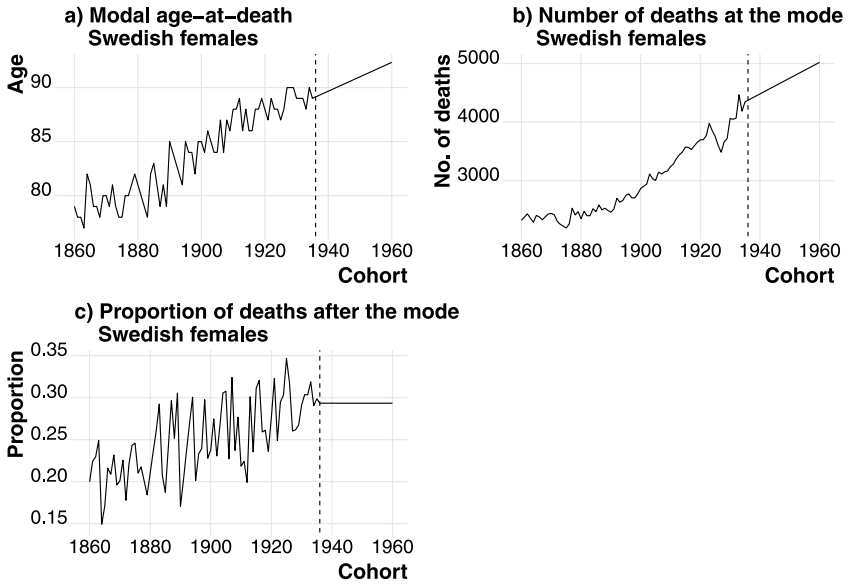

Fig. 2. Demographic data augmentation. Time series of modal age at death (a), number of deaths at the mode (b), and proportion of deaths after the mode (c) of Swedish females by cohorts 1860–1960. Forecast with ARIMA model starting with the 1936 cohort. Source: Human Mortality Database (HMD).

follow the functional form $\boldsymbol{\gamma} = e^{\boldsymbol{X\beta}}$ and thus the maximization process determines the $\boldsymbol{\beta}$ values (Rizzi et al., 2015). A roughness penalty is imposed to ensure smooth death distribution forecasts. We follow the method proposed by Eilers (2007) and use a roughness penalty with an order of difference of two or three. The second-order difference can be written as $\boldsymbol{P} = \lambda\|\boldsymbol{D}_2\beta\|^2 = \lambda\beta'\boldsymbol{D}_2'\boldsymbol{D}_2\beta$, where $\boldsymbol{D}_2$ is a second-order difference matrix. A third-order difference penalty, $\boldsymbol{D}_3$, can be implemented by taking the third-order differences of β instead of the second. We select the order of differences based on an out-of-sample procedure as described in Section 3.1.1 by selecting the order that provides the lowest RMSE in an out-of-sample scheme.

The penalty is multiplied by the weight, λ, which determines the impact of the penalty and thereby how smooth the resulting death distribution should be. λ is selected by a grid search over different values; the optimal λ chosen returns the lowest AIC value. An additional weight is imposed for T3 to ensure that the estimated death distribution is close to the forecasts for d_{M-}, d_{M+}, and d_M, and close to the zero death count at age 121+. $\boldsymbol{V} = \texttt{diag}(\mathbf{1}, 10^2, 10^8, 10^3, 10^3)$ is the diagonal matrix with weights, and the first $A + 1$ elements equal to $\mathbf{1}$ refer to the observed deaths up to age A. In summary, the penalized log likelihood function for a given cohort c can be written as:

$$l^* = l - \frac{\lambda}{2} = \sum_x (d_x \ ln\ \mu_x - \mu_x) - \frac{\lambda}{2}\boldsymbol{P}. \tag{1}$$

