Thirty years on: A review of the Lee–Carter method for forecasting mortality

Ugofilippo Basellini, Carlo Giovanni Camarda, Heather Booth

Review

Thirty years on: A review of the Lee–Carter method for forecasting mortality

Ugofilippo Basellini [a,b,*], Carlo Giovanni Camarda [b], Heather Booth [c]

[a] *Max Planck Institute for Demographic Research, Rostock, Germany*
[b] *French Institute for Demographic Studies, Aubervilliers, France*
[c] *School of Demography, Australian National University, Canberra, Australia*

ABSTRACT

The introduction of the Lee–Carter (LC) method marked a breakthrough in mortality forecasting, providing a simple yet powerful data-driven stochastic approach. The method has the merit of capturing the dynamics of mortality change by a single time index that is almost invariably linear. This thirtieth anniversary review of its 1992 publication examines the LC method and the large body of research that it has since spawned. We first describe the method and present a 30-year ex post evaluation of the original LC forecast for U.S. mortality. We then review the most prominent extensions of the LC method in relation to the limitations that they sought to address. With a focus on the efficacy of the various extensions, we review existing evaluations and comparisons. To conclude, we juxtapose the two main statistical approaches used, discuss further issues, and identify several potential avenues for future research.

1. Introduction

The story of human mortality and longevity is far from complete, eliciting both intrigue and challenge. How long will we live? How can we know? Efforts to anticipate longevity's trajectory seem to have increased with the trajectory itself. Forecasts of mortality were already being made at the beginning of the twentieth century, when English actuaries started to assess the financial effects of lengthening longevity on the reserves of pension providers (Pollard, 1987). Such impetus has not only grown considerably in more recent decades, but is also magnified by population ageing, leading to an expanding body of research into methods for forecasting the length of our lives.

Until the 1980s, the methods used to forecast mortality were deterministic, being based on mathematical formulae or expert judgment, with the latter also used in scenarios to provide variability (Booth & Tickle, 2008). It was only during the last three decades that mortality forecasting embraced stochastic methods. Much of the attention was first stimulated by the seminal contribution in 1992 of Ronald D. Lee and Lawrence R. Carter introducing a parsimonious model to describe age–time-specific death rates in the United States for use in time-series forecasting. Their article "Modeling and Forecasting U.S. Mortality" in the *Journal of the American Statistical Association* presented what would become the renowned and influential Lee–Carter (LC) method.

At a time characterized by rising concerns about the sustainability of pension schemes and meeting the demand for elder healthcare, the introduction of the LC method spurred new interest in mortality forecasting, not least because it brought the idea of the stochastic approach to wider attention and accessibility. Stochastic mortality forecasting was in its infancy and had previously focused on independent time-series models, either of the parameters of an existing "law of mortality" (e.g.

* Correspondence to: Konrad-Zuse-Str. 1, 18057 Rostock, Germany.
E-mail address: basellini@demogr.mpg.de (U. Basellini).

McNown & Rogers, 1989; Tabeau, 2001) or of selected death rates (Pollard, 1987). Thirty years on, the LC method has become the benchmark approach for forecasting mortality. The original article has – at the time of writing (August 2022) – been cited more than 3700 times.[1] Significantly, academics and private sector practitioners, as well as some national and international statistical offices, rely on the LC method and its extensions to forecast mortality for their populations of interest.

To celebrate the thirtieth anniversary of the LC method, this paper presents a comprehensive critical review focusing on both the merits and limitations of the method, and on the various extensions that have been proposed to improve performance. The paper is organized as follows. We start by describing the LC method in Section 2. In Section 3, we perform a 30-year evaluation of the original LC forecast for the U.S. population as published in 1992, comparing it with observed mortality over its first 30 years. We then, in Section 4, review the LC method in detail by considering each of its technical points and the main ways in which these have been modified by others. This is followed in Section 5 by a review of existing evaluations and comparisons of the forecast accuracy of these different LC extensions. In Section 6, we juxtapose the two main statistical approaches used, discuss further issues, and point to possible avenues for future research.

