Lee–Carter cohort mortality forecasts

Ugofilippo Basellini[1,2] 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*

Draft paper for the EPC 2020 conference

October 29, 2019

Abstract

Mortality forecasting has recently stimulated great interest in academics and financial sector practitioners due to the increasing challenges posed by continuous longevity improvements. Many efforts have been directed to model and forecast mortality from the conventional age-period perspective. Conversely, models to forecast cohort mortality are scarce in the demographic and actuarial literature. In this article, we fill this gap by adapting the seminal model of Lee and Carter (1992) and its most successful extensions to the structure of cohort mortality data, which is characterized by missing data corresponding to periods beyond the last available year of data collection. Our approach allows us to derive a complete age-cohort mortality surface simply by estimating the model's parameters, thereby completing the mortality experience of non-extinct cohorts. We apply our methodology to Swedish female cohort mortality data at ages 0–100 for the cohorts 1900–1987 obtained from the Human Mortality Database.

Keywords: Mortality forecasting · Lee–Carter model · Smoothing · Newton–Raphson algorithm · Cohort life tables

1 Introduction

Forecasting mortality is an essential component for the computation of future pension and health care expenditures. Continuous and widespread increases in life expectancy (Riley, 2001; United Nations, 2019) imply rapidly growing costs for the provisions of retirement products and care for the elderly (Currie, 2016). As such, it is unsurprising that the forecast of mortality has been a subject of great interest for public and private sectors as well as for the academic community.

The last three decades have witnessed the flourishing of mortality forecasting as a research field owing to the introduction of stochastic methodologies to project mortality (Booth and Tickle, 2008). Much of this success has been stimulated by the seminal contribution of Lee and Carter (1992), whose model is still widely used by statistical offices and international agencies today. The vast majority of the recently proposed approaches to forecast mortality, including the Lee–Carter (LC) model, is based on an age-period perspective: the principal goal is to forecast mortality patterns of a given population into the future. For example, a common question that such models try to answer is: "What will be the value of life expectancy at birth for the specific population fifty years from now?".

A different approach to forecasting mortality, which has been largely overlooked and unexplored so far, consists in shifting from an age-period to an age-cohort perspective. Completing the mortality pattern of non-extinct cohorts is relevant for practical purposes: for example, insurance companies are interested in predicting the future longevity of groups of people born in specific cohorts. In such settings, cohort forecasts are conventionally obtained by first forecasting mortality in a period fashion, and then extracting cohort mortality patterns from the diagonals of the projected Lexis surface. Although widely used, this approach is rather counter-intuitive and inefficient, and it can generate implausible prediction intervals (van Raalte et al., 2018).

In addition to practical interest, analyses and forecasts of cohort mortality have one fundamental advantage over those based on the period perspective. Cohort mortality developments are actually observed, as opposed to the hypothetical situation assumed in period life tables. As a result, cohort measures of mortality are not affected by "tempo effects", i.e. distortions of period measures caused by not considering the mortality history of the population (Bongaarts and Feeney, 2002, 2003).

Models for forecasting cohort mortality are relatively few in the literature, mostly due to the heavy data demands that such models require (Booth, 2006). One early attempt in this direction was made by the Continuous Mortality Investigation (2007), that employed two-dimensional P-splines model (Currie et al., 2004) to fit and forecast cohort mortality in England & Wales. Shortly afterwards, Chiou and Müller (2009) proposed to forecast cohort log-hazard functions using a functional data approach based on Swedish mortality data. More recently, three different approaches have been proposed to model and forecast cohort mortality from age-at-death distributions (Zanotto and Mazzuco, 2017; Basellini et al., 2019; Rizzi et al., 2019).

In this article, we propose a more direct and alternative approach to forecast mortality that is solely based on cohort data. Specifically, we propose to model and forecast cohort mortality generalizing the most recent and successful extensions of the LC model. Our contribution in this respect is twofold: first, we provide a comprehensive framework that includes the LC model and its main extensions in a single setting. Second, we adapt the estimation procedure of the LC parameters to suit the structure of cohort mortality data which, by construction,

contain missing values. This allows us to complete the mortality experience of non-extinct cohorts. We show an application of our methodology to Swedish female cohort mortality using data from the Human Mortality Database (HMD, 2019). As a future application, our methodology could be applied, for example, to complete the cohort life tables for all countries and sexes in the HMD.

2 Methods

2.1 Notation and data

In this section, we introduce the notation and the data that we employ throughout this paper. Let $Y_{x,c}$ be a non-negative random variable denoting the death counts in a population at age x and for the cohort c. The realizations of $Y_{x,c}$ are the observed number of deaths $y_{x,c}$ corresponding to the central exposure to the risk of death $e_{x,c}$. The force of mortality and central death rates are denoted by $\mu_{x,c}$ and $m_{x,c}$, respectively, with the empirical estimate of the latter being equal to $\hat{m}_{x,c} = y_{x,c}/e_{x,c}$.

In what follows, we assume that the force of mortality $\mu_{x,c}$ remains constant over each year of age (from age x to $x+1$) and birth cohort (from cohort c to $c+1$). This assumption implies that: (i) $\mu_{x,c}$ approximates the force of mortality at exact age $x+\frac{1}{2}$ and exact cohort $c+\frac{1}{2}$, and (ii) central death rates are the maximum likelihood estimators of the force of mortality (Currie, 2016).