Eilers (2007) showed that the maximization can be solved by an Iteratively Re-Weighted Least Squares (IRWLS) algorithm. The maximizing equation (1) leads to the following system of equations:

$$(\check{\boldsymbol{X}}'\boldsymbol{V}\tilde{\boldsymbol{W}}\check{\boldsymbol{X}} + \boldsymbol{P})\boldsymbol{\beta} = \check{\boldsymbol{X}}'\boldsymbol{V}\tilde{\boldsymbol{W}}[\boldsymbol{V}\tilde{\boldsymbol{W}}^{-1}(\boldsymbol{d} - \tilde{\boldsymbol{\mu}}) + \check{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}], \tag{2}$$

where the matrix $\check{\boldsymbol{X}}$ has elements $\check{x}_{ik} = \sum_j c_{ij}x_{jk}\gamma_j/\mu_i$ and can be considered a 'working X' in the IRWLS algorithm and $\tilde{\boldsymbol{W}} = \texttt{diag}(\tilde{\boldsymbol{\mu}})$. The tilde indicates the current values in the algorithm. See Eilers (2007) for further details on the IRWLS. All computations have been performed with the software R (R Core Team, 2018). Demo R code can be found in the Supplementary Material.

2.5. Comparison with the two-dimensional P-spline method

P-spline methods are well established for smoothing mortality rates (Camarda, 2019; Currie et al., 2004). In a two-dimensional setting for forecasting cohort mortality by Continuous Mortality Investigation Bureau (2006), death and exposure counts are modeled assuming that the deaths are subject to a Poisson process. B-splines are used as a regression basis in a penalized IRWLS algorithm. Death counts and exposures are considered as the response and the offset, respectively. An additional penalty enforces smoothness over both age and time. The forecast is treated as a missing value problem that assumes no information for future years. Therefore, the two-dimensional splines method allows smoothing over both age and time and can be used to forecast mortality by treating the forecast period as a missing variable problem (Camarda, 2012; Currie et al., 2004). The main problem with the two-dimensional spline method is its robustness when used for forecasting. Although the method can be very useful and produces accurate forecasts with a short time horizon, it often fails to produce reliable and robust long-term forecasts (Camarda, 2019). We find that the two-dimensional spline method (without further restrictions) is less useful for long-term forecasting because of these robustness problems.

The PCLM model presented in this article is one-(age)dimensional in the sense that each cohort is estimated independently from the other. However, the time

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

dimensional is modelled via the augmented data, where information on the modal age of death is imposed depending on and forecast in time. The PCLM model, therefore, differs from the two-dimensional spline method by using the death distribution as an input variable instead of death rates, and by incorporating the time dimension by augmenting the data instead of specifying age and time parameters.

2.6. Data

Data from four countries, England, Wales, Sweden, and Switzerland, are used to illustrate the fit and forecast of the model. These four countries have been chosen because they have high quality data and sufficiently long data series of deaths and exposure. Actual cohort death rates and exposures have been downloaded from the Human Mortality Database (HMD) (Human Mortality Database, 2019). For females and males in Sweden and England & Wales, we start with the birth cohort 1860. We begin in 1860 because the data quality for Sweden is considerably lower before 1860 (Glei & Lundström, 2019) and we aim to have equally long time series. Data for Switzerland are only available from the 1876 birth cohort and hence this is the first cohort used for both Swiss males and females. For all countries, data are collected for years up to and through 2016.

3. Application and results

For each country analyzed, cohort death rates are calculated by dividing cohort death counts and exposures from the HMD. From the cohort death rates, cohort life tables are derived. Cohort life tables are completed up to the birth cohort of 1905 and are abridged for all more recent birth cohorts. From the estimated complete cohort death distribution, cohort life expectancy is calculated to provide information on the average period that a specific birth cohort may be expected to live. For each country, and separately for each incomplete cohort death distribution, the PCLM is estimated. For cohorts in T2, the simple PCLM for ungrouping is applied, while in T3 the extended version of the PCLM with augmented data is used. A second-order difference penalty provided the lowest forecast error for all of the considered populations and thus we use this throughout. Forecast errors for the third-order difference penalty are shown in the Supplementary Material.