2. The Lee–Carter method

We describe the Lee–Carter (LC) method, as proposed by Lee and Carter in 1992. In what follows, we distinguish between the LC *model* and the LC *method*. The former refers to the functional form for age–time-specific mortality, while the latter comprises a series of steps to estimate the LC model and fit a time-series model to the time index, along with specific adjustment and estimation procedures.

Let $m_{x,t}$ be the central death rate at age x in year t. The LC model can be expressed as

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

where a_x is the general shape of mortality by age, and k_t is a time index of the general level of mortality. The parameter b_x describes the extent to which mortality at age x changes given the overall temporal change in the general level of mortality: greater values of b_x are associated with faster mortality change. The error term $\varepsilon_{x,t}$ reflects residual age-specific temporal influences not captured by the model, and is assumed to have zero mean and variance σ_ε^2. By modeling logged rates, the fitted and forecast rates (obtained by exponentiation) are constrained to be positive. The logarithmic transformation also serves to counter the heteroscedasticity in the observed rates.

The LC model belongs to the class of "relational" models (Brass, 1971) used in demography (Lee & Miller, 2001). It involves a standard (or reference) mortality age profile a_x that is fixed with respect to time. Mortality dynamics over age and time are described by the interaction of the b_x and k_t parameters which modify the standard profile

a_x. Changes over time in logged death rates are perfectly correlated across age. When k_t declines linearly, each age-specific death rate decreases exponentially at its own independent constant rate. Note that a linear decline in k_t corresponds to a decelerating increase in life expectancy at birth.

The LC model is underdetermined, meaning that a unique solution does not exist. Let A_x, B_x and K_t denote a solution of (1). Then, for any non-zero scalar (denoted c), A_x, $B_x c$, K_t/c, and $A_x - B_x c$, B_x, $K_t + c$ are also solutions. Thus, a_x is determined only up to an additive constant, b_x is determined only up to a multiplicative constant, and k_t is determined only up to a linear transformation. To address this, Lee and Carter introduced two constraints to ensure model identification:

$$\sum_x b_x = 1 \qquad \text{and} \qquad \sum_t k_t = 0. \tag{2}$$

Under these constraints, a_x is the temporal mean of $\ln(m_{x,t})$ (this follows from the constraint on k_t and the assumption of zero mean for $\varepsilon_{x,t}$), and b_x is the proportional pace of change.[2] In practically all applications, k_t declines over time, corresponding to mortality improvement. In this context, b_x is generally positive, reflecting age-specific improvement, but will be negative if mortality has increased at age x over the fitting period (in such cases, the interpretation of b_x should heed counterbalancing across age).

The first step of the LC method is the estimation of the LC parameters. The LC model can be viewed as a non-linear model with age and time treated as factor variables, with an assumed homoscedastic and normally distributed error term. The model cannot be fitted by regression methods, as there are no known covariates: only unknown parameters are present on the right-hand side of (1). Lee and Carter used singular value decomposition (SVD)[3] to derive an ordinary least squares (OLS) solution. After subtracting a_x from each column of $\ln(m_{x,t})$ to produce a "centered" matrix, the first term of the SVD is used to derive b_x and k_t. Specifically, b_x is obtained from the first left singular vector (adjusting for the b_x constraint), and k_t is the product of the leading singular value, the first right singular vector, and the sum of the first left singular vector.

The estimated LC model generally results in fitted numbers of deaths (obtained by applying estimated death rates to central exposures or mid-year populations) that differ from those observed. To address this discrepancy, Lee and Carter adjusted the parameter $\hat{k}_t$ in a second-step estimation procedure, while keeping fixed $\hat{a}_x$ and $\hat{b}_x$, such that the fitted number of total deaths matches observed deaths in each year t.

---

[1] Source: GoogleScholar.

[2] Alternative constraints (e.g. Wilmoth, 1993) would lead to different parameter estimates and interpretation, but unchanged fitted and forecast rates (Currie, 2020).

[3] As an historical note, principal component analysis (PCA) – which is closely related to SVD – had previously been applied in demography by Ledermann and Breas (1959) to analyze life table data from different countries, while Gómez de León (1990) and Bell and Monsell (1991) independently employed PCA to forecast age-specific death rates. Wilmoth (1990) also used SVD.