Furthermore, following the prominent work of Brillinger (1986), a standard assumption that we use throughout this article is that the random variable $Y_{x,c}$ follows a Poisson process with expected values equal to the product of exposure and force of mortality:

$$Y_{x,c} \sim \mathcal{P}(e_{x,c}\,\mu_{x,c})\,. \tag{1}$$

Let us further denote by $\eta_{x,c}$ the linear predictor, which in the Poisson setting is associated to the canonical log link function, i.e. $\eta_{x,c} = \ln(\mu_{x,c})$.

Analyses in this paper are performed on observed death counts $y_{x,c}$ and central exposures $e_{x,c}$, arranged into two matrices $\boldsymbol{Y} = (y_{x,c})$ and $\boldsymbol{E} = (e_{x,c})$. Each matrix has dimensions $m \times n$, where rows are classified by m single ages at death x, and columns by n single birth cohorts c, respectively.

Data are retrieved from the Human Mortality Database (HMD, 2019), which provides free access to detailed, consistent and high quality historical mortality data for 41 different countries or areas (Barbieri et al., 2015). For this extended abstract, we show an application of our methodology to Swedish female cohort mortality, for which we consider the age range 0–100 years and the most recent cohorts 1900–1987.

The illustration of the matrix $\boldsymbol{Y}$ is useful to directly grasp the structure of cohort mortality data. Unlike the more familiar case of age-period data, here the matrices $\boldsymbol{Y}$ and $\boldsymbol{E}$ contain missing data, corresponding to periods beyond the last available year of data collection. For the Swedish data considered here, 2016 is the last available year of collected data. This implies that: (i) data are fully observed for the age-groups $x = 0, \ldots, 29$ for all cohorts, (ii) cohorts $c = 1900, \ldots, 1916$ are fully observed for all ages, and (iii) for the cohorts $c = 1917, \ldots, 1987$, data are increasingly missing from age 100 downwards. Here we illustrate

the $\boldsymbol{Y}$ matrix denoting by "na" the missing data:

$$\boldsymbol{Y} = (y_{x,c}) = \begin{bmatrix} y_{0,1900} & \cdots & y_{0,1916} & y_{0,1917} & y_{0,1918} & \cdots & y_{0,1987} \\ y_{1,1900} & \cdots & y_{1,1916} & y_{1,1917} & y_{1,1918} & \cdots & y_{1,1987} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ y_{29,1900} & \cdots & y_{29,1916} & y_{29,1917} & y_{29,1918} & \cdots & y_{29,1987} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \iddots & na \\ y_{98,1900} & \cdots & y_{98,1916} & y_{98,1917} & y_{98,1918} & \iddots & \vdots \\ y_{99,1900} & \cdots & y_{99,1916} & y_{99,1917} & na & \cdots & na \\ y_{100,1900} & \cdots & y_{100,1916} & na & na & \cdots & na \end{bmatrix}. \tag{2}$$

2.2 Choice of Lee–Carter models

Here, we describe the four models that we employ for modelling and forecasting cohort mortality. In the presentation of the models, which have been originally proposed for the conventional age-period structure of mortality data, we adapt notation to suit the novel age-cohort perspective employed in this article.

The first model that we consider, and that we denote M1 in the remainder of the article, is the original Lee–Carter (LC, 1992) model, which describes a matrix of logged age-specific central death rates with the linear predictor:

$$\eta_{x,c} = \alpha_x + \beta_x \kappa_c \,, \tag{3}$$

where α_x captures the average shape of age-specific mortality, β_x the rate of mortality improvement at age x, and κ_c the general level of mortality for the cohort c. Since the model is undetermined, two standard constraints are used to ensure model identification:

$$\sum_x \beta_x = 1 \quad \text{and} \quad \sum_c \kappa_c = 0 \,. \tag{4}$$

Secondly, we consider the extension of the LC linear predictor proposed by Renshaw and Haberman (2003): we go beyond a single principal component, i.e. the age-time interaction term $\beta_x \kappa_c$. Here, we focus on the case of two principal components; as such, the linear predictor of our second model, denoted M2, can be expressed as:

$$\eta_{x,c} = \alpha_x + \beta_x \kappa_c + \beta_x^{(2)} \kappa_c^{(2)} \,, \tag{5}$$

where the parameters have the same interpretation of the LC model: $\kappa_c^{(2)}$ accounts for the second main trend in mortality over cohorts and $\beta_x^{(2)}$ modulate this trend across ages. In addition to Equation (4), two further constraints are needed to ensure model identifiability:

$$\sum_x \beta_x^{(2)} = 1 \quad \text{and} \quad \sum_c \kappa_c^{(2)} = 0 \,. \tag{6}$$

Both models M1 and M2 can be further extended to include an additional parameter vector that accounts for period effects. Let γ_t denote the period effect at time $t = c + x$, where the periods considered in the data can be easily derived from inverting the well-known age-period-cohort relationship $cohort = period - age$. Note that (i) period effects can be considered as the counterpart of cohort effects in the analysis of age-period data (see Renshaw and Haberman, 2006), (ii) a varying number of available data points are associated to each γ_t. The linear

predictors in Equations (3) and (5) can be augmented with the inclusion of this additional parameter vector, which is subject to the constraint:

$$\sum_{t=c+x} \gamma_t = 0. \tag{7}$$

Table 1 summarizes the four models just presented in terms of their linear predictors and number of constraints needed to ensure the model's identifiability.