In this section, we report results for Swedish women only as an illustration. Extensive results by country and gender are presented in the Supplementary Material. Fig. 3 shows the age-at-death distribution of Swedish females by cohort: the blue part of the distributions represents the actual raw data (smoothed with the PCLM), whereas the red part of the distributions shows the estimate of the PCLM that allocates the sum of the remaining cohort deaths in single ages up to age 110. Up to the cohort of 1935, the PCLM successfully redistributes the death counts in the death distribution. For the younger cohorts, the PCLM is fed information on the forecast modal age at death and the number of deaths at the mode, as well as

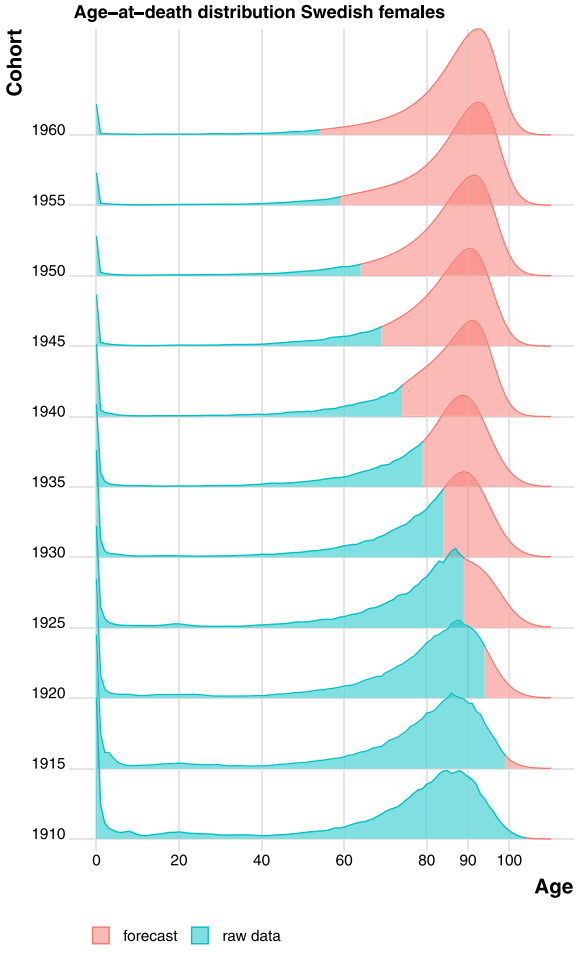

Fig. 3. PCLM d_x forecast. The age-at-death distribution (d_x) of Swedish females by cohorts 1910-1960: raw data and forecast. Cohorts 1910–1935 are forecast with PCLM; cohorts 1936–1960 are forecast with PCLM and augmented data. Source: Human Mortality Database (HMD).

before and after the mode. With such augmented data, the PCLM is proven to be able to estimate cohort death distributions up to the 1960 cohort.

The life expectancy at birth for Swedish females, computed with the estimates of the complete death distributions, is illustrated in Fig. 4. The life expectancy is smoothed on a small scale such that the overall patterns are more visible, and a local polynomial regression fitting smoothing procedure is applied. For both time windows T2 and T3, life expectancy at birth increases over cohorts.

3.1. Model validation

3.1.1. Validation tests leaving out data for T2

The proposed modeling of cohort mortality is validated by an out-of-sample comparison of observed and forecast life expectancies. Because actual life expectancy can only be calculated for complete cohorts, only data until the 1905 birth cohort can be used for validation. Death distribution forecasts are constructed, fitting the PCLM to a restricted part of the data and leaving out 10, 15, 20, 25,

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

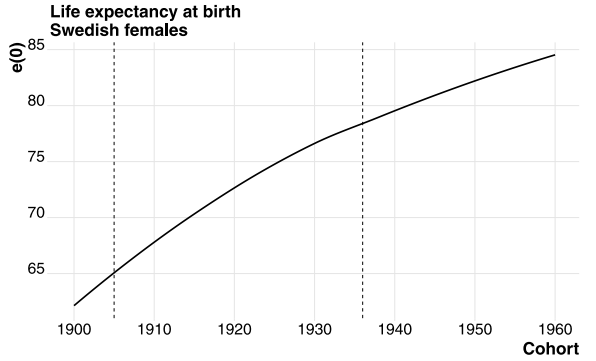

Fig. 4. PCLM $e(0)$ forecast, smoothed. Life expectancy at birth $e(0)$ of Swedish females by cohorts 1900–1960. Cohorts 1905–1935 are forecast with PCLM; cohorts 1936–1960 are forecast with PCLM and augmented data. Source: HMD.