In the third step, the adjusted $\hat{k}_t$ are modeled and forecast with an appropriate ARIMA time-series model. In what follows, we refer to the modeled and forecast adjusted $\hat{k}_t$ as k_t in order to ease notation. In the context of U.S. mortality for the years 1933–1989, Lee and Carter found a random walk with drift to be the most appropriate model. This time-series model is

$$k_t = k_{t-1} + \delta + \beta D_f + e_t \qquad (3)$$

where δ denotes the drift, and e_t is a white noise process – or innovation – with mean zero and variance σ_e^2. Lee and Carter took into account the spike in mortality caused by the 1918 Spanish influenza pandemic, using the dummy variable D_f (with associated coefficient β). As the authors note, this inclusion only affects the variance estimate and prediction intervals (while the drift estimate and point forecast remain unchanged).

Fitted and forecast rates are derived from the fitted and forecast time-series values of k_t and the estimated age-varying parameters $\hat{a}_x$ and $\hat{b}_x$. At the oldest ages, where data are unreliable, the forecast rates are not used. Instead, in the fourth step, Lee and Carter employ the Coale and Guo (1989) procedure to derive death rates at ages 85+ from those at younger ages. Standard life table methods are then used to obtain other mortality functions (density, survival, life expectancy, etc.).

The variance of fitted and forecast mortality depends on the error in estimating the parameters a_x and b_x, the LC model error term $\varepsilon_{x,t}$, and the error in modeling k_t. The errors associated with estimating the LC model are relatively small and not considered by Lee and Carter. The error in modeling k_t as in (3) is composed of two independent terms: the innovation and the error in estimating the drift (the minor error in estimating β is ignored). Since the former far exceeds the latter, Lee and Carter focused only on the innovation. The authors dedicate an Appendix to the treatment of errors, demonstrating the extent to which error due to k_t increasingly dominates the other sources of error in forecast death rates and life expectancy as the forecast horizon increases. They conclude that prediction intervals for life expectancy at birth can be computed solely from $\hat{\sigma}_e^2$, while for age-specific death rates, 95% prediction intervals can be derived by multiplying the point forecast by a factor of $\exp\left(\pm 2\, b_x\, \hat{\sigma}_e\, \sqrt{h}\right)$, where h is the forecast horizon. As the width of the prediction interval for k_t increases by $\sqrt{h}$, the interval widens most rapidly in the early years of the forecast.

Several features have contributed to the success of the LC method and its wide application:

- the method is simple and powerful: age-specific mortality forecasts are derived by modeling a single time index by standard time-series methods;
- in many applications, a linear time index adequately captures the historical decline in mortality;
- the approach is stochastic, permitting the derivation of probabilistic prediction intervals for all mortality quantities; and
- the approach is purely extrapolative: forecasts are based on observed data without subjective expert judgment or external information. The only decision required of the user is the choice of fitting period.

3. The original LC forecast: A 30-year evaluation

We perform an ex post evaluation of the forecast performance of the LC method by comparing the mortality forecast made by Lee and Carter for the U.S. population with observed mortality. Data for both sexes combined were analyzed by five-year age groups [0, 1–4, 5–9, . . . , 105–109]. The fitting period was 1933–1989, and the forecast period was 1990–2065. For this evaluation we consider the 30-year time period from 1990 to 2019. Note that this time period does not include the effects of the Covid-19 pandemic, which began in 2020. Our evaluation extends the earlier evaluation of Lee and Miller (2001) by 20 years.

In order to derive mortality forecasts as close as possible to those of Lee and Carter, we retrieve the estimated $\hat{a}_x$ and $\hat{b}_x$ from Table 1 of their paper, and the forecast values of k_t for the period of interest from Table 2. These three parameters allow us to construct the forecast mortality surface for the years 1990–2019. Following Lee and Carter, we derive forecast death rates at ages 85+ from those at younger ages by using the procedure described in Coale and Guo (1989, pp. 614–615). From Table 2 we also obtain the estimated standard deviation of forecast k_t, from which we construct the prediction intervals for life expectancy at birth (e_0) and age-specific death rates following the procedures described by Lee and Carter (1992) (see Section 2). Observed mortality data for the period of interest were retrieved from the Human Mortality Database (2022). These analyses are fully reproducible from an open-access repository.[4]

Fig. 1 shows the results of this evaluation exercise. The left panel reports the observed e_0 from 1985 to 2019, along with the first 30 years of the LC forecast with 95% prediction intervals. The right panel compares the observed and forecast death rates (on a logarithmic scale) in 2019.