Model	Linear predictor $\eta_{x,c}$	Constraints
M1	$\alpha_x + \beta_x \kappa_c$	2
M2	$\alpha_x + \beta_x \kappa_c + \beta_x^{(2)} \kappa_c^{(2)}$	4
M1 with period effects	$\alpha_x + \beta_x \kappa_c + \gamma_t$	3
M2 with period effects	$\alpha_x + \beta_x \kappa_c + \beta_x^{(2)} \kappa_c^{(2)} + \gamma_t$	5

Table 1. The four Lee–Carter models considered for modelling and forecasting cohort mortality data, with their linear predictor $\eta_{x,c}$ and the number of constraints needed to ensure each model's identifiability.

2.3 Parameters' estimation

Traditionally, the original LC model is estimated by ordinary least-squares using singular value decomposition (SVD, see Lee and Carter, 1992). The main drawback of this approach is the assumption of homoskedastic and normally distributed errors, which is a fairly unrealistic assumption for human mortality. The force of mortality is indeed more variable at older than at younger ages because of the smaller number of deaths (Brouhns et al., 2002). Moreover, SVD is a plain mathematical approach and it cannot be implemented in the presence of missing data, which is the case for the cohort data that we analyse.

Consequently, as suggested by Alho (2000) and implemented by Brouhns et al. (2002) in the original LC setting, we work within a Poisson framework (cf. Equation (1)), and we derive maximum-likelihood estimates of the models' parameters. Specifically, given a linear predictor $\eta(\boldsymbol{\theta})_{x,c} = \ln\left(\mu(\boldsymbol{\theta})_{x,c}\right)$ that depends on a vector of parameters $\boldsymbol{\theta}$, estimation can be achieved by maximising the Poisson log-likelihood:

$$\ln \mathcal{L}\left(\boldsymbol{\theta} | \, y_{x,c}, e_{x,c}\right) \propto \sum_{x,c} \left\{ y_{x,c} \, \eta(\boldsymbol{\theta})_{x,c} - e_{x,c} \, \mu(\boldsymbol{\theta})_{x,c} \right\}, \tag{8}$$

where the linear predictor $\eta(\boldsymbol{\theta})_{x,c}$ can follow any of the four models in Table 1, and the parameters are subject to the model's specific constraints. The advantage of this assumption is that the error terms now follow a non-additive heteroscedastic error structure, which is a more suitable assumption for modeling human mortality.

Here, we propose two modifications to the log-likelihood in Eq. (8). First, in order to account for the structure of cohort data that displays missing data, we employ a matrix of weights $\boldsymbol{W} = (w_{x,c})$, whose elements are equal to one if data are observed, and zero if data are missing, i.e. where we have "na" in (2). Second, following the suggestion of Delwarde et al. (2007), we include penalization terms for the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ schedules to obtain smooth parameters. These, in turn, avoid the irregularities typically found in LC fitted and projected life tables.

Let us consider the original LC model (M1 in Table 1), so that $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}]$. We can then rewrite the penalised log-likelihood as:

$$\ln \mathcal{L}^{\star}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa} | \, y_{x,c}, e_{x,c}, w_{x,c}) \propto \sum_{x,c} w_{xc} \{ y_{x,c} \, \eta(\boldsymbol{\theta})_{x,c} - e_{x,c} \, \mu(\boldsymbol{\theta})_{x,c} \} - \frac{1}{2} \lambda_{\alpha} \boldsymbol{\alpha}' \boldsymbol{D}' \boldsymbol{D} \boldsymbol{\alpha} - \frac{1}{2} \lambda_{\beta} \boldsymbol{\beta}' \boldsymbol{D}' \boldsymbol{D} \boldsymbol{\beta}, \tag{9}$$

where the smoothing parameters λ_{α} and λ_{β} control the amount of smoothness in the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively, and $\boldsymbol{D}$ is the second order difference $m \times (m-2)$ matrix:

$$\boldsymbol{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots \\ 0 & 1 & -2 & 1 & 0 & \cdots \\ 0 & 0 & 1 & -2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{10}$$

It should be pointed out that the order of difference (second order here) does not play a relevant role in this context. For the model M2 (without and with period effects), we include an additional penalization term for $\boldsymbol{\beta}^{(2)}$, while the parameters $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are always left unpenalised to better capture the cohort- and time-specific fluctuations of mortality.

The uni-dimensional iterative Newton-Raphson method works well to maximise Eq. (9): at each iteration step ν+1, a single set of parameters is updated while keeping fixed all remaining parameters at their current estimates using the formula:

$$\hat{\theta}^{(\nu+1)} = \hat{\theta}^{(\nu)} - \frac{\partial \ln \mathcal{L}^{(\nu)} / \partial \theta}{\partial^2 \ln \mathcal{L}^{(\nu)} / \partial \theta^2}. \tag{11}$$

The iterative procedure is stopped once a very small increase in the log-likelihood function is achieved (e.g. an increase smaller than 10^{-6}). Analytic solutions for the four models are provided in Appendix A.