and 30 years to be forecast. From the perspective of the cohort, when 10 years are omitted, this means that observations for the 10 oldest ages are left out for validation of the 1905 cohort, the 9 oldest ages for the 1904 cohort, and so on until the 1896 cohort. Life expectancies are then calculated for the cohorts for which data are left out or until no earlier cohort is available. For example, when 10 years are left out from the analysis, life expectancies for cohorts from 1896 to 1905 are computed. A similar procedure is performed, leaving 15, 20, 25, and 30 years out. Observed and forecast life expectancies are compared by calculating the root mean squared error (RMSE):

$$RMSE = \sqrt{\sum_{i=1}^{H}(\hat{e}_{x,i} - e_{x,i})^2/H}, \tag{3}$$

where $\hat{e}_{x,i}$ is the life expectancy forecast, $e_{x,i}$ is the one observed, and H is the number of years left out. Results are reported in Table 1 for life expectancies at age 0, 50, and 65 years.

When calculating life expectancy, a combination of in-sample fitted death distributions and out-of-sample death distribution forecasts are used. Because life expectancy at age x only depends on ages higher than or equal to x, the RMSE at age 65 will depend more on out-of-sample forecast values compared to the RMSE at ages of 0 and 50 years. A similar argument holds for the RMSE at age 50 compared to age 0.

We also measure the forecast bias of the model by calculating the mean forecast error using a similar scheme as with the RMSE; that is:

$$ME = \sum_{i=1}^{H}(\hat{e}_{x,i} - e_{x,i})/H. \tag{4}$$

Table 1 shows the accuracy of the results for all analyzed populations; i.e. England & Wales, Sweden and Switzerland, females, and males. The PCLM in general forecasts life expectancy accurately. Even when 30 years are left out, the RMSE error remains below 1 year. The RMSE error for life expectancy increases when the forecast horizon widens, which is to be expected because longer and less-certain forecasts are being calculated. The PCLM fits all the analyzed countries equally well, with no systematically lower accuracy errors found for one particular country or sex. The only difference occurs with the 30-years analysis, for which larger errors are found for the female population, but we have no reason to expect this to be a general tendency.

Table 1
Accuracy measures: RMSE observed vs. fitted and forecast life expectancy for selected ages and out-of-sample periods.

Years left out	10 years	15 years	20 years	25 years	30 years
Switzerland, females					
RMSE for $e_{0,i}$	0.0016	0.0046	0.0324	0.0389	0.3357
RMSE for $e_{50,i}$	0.0040	0.0081	0.0429	0.0520	0.2858
RMSE for $e_{65,i}$	0.0035	0.0074	0.0495	0.0595	0.2953
Sweden, females					
RMSE for $e_{0,i}$	0.0028	0.0043	0.0120	0.0378	0.3380
RMSE for $e_{50,i}$	0.0047	0.0069	0.0168	0.0503	0.3142
RMSE for $e_{65,i}$	0.0045	0.0063	0.0173	0.0558	0.3305
England & Wales, females					
RMSE for $e_{0,i}$	0.0096	0.0637	0.2164	0.3665	0.4231
RMSE for $e_{50,i}$	0.0091	0.0481	0.1515	0.2498	0.3120
RMSE for $e_{65,i}$	0.0086	0.0471	0.1466	0.2403	0.3094
Switzerland, males					
RMSE for $e_{0,i}$	0.0007	0.0018	0.0098	0.0240	0.1731
RMSE for $e_{50,i}$	0.0037	0.0040	0.0143	0.0366	0.1305
RMSE for $e_{65,i}$	0.0020	0.0031	0.0175	0.0454	0.1312
Sweden, males					
RMSE for $e_{0,i}$	0.0010	0.0207	0.0564	0.1289	0.2381
RMSE for $e_{50,i}$	0.0031	0.0160	0.0420	0.0988	0.1775
RMSE for $e_{65,i}$	0.0019	0.0171	0.0420	0.0995	0.1785
England & Wales, males					
RMSE for $e_{0,i}$	0.0014	0.0041	0.0403	0.1521	0.2359
RMSE for $e_{50,i}$	0.0018	0.0039	0.0326	0.1152	0.1780
RMSE for $e_{65,i}$	0.0011	0.0046	0.0344	0.1190	0.1835

Table 2 shows the mean forecast error (ME) for all the analyzed populations. The ME shows a similar result to the RMSE in terms of accuracy, and displays no systematic bias in the model as negative and positive ME errors are found.