The left panel of Fig. 1 reveals that the LC forecast was optimistic: forecast e_0 is greater than the observed values for all years. However, the fact that the LC forecast of e_0 is already upwardly biased by approximately half a year in the first year of the forecast (as already noted by Bell (1997), Lee and Miller (2001); see also 'Jump-off error') indicates that this is the main source of overestimation, as the forecast slope is generally in keeping with the observation. Only in the last five years of the evaluation does the forecast error increase substantially, to about 1.4 years in 2019, owing to the recent stagnation in observed U.S. mortality. The 95% prediction interval amply contains all of the observed data points. We note that this (single) evaluation does not accord with the finding that the LC method underpredicts life expectancy (Bergeron-Boucher et al., 2017; Booth et al., 2006; Lee & Miller, 2001).

The right panel of Fig. 1 shows that the optimism in e_0 stems from death rates at ages below 60 years, which were forecast to decline at a much faster rate than actually occurred. Decomposition results (available from

---

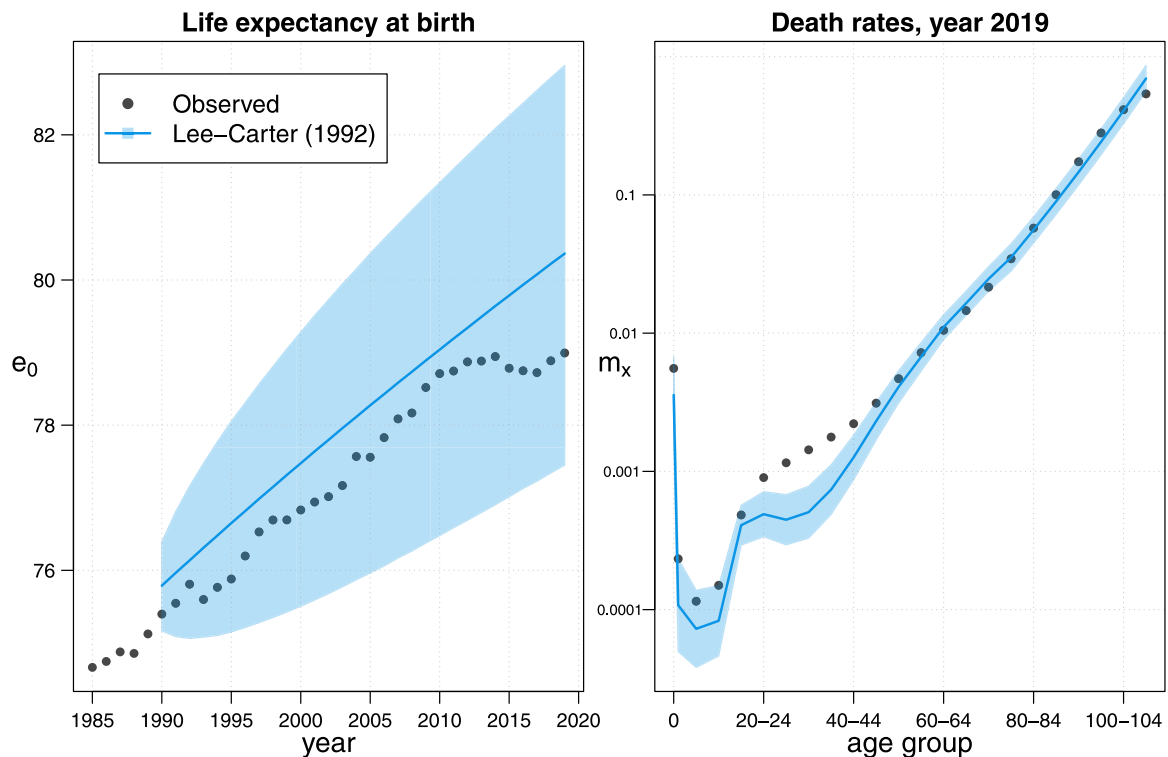4 Available at: https://github.com/ubasellini/LC-30anniversary