2.4 Estimation and forecasting procedure

An essential advantage of the proposed approach is that models M1 and M2 allows us to complete cohort mortality without conducting any forecasting routine: not available data are directly filled within the estimation procedure. On the contrary, models with period effects require the forecast of the (erratic) period effect to reconstruct the mortality surface. However, both models with period effects are introduced here to identify, and eventually disregard, specific years that will likely jeopardize the final outcomes.

We thus propose a two-step approach to fit and forecast cohort mortality.

The first step consists in identifying period outlier effects, which we wish to exclude from the estimation of the model's parameters. Indeed, we do not want the reconstruction and completion of cohort mortality data to be affected by outlier period effects. Note that Lee and Carter (1992) proposed a similar yet different approach to ours: they included a dummy variable in the time-series of their κ_t corresponding to the influenza pandemic (i.e. the Spanish flu) of 1918, so that mortality forecasts are not influenced by this outlier effect. Unlike the original LC model, we carry out a maximum likelihood estimation and the advantage of our approach is that we can exclude outlier periods within the estimation procedure; as such, all parameters (and not only the time index) are not influenced by outlier events.

In order to pinpoint such outliers, we fit the augmented models M1 and M2 with period effects, and we analyse the time-series of the estimated parameter $\hat{\gamma}_t$ for each model separately.

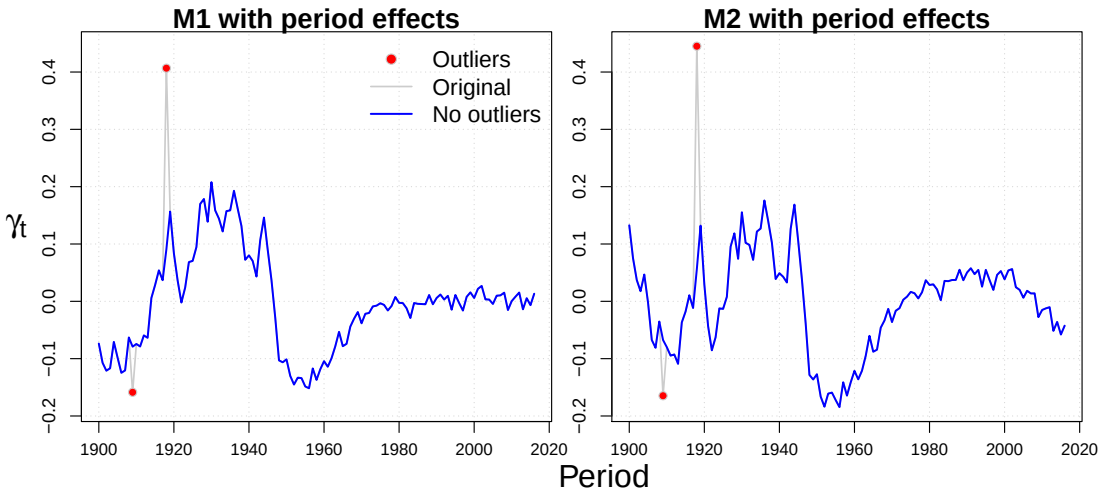

Figure 1. Estimated period effects γ_t (grey lines), detected outliers (red points) and re-estimated time series without outliers (blue lines) for the models M1 and M2 with period effects.
Source: authors' own elaborations using data of the Human Mortality Database (2019).

We identify additive outliers in the time-series of γ_t by using the procedure proposed by Chen and Liu (1993), which is readily implemented in the `R` package `tsoutliers` (López-de Lacalle, 2019).

In the second step, we fit the two models M1 and M2 excluding the outlier periods identified in the first step. This can be easily achieved by assigning zero weights to the (diagonal) elements of $\boldsymbol{W}$ in Equation (9) corresponding to the outlier years. The resulting estimated parameters of the two models readily allow to obtain an estimation of the complete mortality surface over all ages and cohorts, i.e. we complete the mortality experience of partially observed cohorts without forecasting any time-series.

For all the models considered in these two steps, model selection is performed by finding the optimal values of the $\lambda_{\boldsymbol{\alpha}}$ and $\lambda_{\boldsymbol{\beta}}$ (and $\lambda_{\boldsymbol{\beta}^{(2)}}$ for M2) that minimize the Bayesian Information Criterion (BIC, Schwarz, 1978). Minimization is achieved by performing a multidimensional grid search over different combinations of $\lambda_{\boldsymbol{\alpha}}$ and $\lambda_{\boldsymbol{\beta}}$ (and $\lambda_{\boldsymbol{\beta}^{(2)}}$ for M2).

3 Results

In this section, we show an application of our methodology to Swedish female cohort mortality using data from the Human Mortality Database (HMD, 2019). Specifically, we consider the most recent cohorts 1900–1987 and the age range 0–100 years.

As discussed in Subsection 2.4, we start by fitting the augmented M1 and M2 models with period effects, and we analyse the time series of the estimated $\hat{\gamma}_t$ to identify outlier effects. Figure 1 shows the originally estimated $\hat{\gamma}_t$ for the two models as well as the detected outliers. In both cases, 1909 and 1918 were recognized as outlier observations. Whereas the reason behind 1918 is immediate (i.e. Spanish Flu), further research will be necessary to better grasp the detection of 1909 as outlier year.

