Lee–Carter cohort mortality forecasts

Ugofilippo Basellini*[1,2] and Carlo Giovanni Camarda[2]

[1] *Max Planck Institute for Demographic Research (MPIDR), Rostock*
[2] *Institut national d'études démographiques (INED), Aubervilliers*

Draft paper for the IPC 2021 conference

May 16, 2021

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 model's parameters, thereby completing the mortality experience of non-extinct cohorts without any extrapolation. 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 · Mortality modelling · Lee–Carter model · Smoothing · Newton–Raphson algorithm · Cohort life tables

A longer version of this four-page extended abstract is available upon request from the authors.

*Corresponding author: basellini@demogr.mpg.de
Address: Konrad-Zuse-Str. 1, 18057 Rostock, Germany

1 Introduction

Forecasting mortality is an essential component for the computation of future pension and health care expenditures as well as a crucial component in population projections. 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.

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. Typically, cohort forecasts are obtained by 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). Models for forecasting cohort mortality are relatively few in the literature, but some recent efforts have been made in this direction (Basellini et al., 2020; Chiou and Müller, 2009; Continuous Mortality Investigation, 2007; Rizzi et al., 2021; Zanotto and Mazzuco, 2017). In this article, we propose a new approach to forecast cohort mortality that generalizes the most recent and successful extensions of the LC model. Our contribution 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, 2021). As a future application, our methodology could be applied to complete the cohort life tables for all countries and for both sexes in the HMD.

2 Methods

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 $\hat{m}_{x,c} = y_{x,c}/e_{x,c}$. Following Brillinger (1986), we assume 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 $\boldsymbol{E} = (e_{x,c})$. 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 triangle:

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

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:

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

For both M1 and M2, 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 maximizing the Poisson log-likelihood:

$$\ln \mathcal{L}\left(\boldsymbol{\theta} | \, y_{x,c} \,, e_{x,c}\right) \propto \sum_{x,c} \left\{ y_{x,c} \, \eta(\boldsymbol{\theta})_{x,c} - e_{x,c} \, \mu(\boldsymbol{\theta})_{x,c} \right\} \,, \tag{7}$$

where the linear predictor $\eta(\boldsymbol{\theta})_{x,c}$ can follow either Eq. (3) or (5), and the parameters are subject to the model's specific constraints. The advantage of this assumption is that the error terms 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. (7). 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. Second, following the suggestion of Currie (2013), we include penalization terms for both $\boldsymbol{\alpha}$ 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), so that $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}]$. We can then rewrite the penalised log-likelihood as:

$$\begin{aligned} \ln \mathcal{L}^{\star}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa} | \, y_{x,c} \,, e_{x,c} \,, w_{x,c}\right) \propto \\ \sum_{x,c} w_{xc} \left\{ y_{x,c} \, \eta(\boldsymbol{\theta})_{x,c} - e_{x,c} \, \mu(\boldsymbol{\theta})_{x,c} \right\} - \frac{1}{2} \lambda_\alpha \, \boldsymbol{\alpha}' \, \boldsymbol{D}' \boldsymbol{D} \, \boldsymbol{\alpha} - \frac{1}{2} \lambda_\beta \, \boldsymbol{\beta}' \, \boldsymbol{D}' \boldsymbol{D} \, \boldsymbol{\beta} \,, \end{aligned} \tag{8}$$

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. For the model M2 , we include an additional penalization term for $\boldsymbol{\beta}^{(2)}$, while the parameters $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$ are always left unpenalised to better capture the cohort-specific fluctuations of mortality. The uni-dimensional iterative Newton-Raphson method is then adapted to maximise Eq. (8).

3 Results

Figure 1 shows the estimated parameters of the models M1 and M2 for Swedish females aged 0–100 years in the cohorts 1900–1987. The parameters 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 $\beta_x^{(2)}$.

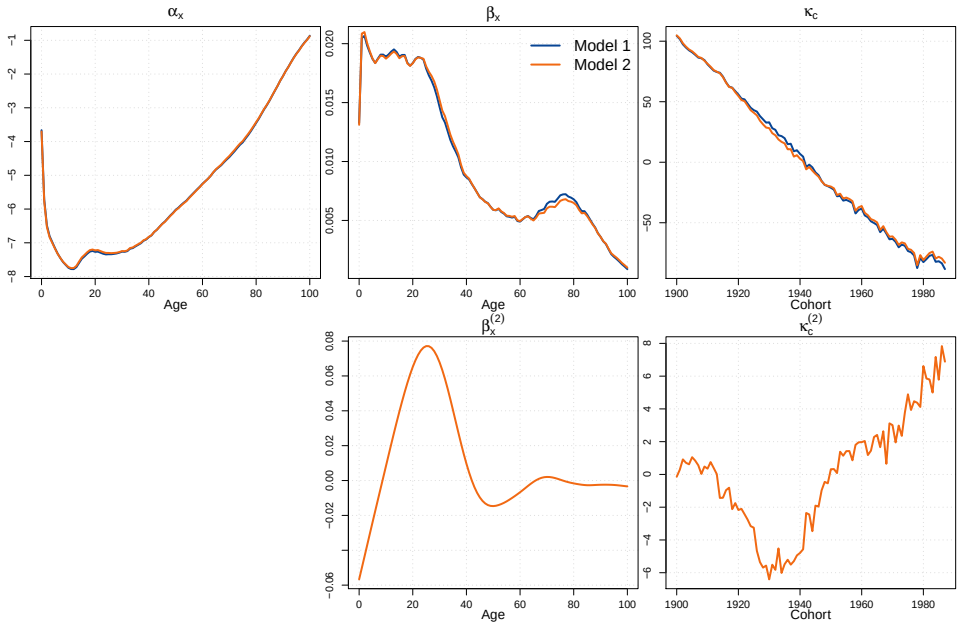

Figure 1. Estimated parameters of the M1 (blue lines) and M2 (orange lines) models.