3.1.2. Sensitivity analysis for T2

In addition to the out-of-sample test, we validate the model by comparing actual and fitted death distributions. For such a comparison, again, only complete cohorts can be analyzed. Therefore, data from the first cohort available (see Section 2.6 for country- and gender-specific starting cohorts) up to the 1905 cohort are used. For each cohort analyzed, we assume that the last observed age is 65, 70, 75, and 80. We apply the PCLM to the artificially uncompleted cohorts and compare the fit with the actual death distributions. See Fig. 5 for an example of results; i.e., Swedish females born in 1901 with an assumed last age observed of 75.

To further assess the quality of the results, we use the Kullback–Leibler (KL) divergence for each cohort:

$$KL = \int f(x) \ln\left[\frac{f(x)}{\hat{f}(x)}\right] dx, \tag{5}$$

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

Table 2
Mean forecast error (ME) observed vs. fitted and forecast life expectancy for selected ages and out-of-sample periods.

Years left out	10 years	15 years	20 years	25 years	30 years
Switzerland, females					
ME for $e_{0,i}$	0.000	0.003	0.001	0.013	−0.043
ME for $e_{50,i}$	0.000	0.003	0.002	0.018	−0.059
ME for $e_{65,i}$	0.001	0.004	0.002	0.020	−0.067
Sweden, females					
ME for $e_{0,i}$	−0.002	0.001	−0.002	0.005	−0.017
ME for $e_{50,i}$	0.000	0.001	−0.002	0.007	−0.021
ME for $e_{65,i}$	0.000	0.001	−0.002	0.007	−0.024
England & Wales, females					
ME for $e_{0,i}$	0.000	0.001	0.004	0.012	0.032
ME for $e_{50,i}$	0.000	0.001	0.005	0.016	0.019
ME for $e_{65,i}$	0.000	0.001	0.007	0.019	0.017
Switzerland, males					
ME for $e_{0,i}$	0.000	0.000	0.000	0.011	−0.007
ME for $e_{50,i}$	0.000	0.001	0.001	0.016	−0.010
ME for $e_{65,i}$	0.000	0.001	0.001	0.020	−0.013
Sweden, males					
ME for $e_{0,i}$	0.000	0.001	0.002	0.003	0.011
ME for $e_{50,i}$	0.000	0.001	0.002	0.004	0.016
ME for $e_{65,i}$	0.000	0.001	0.002	0.005	0.019
England & Wales, males					
ME for $e_{0,i}$	0.000	0.000	0.001	0.005	0.009
ME for $e_{50,i}$	0.000	0.000	0.001	0.008	0.014
ME for $e_{65,i}$	0.000	0.000	0.002	0.010	0.018

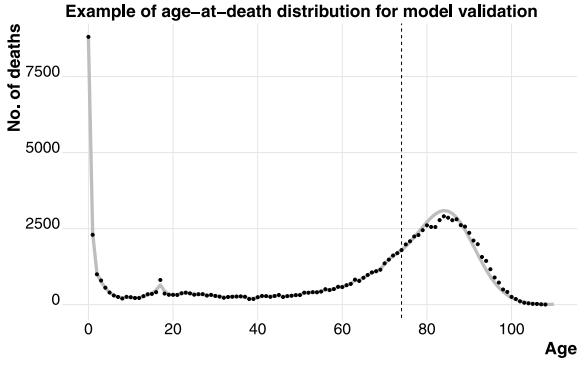

Fig. 5. Sensitivity analysis. Example of an age-at-death distribution of one cohort: Fitted distribution with PCLM as if the last age observed is 75 (smooth line) and actual observed data (dots) for Swedish females. Source: HMD.

where $f(x)$ stands for the density of the truly observed age-at-death distribution, and $\hat{f}(x)$ is the density of the forecast age-at-death distribution. The KL divergence measures the information lost when the fitted density is used to approximate the true one. The minimum of the divergence equals 0. For each country and gender, an average of the KL for each cohort analyzed is reported in Table 3: as a general pattern, the KL divergence decreases as the last age observed of a cohort increases.