In the second step, we fit models M1 and M2 removing the two outliers from the estimation procedure. Figure 2 shows the estimated parameters of the two models. The parameters

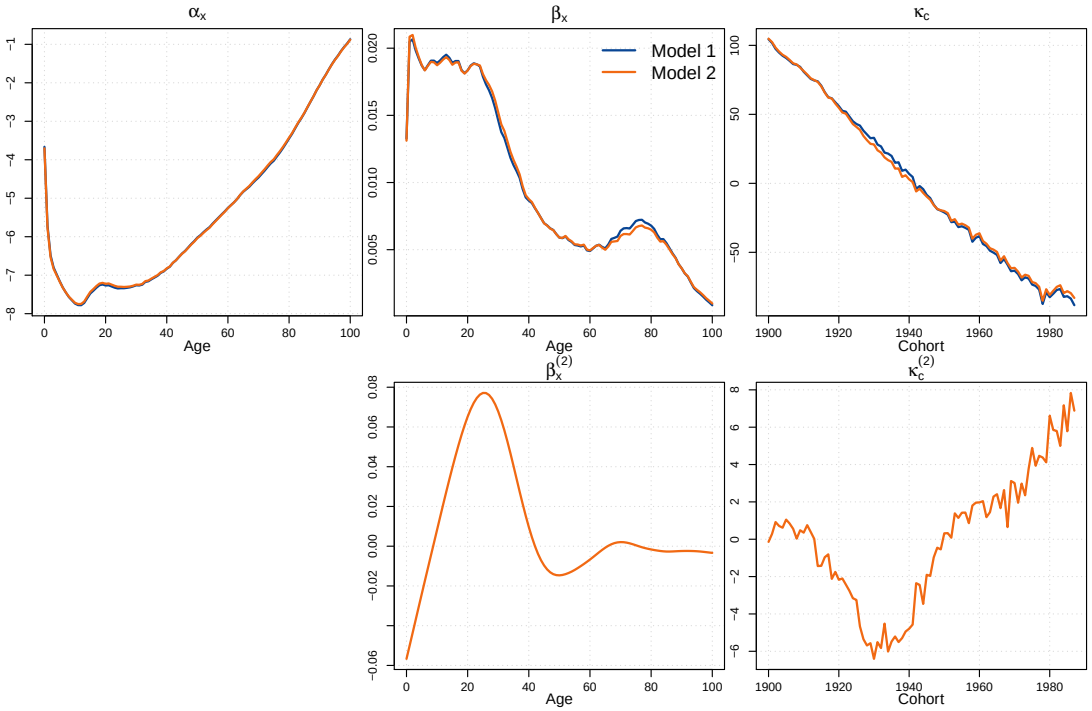

Figure 2. Estimated parameters of the M1 (blue lines) and M2 (orange lines) models. *Source*: authors' own elaborations using data of the Human Mortality Database (2019).

α_x, β_x and κ_c display the typical shapes generally found when fitting the LC model to mortality data; moreover, the parameter $\kappa_c^{(2)}$ captures the second main cohort mortality trend, modulated over ages by $\beta_x^{(2)}$.

Figure 3 shows the observed, fitted and forecast central death rates (in $\log_{10}$ scale) of the two models for Swedish females at selected ages for the cohorts 1900–1987. The Figure highlights that employing the second principal components (M2) provides a superior fit compared to using only one principal component (M1). BIC values, as shown in Table 2 further confirm that M2 outperforms M1: the increased model flexibility provided by the additional parameters $\beta_x^{(2)}$ and $\kappa_c^{(2)}$ (shown by an higher Effective Dimensions of the model) is statistically sound and justifiable.

Model	DEV	ED	BIC
M1	17715	**260**	19991
M2	**10950**	359	**14088**

Table 2. Poisson Deviance (DEV), Effective Dimensions (ED) and Bayesian Information Criterion (BIC) of the models M1 and M2 fitted on Swedish female mortality for ages 0–100 and cohorts 1900–1987. Lower values of the DEV, ED and BIC (in bold) correspond to better fit, more parsimony and better model, respectively.

Figure 3 further shows that, while the fitted rates of the two models are different, mortality forecasts are instead rather similar. Similar conclusions can be drawn from the analyses of death rates over age for selected cohorts, which are shown in Figure 4: the main difference between the two models appears in the fitting rather than forecast regions. The two figures further exemplify how the plain estimation of the two models directly allows to complete the mortality experience of partially observed cohorts.