Figure 2 shows the observed, fitted and forecast central death rates 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 (not shown here) further confirm this finding. The figure 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 (not shown here). Moreover, the figure exemplifies how the plain estimation of the two models directly allows to complete the mortality experience of partially observed cohorts.

Finally, Figure 3 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. 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 3 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.

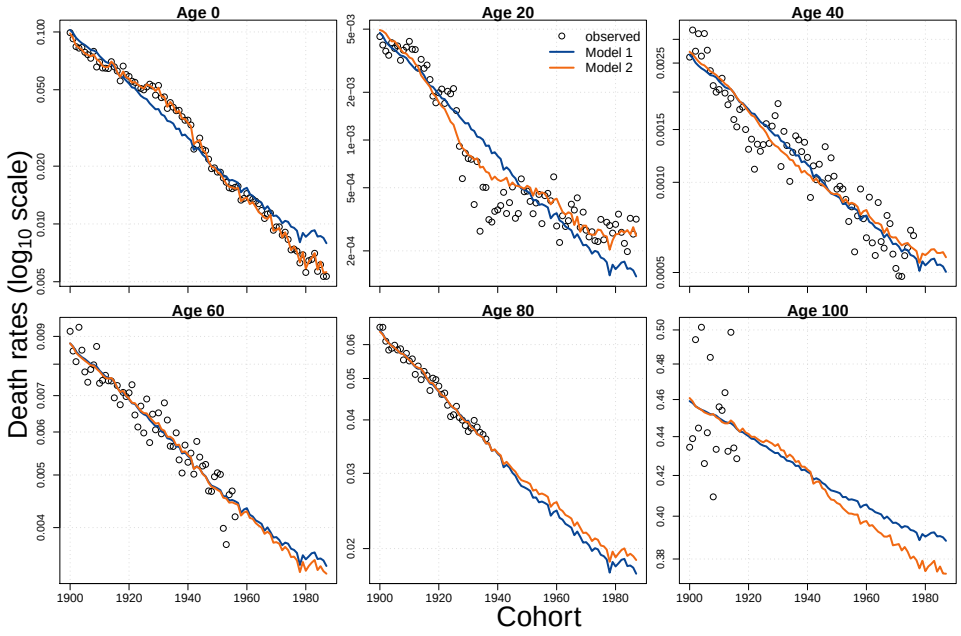

Figure 2. Observed, fitted and forecast central death rates (in $\log_{10}$ scale) for Swedish females at selected ages for the cohorts 1900–1987 for the models M1 (blue lines) and M2 (orange lines).
Source: authors' own elaborations using data of the Human Mortality Database (2021).

4 Discussion

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) model. Specifically, we employ and compare the original and the two-component LC models (firstly implemented by Renshaw and Haberman, 2003). Unlike most of the LC variants, 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 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 (2021).

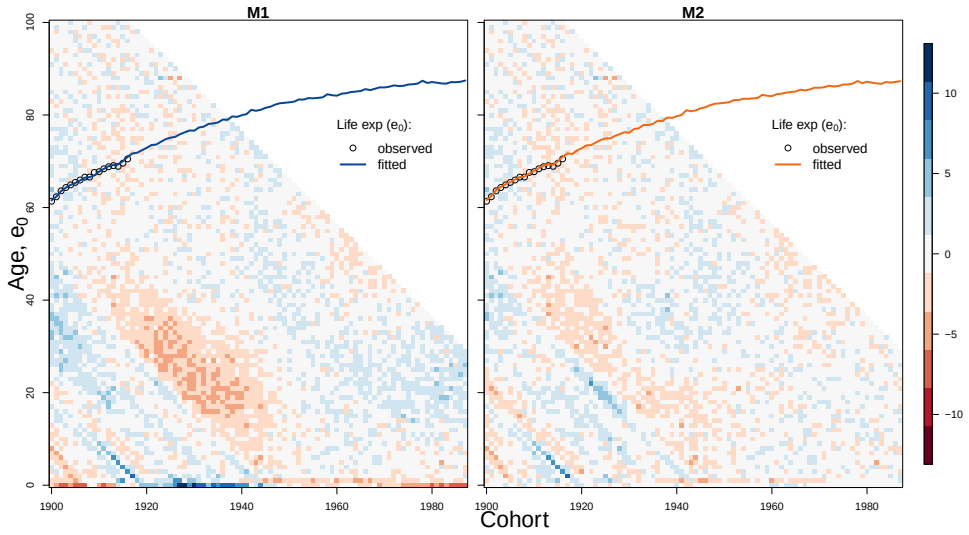

Figure 3. Poisson Deviance residuals, observed, fitted and completed life expectancy at birth (e_o) for the models M1 and M2 on Swedish females, ages 0–100 and cohorts 1900–1987.
Source: authors' own elaborations using data of the Human Mortality Database (2021).

References

Basellini, U., Kjærgaard, S., and Camarda, C. G. (2020). An age-at-death distribution approach to forecast cohort mortality. *Insurance: Mathematics and Economics*, 91:129–143.

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.

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. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, 13(1):69–93.

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

Human Mortality Database (2021). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

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

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

Rizzi, S., Kjærgaard, S., Boucher, M.-P. B., Camarda, C. G., Lindahl-Jacobsen, R., and Vaupel, J. W. (2021). Killing off cohorts: Forecasting mortality of non-extinct cohorts with the penalized composite link model. *International Journal of Forecasting*, 37(1):95–104.

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*.