Additionally, we checked how accurately the PCLM estimates the mode when the last age assumed to be observed for a cohort is 75. We found that the model correctly sets the modal age at death within ±4 years

Table 3
Divergence measures: KL divergence observed vs. fitted and forecast age-at-death distributions for assumed last cohort ages.

Last cohort age	age 65	age 70	age 75	age 80
Switzerland, females				
KL for $\boldsymbol{d}$	0.0656	0.0319	0.0149	0.0091
Sweden, females				
KL for $\boldsymbol{d}$	0.0284	0.017	0.0144	0.0108
England & Wales, females				
KL for $\boldsymbol{d}$	0.0167	0.0187	0.0101	0.0039
Switzerland, males				
KL for $\boldsymbol{d}$	0.0272	0.0108	0.0103	0.0098
Sweden, males				
KL for $\boldsymbol{d}$	0.0133	0.0082	0.0116	0.0112
England & Wales, males				
KL for $\boldsymbol{d}$	0.0097	0.008	0.0064	0.0066

of age and estimates the number of deaths at the mode with a difference from the raw numbers within ±400 deaths (out of 100.000). We found a slight tendency of the PCLM to overestimate the modal age at death and underestimate the number of deaths at the mode when the last age assumed to be observed for a cohort is 75, as reported in Fig. 5.

4. Discussion and conclusion

We have presented a novel method to forecast age-at-death distributions of cohorts not yet extinct based on the penalized composite link model. The basis for the forecast is the death counts of a cohort life table. For the studied populations, we constructed cohort life tables using cohort mortality data from the HMD. The grounding assumptions of the proposed method are (i) the Poisson distributed death counts d_i of a cohort life table, (ii) the smoothness of the forecast age-at-death distribution $\boldsymbol{\gamma}$, and (iii) no deaths occurring after age 120.

The PCLM applied to T2 cohorts - cohorts where the last observed age of death is not far from the modal age at death - only makes use of these three modest assumptions. No target parametric model of the forecast age-at-death distribution is given. Hence, the PCLM smoothly redistributes the remaining deaths after the last age of death is observed. In the presented application, we are able to forecast with the simple PCLM age-at-death distributions of cohorts born up to 1935; i.e., cohorts where the last ages of deaths observed correspond to around 30 years before the cohorts' natural extinction. For younger cohorts - T3 cohorts - where the modal age at death is far from being observed, additional assumptions are required. We set up the necessary constraints by augmenting the input data from the PCLM with prior information on (iv) the position of the modal age at the death of the distribution, and (v) the proportion of deaths before and after the mode. In the presented study, the modal age at death with the corresponding number of death counts are estimated with a LLT model. This works accurately for the populations we studied when both the

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.

modal age at death and the number of deaths at the mode show a linearly increasing pattern over calendar years. The PCLM with augmented data succeeds in forecasting age-at-death distributions up to the birth cohort of 1960; i.e., cohorts where the last age of death observed corresponds to around 50 years before the cohorts' natural extinction and have therefore not yet reached retirement age. The proposed model can be applied both to forecast mortality for country populations and sub-populations; for instance, to groups of insured individuals.

The suggested forecasting method for T3 could show limitations when the mortality pattern over time is not relatively stable. For example, if an age-at-death distribution for a particular cohort deviates from the pattern of the neighboring cohorts, the predicted deaths and the modal age at death of that particular cohort might become inaccurate. Weights in the PCLM fitting methodology allow for some flexibility but cannot handle large deviations between neighboring cohorts, and odd fitted death distributions can occur in such situations.

Most mortality forecasts are based on period mortality data, despite the demand from public and private institutions to obtain forecasts for actual cohorts rather than synthetic populations. True cohort forecasts are rare in the literature and this article is one of the few that successfully forecasts cohort mortality. Here, we have introduced a novel method to complete the mortality trajectories of non-extinct cohorts. The completion of cohort mortality is not only useful for forecasting but also in comparisons of cohorts over time.

Acknowledgments

The research and publication of this paper was supported by AXA Research Fund through the research programme AXA Chair Longevity Research.

The authors wish to thank the reviewers for their useful comments and Maarten J. Wensink for critical comments and suggestions.