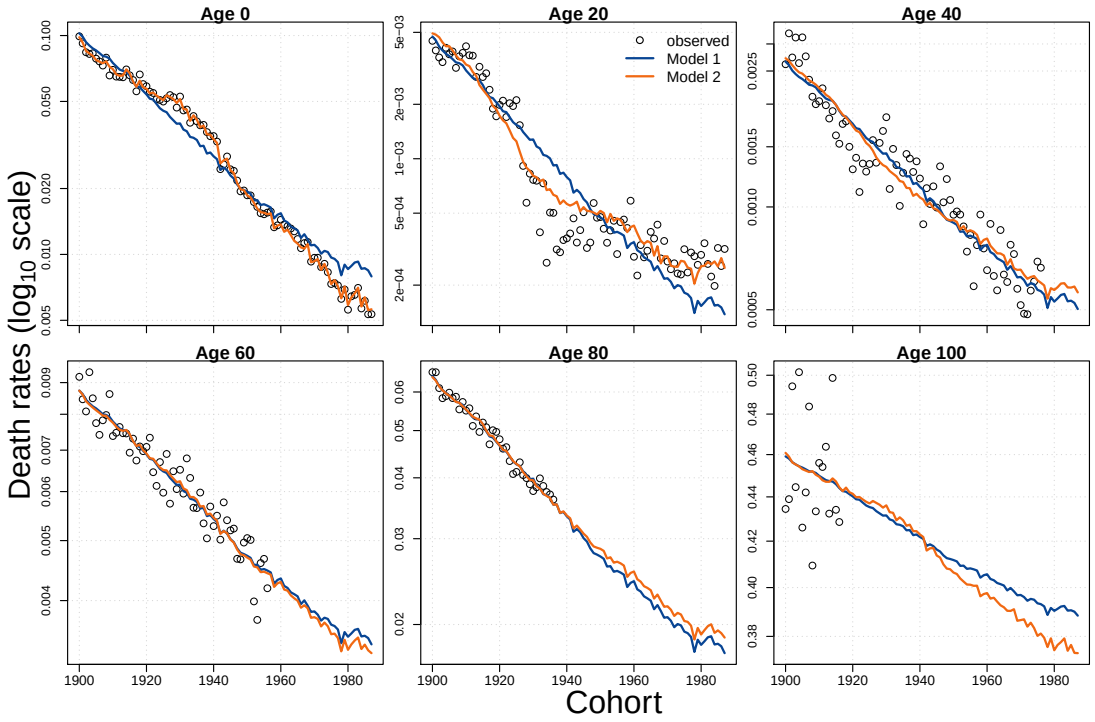

Figure 3. Observed, fitted and forecast central death rates (in $\log_{10}$ scale) for Swedish females at selected ages for the cohorts 1900–1987 for the models M1 (blue lines) and M2 (orange lines).
Source: authors' own elaborations using data of the Human Mortality Database (2019).

Finally, Figure 5 shows the deviance residuals of the two models. In these shaded contour maps, we can identify for which ages, cohort and years, a model under- and over-estimates observed mortality patterns. In theory we seek for a model that presents a large number of residuals close to zero and without recognizable patterns. On the one hand, both models present areas which misfit the data, e.g. the diagonal red bend corresponding to World War II years. On the other, these plots confirm that M2 clearly outperforms M1: colours on the right panel are generally lighter, especially for infant mortality.

Along with the residuals, Figure 5 presents observed, fitted and completed life expectancy at birth (e_o) from both models. The graph shows that both models successfully fit the observed e_o of the fully observed cohorts (1900–1916). Furthermore, the estimated age-cohort mortality surface allows us to compute e_o for all non-extinct cohorts.

4 Discussion

Mortality forecasting has recently been a subject of great interest among academics and financial sector practitioners due to the increasing challenges posed by sustained longevity improvements. The great majority of the advancements in the field have been made from the conventional age-period analysis of mortality data: the most innovative and successful techniques to forecast mortality are indeed based on different functions of period life tables (see, for example, Lee and Carter, 1992; Cairns et al., 2006; Raftery et al., 2013). Extensions of these models have been proposed to consider cohort effects (Renshaw and Haberman, 2006; Cairns et al., 2009; Plat, 2009); however, very few attempts have been made to forecast mortality using an age-cohort perspective.