Fig. 1. Comparison of the original LC forecast with observed data: life expectancy at birth in 1985–2019 (left panel) and age-specific death rates in 2019 (log scale, right panel) with 95% prediction intervals, U.S. population (both sexes). *Source*: Own calculations using estimates from Lee and Carter (1992); observed data from the Human Mortality Database (2022).

the open-source repository) reveal that the ages 0 and 25–44 years contributed the most to the difference in 2019 life expectancy. In terms of the interval forecast, only 16 age-specific death rates (out of 23) are contained within the 95% prediction interval.

4. Limitations and extensions

Despite its several strengths and advantages, the LC method is not without limitations. Several efforts have been made in the subsequent literature to address these limitations by extending the LC method to include useful modified or additional features. Some of these extensions are included in previous broader mortality forecasting reviews (see, e.g., Booth & Tickle, 2008; Janssen, 2018; Stoeldraijer et al., 2013). In this section, we provide a dedicated review of the most prominent and widely employed LC extensions. We organize our review by the features and limitations of the method. We see this as a more useful approach for the reader than to describe, perhaps chronologically, a series of extensions with the risk of omitting explicit reference to their purpose. As many extensions address more than one limitation, we make use of cross-referencing to link their different features. We endeavor to cover the most prominent, if not all, extensions.

4.1. Adjusting $\hat{k}_t$

One of the most celebrated advantages of the LC model of mortality is its ability to capture almost all of the

variation by the single dominant term of the SVD. By using this rank-1 solution, all other variation, which is captured by the remaining terms, is relegated to the residual $\varepsilon_{x,t}$ (note that the full SVD would reproduce rates exactly). Thus, the fitted model does not match observed rates and, consequently, observed deaths. This limitation is augmented by the fact that in modeling logged death rates, greater weight is given to low rates and lesser weight to high rates than would be the case if rates were modeled (noting that the SVD solution is unweighted). Thus, larger absolute errors occur where rates are highest, and as higher rates prevail at older ages where the population is relatively large (as a result of population ageing), the error in terms of deaths is magnified (see also 'The statistical framework').

To address errors in fitted deaths, Lee and Carter incorporated a second-step adjustment of $\hat{k}_t$ to reproduce, for each year of the fitting period, the observed total number of deaths (see Section 2). By adjusting $\hat{k}_t$ to achieve observed deaths, greater weight is given to ages at which numbers of deaths are large, thereby counterbalancing (to some degree) the effect of the log transform. However, the weighting is complex and essentially unknown, as it does not conform to standard statistical criteria (Booth et al., 2002, p. 327).

Three alternative second-step adjustments of $\hat{k}_t$ have been proposed:

- Lee and Miller (2001) adjusted $\hat{k}_t$ to match observed life expectancy at birth in each year. This adjustment has the unique advantage of eliminating the

need for population data (as deaths are not employed), thereby expanding the applicability of the method to situations where only death rates are available. However, the underlying weighting procedure is even less intelligible than in the LC method.

- Booth et al. (2002) adjusted $\hat{k}_t$ to match observed age-specific deaths by using Poisson regression and minimization of the deviance. This conditional maximum likelihood procedure with known weights gives greater weight to ages at which more deaths occur. The authors argue that, compared with the LC adjustment, this procedure is likely to reduce the error in the last year of the fitting period (jump-off year), thereby improving forecast accuracy (see 'Forecasting evaluations and comparisons').
- Rabbi and Mazzuco (2021) adjusted $\hat{k}_t$ to match the variability of the age-at-death distribution rather than the mean (i.e. life expectancy at birth). The variability is computed using the lifespan disparity measure (Goldman & Lord, 1986; Vaupel, 1986; Vaupel & Canudas-Romo, 2003). The authors argue that this adjustment makes the time index more reflective of mortality improvements.