Appendix A. Supplementary data

Supplementary material Demo R code; Extensive methods description LLT; Extensive results related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2020.03.003.

References

Barker, D. (2004). The developmental origins of adult disease. *Journal of the American College of Nutrition*, *23*, 588S–595S.

Basellini, U., & Camarda, C. (2019). Modelling and forecasting adult age-at-death distributions. *Population Studies*, *73*, 119–138.

Basellini, U., Kjærgaard, S., & Camarda, C. G. (2019). An age-at-death distribution approach to forecast cohort mortality. In *Population association of America 2019 annual meeting program*.

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

Bergeron-Boucher, M. P., Simonacci, V., Oeppen, J., & Gallo, M. (2018). Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: An application to Canadian provinces and territories. *North American Actuarial Journal*, *22*, 92–118.

Booth, H., Maindonald, J., & Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, *56*, 325–336.

Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, *3*, 3–43.

Cairns, A. J. G., Blake, D., & Dowd, K. (2006). A two factor model for stochastic mortality with parameter uncertainty: Theory and Calibration. *The Journal of Risk and Insurance*, *73*, 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., et al. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, *13*, 1–35.

Camarda, C. (2012). Mortalitysmooth: An R package for smoothing Poisson counts with P-splines. *Journal of Statistical Software, Articles*, *50*, 1–24.

Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Population Studies*, *41*, 1091–1130.

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

Chiou, J. M., & Müller, H. G. (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *Journal of the American Statistical Association*, *104*, 572–585.

Continuous Mortality Investigation Bureau (2006). Stochastic projection methodologies: further progress and p-spline model features, example results and implications. URL https://www.actuaries.org.uk/system/files/documents/pdf/wp20revised.pdf.

Currie, I. (2012). Forecasting with the age-period-cohort model?. In *27th International workshop on statistical modelling*. Prague.

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

Doll, R., & Hill, A. (1950). Smoking and carcinoma of the lung. *British Medical Journal*, *2*, 739–748.

Doyle, J. T., Dawber, T. R., Kannel, W. B., Heslin, A. S., & Kahn, H. A. (1962). Cigarette smoking and coronary heart disease. *New England Journal of Medicine*, *266*, 796–801.

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press.

Eilers, P. H. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. *Statistical Modelling*, *7*, 239–254.

Elo, I. T., & Preston, S. H. (1992). Effects of early-life conditions on adult mortality: A review. *Population Index*, *58*, 186–212.

Glei, D., & Lundström, J. (2019). About mortality data for Sweden.

Human Mortality Database (2019). Human mortality database. 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 15.01.2019).

Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography*, *50*, 261–283.

Janssen, F., & Kunst, A. (2005). Cohort patterns in mortality trends among the elderly in seven European countries, 1950–99. *International Journal of Epidemiology*, *34*, 1149–1159.

Kjærgaard, S., Ergemen, Y. E., Kallestrup-Lamb, M., Oeppen, J., & Lindahl-Jacobsen, R. (2019). Forecasting causes of death by using compositional data analysis: the case of cancer deaths. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *68*, 1351–1370.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U. S. Mortality. *Journal of the American Statistical Association*, *87*, 659–671.

Lee, R. D., & Miller, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, *38*, 537–549.

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

Preston, S., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and modeling population processes*. Oxford: Blackwell Publishers.

Preston, I. H., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography*, *43*, 631–646.

R Core Team (2018). *R: A language and environment for statistical computing.R version 3.5.1*. Vienna, Austria: R Foundation for Statistical Computing, URL http://www.R-project.org/.

Renshaw, A., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance Mathematics and Economics*, *38*, 556–570.

Rizzi, S., Gampe, J., & Eilers, P. H. C. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology*, *182*, 138–147.

Stoeldraijer, L., van Duin, C., van Wissen, L. J. G., & Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the netherlands. *Demographic Research*, *29*, 323–354.

Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., Riffe, T., Boe, C., et al. (2017). *Methods protocol for the human mortality database*: *Technical report*, Human Mortality Database.

Please cite this article as: S. Rizzi, S. Kjærgaard, M.-P. Bergeron Boucher et al., Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. International Journal of Forecasting (2020), https://doi.org/10.1016/j.ijforecast.2020.03.003.