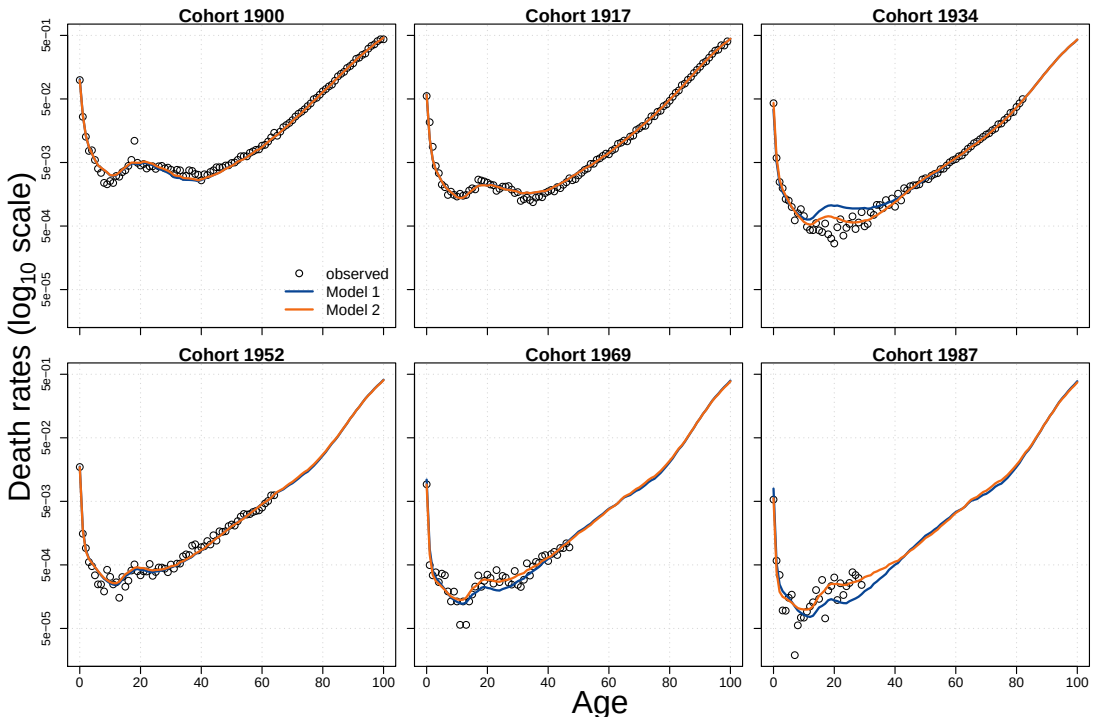

Figure 4. Observed, fitted and forecast central death rates (in $\log_{10}$ scale) for Swedish females at selected cohorts for the ages 0–100 for the models M1 (blue lines) and M2 (orange lines).
Source: authors' own elaborations using data of the Human Mortality Database (2019).

In this article, we propose a methodology to directly model and forecast cohort mortality. Our procedure is based on the most successful extensions of the prominent Lee–Carter (LC, Lee and Carter, 1992) model. Specifically, we employ and compare the original and the two-component LC models (firstly implemented by Renshaw and Haberman, 2003). Both models are embedded in the conventional Poisson framework for death counts (Brillinger, 1986), which allows to define and compute maximum likelihood estimators of the LC parameters. This further allows us to: (i) accommodate the presence of missing data that characterize cohort mortality data, and (ii) obtain smooth LC parameters, by including penalization terms in the Poisson log-likelihood function.

The plain estimation of these two models directly provides us with a complete age-cohort mortality surface. As such, the completion of the mortality developments of non-extinct cohorts is achieved without forecasting any parameter. In this extended abstract, we have shown an application of our methodology to Swedish female mortality at ages 0–100 and birth cohorts 1900–1987; the more flexible model with two principal components fits the observed mortality patterns better than the original LC model. Mortality forecasts of the two models are instead quite similar.

Future work is currently foreseen during the next months along different directions. The first step will be to compute the confidence intervals of the models' estimates. In particular, two different approaches could be employed for this purpose: first, it could be possible to numerically compute the Hessian matrix of the estimated LC parameters. This should be possible given the Poisson framework and maximum likelihood estimation procedure. The delta method could then be used to compute confidence intervals for the cohort mortality rates. If the approximation of the Hessian matrix turned out to exceed the computational power (due to the large number of parameters in the models), a (residual) bootstrapping

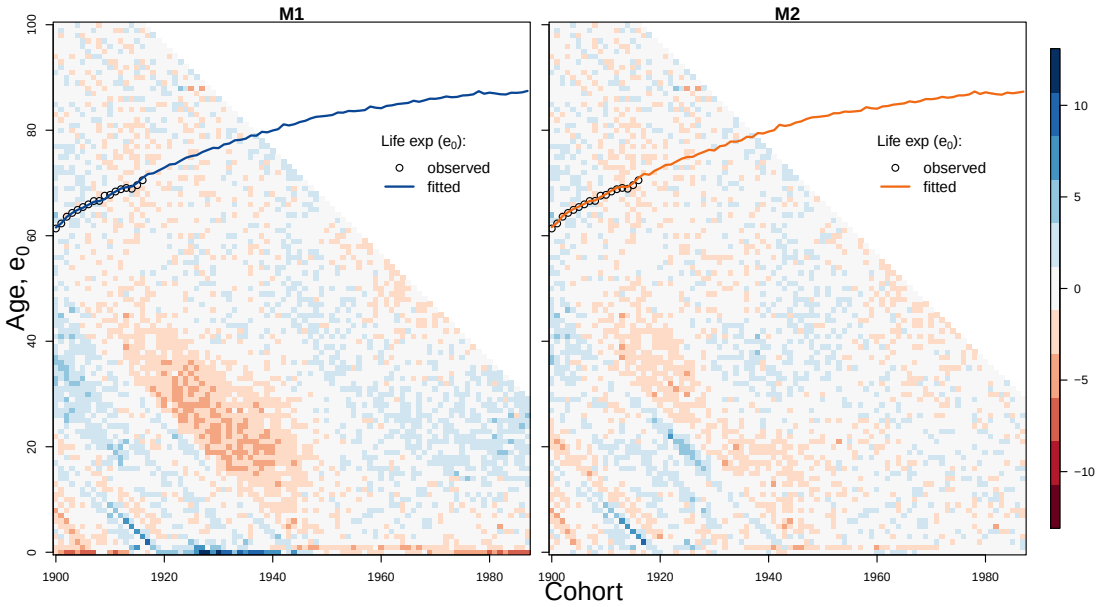

Figure 5. Poisson Deviance residuals, observed, fitted and completed life expectancy at birth (e_o) for the models M1 and M2 on Swedish females, ages 0–100 and cohorts 1900–1987. *Source*: authors' own elaborations using data of the Human Mortality Database (2019).

approach could be employed instead.

In addition to this aspect, we plan to compare the outcomes of our approach with those of other models. For example, the two-dimensional P-splines model (Currie et al., 2004), which has already been employed to forecast cohort mortality data (Continuous Mortality Investigation, 2007). It would be further interesting to compare our results with the cohort forecasts obtained by extracting the diagonals of the LC age-period projected surface. Finally, we will extend our analyses to several different populations of the Human Mortality Database (2019).