The size of the adjustment of $\hat{k}_t$ needed to reproduce deaths or life expectancy depends on the fit of the LC model. Subsequent evaluations found that the first two of these alternative adjustments have only a negligible effect on forecast accuracy (Booth et al., 2005). It should be noted that the second-step adjustment of $\hat{k}_t$ is no longer necessary when a more appropriate statistical framework is used (see 'The statistical framework').

4.2. Jump-off error

Lee and Carter (1992) noted that their model would not fit the observed death rates exactly in the jump-off year, and that this jump-off error could be removed by setting a_x equal to observed logged rates (and $k_t = 0$) in the jump-off year. However, the authors chose not to do this because it would reduce the overall goodness-of-fit and would have little effect on life expectancy because the largest relative errors occur when rates are lowest. With the benefit of hindsight, Lee and Miller (2001) noted that using fitted jump-off rates in the original LC forecast had in fact produced a 0.6-year jump-off error in U.S. life expectancy at birth, causing significant bias in the forecast (as also noted by Bell, 1997). These authors thus use observed jump-off rates in their extension of the LC method, removing the bias.

The magnitude of the jump-off error depends solely on the goodness-of-fit of the LC model with the adjustment of $\hat{k}_t$. This depends on how well the LC model assumptions (fixed b_x and homoscedastic errors with mean zero) are met, which in turn depends partly on the choice of fitting period. Booth et al. (2002) chose to use fitted jump-off rates after achieving a substantial reduction in the jump-off error for Australian data through a shorter fitting period and an improved goodness-of-fit of the LC model.

In order to avoid reproducing in the forecast any idiosyncrasies in the observed data, jump-off rates may be smoothed either independently or as part of the estimation procedure (see, e.g., Basellini & Camarda, 2022; Bergeron-Boucher et al., 2017). Stoeldraijer et al. (2018) analyzed the accuracy and robustness of mortality forecasts using the LC model and six different types of jump-off rates. Focusing on males and females in eight European countries from 1960–2014, the authors found a trade-off between accuracy and robustness: greater accuracy was achieved by using the last observed year as jump-off rates, while greater robustness was obtained by using an average of recent observed years as jump-off rates.

4.3. The statistical framework

Perhaps the most significant limitation of the LC method is its statistical framework. The LC model is a model of death rates and lacks an explicit underlying stochastic process. As the SVD estimation procedure is equivalent to obtaining an OLS solution, which is unweighted and where the error term is assumed homoscedastic and normally distributed, the Gaussian statistical framework is implicit. Several authors have noted that these Gaussian assumptions are unrealistic for human mortality, as death rates are more variable at older adult ages (because of higher rates and decreasing population denominators), and that greater weight should ideally be assigned to ages where numbers of deaths are largest.

Wilmoth (1993) addressed these limitations in two alternative ways, which produced very similar parameter estimates for Japanese females from 1951 to 1990. First, he proposed an SVD estimation procedure based on weighted least squares with weights equal to the number of deaths, thus giving greater weight to ages at which most deaths occur and also avoiding the problem of zero death rates (the log of zero being minus infinity) by assigning zero weight. In the second alternative, he adopted the assumption of Poisson deaths, an approach also adopted by Brouhns et al. (2002) and later by others. Brillinger (1986) had shown that, in a variety of cases, death counts can be approximated by a Poisson distribution. By doing so, the statistical framework is effectively switched from Gaussian to Poisson. In this approach, the LC model is embedded within a Poisson regression for death counts with a heteroscedastic error structure. Thus observed deaths are modeled directly, with population (which is taken as given) as an offset. As the error term applies directly to deaths, and the total number of fitted deaths equals total observed deaths by construction, the second-step adjustment of $\hat{k}_t$ is not needed. In this "Poisson-LC" approach, parameter estimation is derived by maximum likelihood estimation (MLE) for which Brouhns et al. (2002) provided Newton–Raphson updating formulae.[5] The a_x schedule is no longer the temporal mean, though the normalization remains. Moreover, the three parameter schedules are estimated

---

5 The Newton–Raphson method is a root-finding algorithm which produces successively better approximations to the roots (or zeroes) of a real-valued function.

simultaneously, rather than in the two separate steps (first a_x, and then b_x and k_t via SVD).

Several authors later employed the Poisson assumption within a generalized linear model (GLM) framework in their LC extensions (e.g. Camarda & Basellini, 2021; Delwarde et al., 2007a; Renshaw & Haberman, 2003b). Moreover, Currie (2013) extended the GLM framework to include smoothing and constraints within the estimation procedure, and showed that the LC model can be linearized with respect to its parameters. Consequently the Poisson-LC can be estimated by fitting three different GLMs (each keeping two parameters fixed and estimating the third parameter) successively until reaching convergence. The author further simplified this approach, showing how fitting only two different GLMs (one fixing b_x, and the other estimating a_x and k_t, and vice versa) can achieve the same result.

While being more realistic than the Gaussian assumption in modeling human mortality, the Poisson distribution may not be optimal, as death counts are generally overdispersed, especially at older ages. In other words, the variance may exceed the mean, violating the Poisson property of equal mean and variance. Overdispersion can arise from heterogeneity in the population: some individuals are more frail (i.e. have a greater mortality risk) than others because of unobserved differences (Vaupel et al., 1979).

However, the Poisson distribution can be generalized to account for overdispersion. Renshaw and Haberman (2003a) proposed an overdispersed Poisson quasi-likelihood estimation procedure. Moreover, these authors suggested modeling the parameter k_t with time as a known covariate, corresponding to assuming a linear function for the time index, with the additional possibility of capturing a structural break by a "hinge". Delwarde et al. (2007b) employed the negative binomial distribution to accommodate the overdispersion in death-count data. Using data for three countries by sex, the authors find that their approach produces a better fit compared with the Poisson-LC, but that life expectancy forecasts remained largely unchanged. While Delwarde et al. (2007b) employ an overall overdispersion parameter, Li et al. (2009) extended the approach to age-specific overdispersion. These authors account for heterogeneity by using a gamma distribution for frailty in a Poisson setting, which generates a negative binomial distribution with age-specific overdispersion parameters, and using data for the U.S. and Canada, they showed that the width of the prediction intervals for rates increased by 18% to 180% compared with the Poisson-LC intervals.

Hyndman and Ullah (2007) addressed the problem of heteroscedasticity by smoothing the death rates (see 'Smoothing'). These authors assumed an approximate binomial distribution for deaths, but in later work, Hyndman and Booth (2008) adopted the Poisson distribution. The random error in observed deaths is removed before model estimation and added back after estimation as one of the components of total error. Without such random error, the Hyndman–Ullah approach employs the Gaussian framework, which is also supported by the inclusion of several terms of the SVD (Hyndman & Ullah, 2007,

pp. 4947–4948) (see 'A single k_t'). Rabbi and Mazzuco (2021) also employ the SVD approach with smoothed rates (see 'Smoothing').

Regardless of the assumed stochastic process, observed death rates have the limitation of being unbounded at the oldest age. Three alternative age-specific measures have been employed to address this limitation, while retaining the LC parameterization. First, Currie (2013) modeled the probability of death $q_{x,t}$, with the advantage that its value is 1 at the oldest age. This is achieved within the GLM framework by using a logit link function, a binomial error structure for death counts, and population data referring to the beginning of the year. The goodness-of-fit of this "binomial-LC" was a substantial improvement over the Poisson-LC for male mortality at ages 40–90 years in England and Wales from 1961 to 2009. Second, Bergeron-Boucher et al. (2017) employed the life-table age-at-death distribution (the probability density function) $d_{x,t}$, which is naturally constrained to sum to one. Motivated by Oeppen (2008), the authors employ compositional data analysis (CoDa) techniques to model the death distribution while maintaining this constraint, with forecasts derived using SVD and ARIMA time-series models, as in the LC method. Third, Bergeron-Boucher et al. (2019) explored the survival function $\ell_{x,t}$. Based on data for 18 countries by sex, a comparison by Bergeron-Boucher et al. (2019) of forecasts made using the LC model with different measures shows that modeling $m_{x,t}$ and $q_{x,t}$ produces more pessimistic forecasts than models based on $d_{x,t}$ and $\ell_{x,t}$. The authors state that this pessimism "seems to be rooted in the indicator and/or transformation used in these models" (Bergeron-Boucher et al., 2019, p. 1249). Renshaw and Haberman (2003c) also departed from modeling rates by employing mortality improvement factors (the ratio of rates in years $t+1$ and t). These are modeled without the log-transform, retaining the LC parameterization and SVD (or an approximation). This method is illustrated using data for two male populations in the U.K.