References

Alho, J. M. (2000). "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications", Ronald Lee, January 2000. *North American Actuarial Journal*, 4(1):91–93.

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., Kjærgaard, S., and Camarda, C. G. (2019). An age-at-death distribution approach to forecast cohort mortality. In *84th Annual Meeting of the Population Association of America. Austin, TX (USA).*

Bongaarts, J. and Feeney, G. (2002). How long do we live? *Population and Development Review*, 28(1):13–29.

Bongaarts, J. and Feeney, G. (2003). Estimating mean lifetime. *Proceedings of the National Academy of Sciences*, 100(23):13127–13133.

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

Booth, H. and Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(1-2):3–43.

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

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

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

Chen, C. and Liu, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88(421):284–297.

Chiou, J.-M. and 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(486):572–585.

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. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 4:356–383.

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

Delwarde, A., Denuit, M., and Eilers, P. (2007). Smoothing the Lee–Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach. *Statistical Modelling*, 7(1):29–48.

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 30 August 2019).

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

López-de Lacalle, J. (2019). *tsoutliers: Detection of Outliers in Time Series*. R package version 0.6-8.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.

Raftery, A. E., Chunn, J. L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50(3):777–801.

Renshaw, A. E. and Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.

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

Rizzi, S., Kjærgaard, S., Bergeron-Boucher, M.-P., Lindahl-Jacobsen, R., and Vaupel, J. (2019). Forecasting mortality of not extinct cohorts. In *21st Nordic Demographic Symposium, Reykjavik (Iceland)*.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

United Nations (2019). *World Population Prospects 2019: Highlights. ST/ESA/SER.A/423*. Department of Economic and Social Affairs, Population Division, New York, NY.

van Raalte, A. A., Nepomuceno, M. R., Camarda, C. G., and Myrskylä, M. (2018). How is lifespan variation developing across cohorts? In *Tel Aviv Workshop on Inequality and Uncertainty in Length of Life*.

Zanotto, L. and Mazzuco, S. (2017). Reconstruction of cohort data via EM algorithm. In *XXVIII International Population Conference*.

A Updating formulas

Here, we present the analytic solutions of the four models that we employ in this article. Similar relationships can be found in Brouhns et al. (2002), Renshaw and Haberman (2003) and Renshaw and Haberman (2006).

For model 1, the three parameters updating relationship are:

$$\begin{aligned}
\hat{\alpha}_x^{(\nu+1)} &= \hat{\alpha}_x^{(\nu)} - \frac{\sum_c w_{xc}\left(d_{xc} - \hat{d}_{xc}\right)}{\sum_c w_{xc}\,\hat{d}_{xc}} \\
\hat{\beta}_x^{(\nu+1)} &= \hat{\beta}_x^{(\nu)} - \frac{\sum_c w_{xc}\left(d_{xc} - \hat{d}_{xc}\right)\hat{\kappa}_c}{\sum_c w_{xc}\,\hat{d}_{xc}\,\hat{\kappa}_c^2} \\
\hat{\kappa}_t^{(\nu+1)} &= \hat{\kappa}_t^{(\nu)} - \frac{\sum_x w_{xc}\left(d_{xc} - \hat{d}_{xc}\right)\hat{\beta}_x}{\sum_x w_{xc}\,\hat{d}_{xc}\,\hat{\beta}_x^2}
\end{aligned} \tag{A.1}$$

where $\hat{d}_{xc} = e_{xc}\,\hat{\mu}(\boldsymbol{\theta})_{x,c}$ are the fitted deaths from the model.

For model 2, two additional relationships are required in addition to those in Equation (A.1) for updating the second principal component parameters:

$$\begin{aligned}
\left(\hat{\beta}_x^{(2)}\right)^{(\nu+1)} &= \left(\hat{\beta}_x^{(2)}\right)^{(\nu)} - \frac{\sum_c w_{xc}\left(d_{xc} - \hat{d}_{xc}\right)\hat{\kappa}_c^{(2)}}{\sum_c w_{xc}\,\hat{d}_{xc}\left(\hat{\kappa}_c^{(2)}\right)^2} \\
\left(\hat{\kappa}_c^{(2)}\right)^{(\nu+1)} &= \left(\hat{\kappa}_c^{(2)}\right)^{(\nu)} - \frac{\sum_x w_{xc}\left(d_{xc} - \hat{d}_{xc}\right)\hat{\beta}_x^{(2)}}{\sum_x w_{xc}\,\hat{d}_{xc}\left(\hat{\beta}_x^{(2)}\right)^2}
\end{aligned} \tag{A.2}$$

Finally, for the two augmented models M1 and M2 with period effects, an additional relationship is required in addition to those in Equation (A.1) and (A.2) for updating the period effect parameter:

$$\hat{\gamma}_t^{(\nu+1)} = \hat{\gamma}_t^{(\nu)} - \frac{\sum_t w_{xc} \left(d_{xc} - \hat{d}_{xc} \right)}{\sum_t w_{xc} \hat{d}_{xc}} \tag{A.3}$$

where the summation is over all $t = x + c$ considered in the data.