A limitation of the LC method is the sequential estimation of the LC model and time-series model. Czado et al. (2005) noted that this procedure may lead to "incoherence" (p. 262). The authors thus embed LC in a Bayesian Poisson setting, where parameter estimation and forecasting are integrated, and the Bayesian framework allows the derivation of credible intervals for fitted and projected parameters and death rates. Parameter estimates from this Bayesian approach are very close to those from the Poisson-LC for Belgian mortality data from 1950 to 2000, but the Bayesian forecast rates are higher (more pessimistic) than those obtained from the LC method or Poisson-LC. de Jong and Tickle (2006) integrate the estimation of the age-specific parameters a_x and b_x, and the random walk with drift model for k_t, as well as smoothing over both age and time (see 'Smoothing'). This "LC(smooth)" extension is first illustrated using OLS, while ML estimates are obtained using three Kalman filters[6] with a smoothing filter for k_t (see 'Smoothing'). This

---

[6] The Kalman filter is an optimal estimation algorithm that uses uncertain observed data to estimate unobserved variables (Harvey, 1990).

state-space framework can incorporate additional features, such as other time-series models and heteroscedasticity. Data for Australian females are used to demonstrate the approach.

The LC model has been extended to a three-dimensional structure and estimated using tensor decomposition (an extension of two-dimensional SVD). Russolillo et al. (2011) used a Tucker decomposition (a type of tensor decomposition) of an array structured by age, time, and country. Analyzing data for ten European countries, the authors obtained country-specific estimates of the b_x and k_t parameters, as well as a synthesized overall mortality index. Dong et al. (2020) extended this approach by employing several tensor decompositions of the three-dimensional data (age, time, and either country or sex). Four methods were used: canonical polyadic decomposition, and three more flexible Tucker decompositions.

Machine learning techniques have been implemented within the LC framework. Deprez et al. (2017) applied a regression tree boosting machine approach to improve the goodness-of-fit of the LC model and of the cohort-based extension of Renshaw and Haberman (2006). Levantesi and Pizzorusso (2019) extended this work by considering three alternative tree-based machine learning techniques. Moreover, these authors introduce an LC model enhanced by machine learning, whereby an additional set of LC parameters, derived from machine learning, is included in the LC model. Nigri et al. (2019) introduced a recurrent neural network (RNN) approach to model and forecast k_t; and Marino et al. (2022) extended this approach to derive prediction intervals. Richman and Wüthrich (2021) employed neural networks to estimate the parameters of the multipopulation coherent (Li & Lee, 2005) method (see 'Coherent mortality forecasts').

A final note concerns the interpretation of the LC parameters. While b_x and k_t have a demographically meaningful interpretation for a given population and time period (under the LC normalization), they cannot be used for comparison between different populations, because they have only a *relative* interpretation. To overcome this, de Jong et al. (2020) proposed a novel parameterization of the LC model, where the b_x and k_t parameters have an absolute interpretation and are – with appropriate normalization – comparable across populations. The model is expressed in terms of the "needed exposure", and the parameters have a meaningful interpretation related to this new mortality metric. The approach is illustrated using data for six countries by sex. As an alternative, tensor decomposition (e.g. Dong et al., 2020; Russolillo et al., 2011) avoids the limitation of the lack of parameter comparability across different populations.

4.4. Fixed b_x

A central assumption of the LC model is that the b_x schedule is fixed over time. Lee and Miller (2001) noted that "a number of observers have suggested that the b_x coefficients might vary over time" (p. 543). Specifically, while in the first part of the twentieth century, mortality improvements were rapid at young ages and slow at old ages, the second part of the century witnessed a reversal

of this pattern (Kannisto et al., 1994; Rau et al., 2008). This poses a challenge for the LC model, which cannot capture changing age patterns of mortality decline over time.

Clearly, the suitability of the fixed b_x assumption is closely related to the length of the fitting period. Following Tuljapurkar et al. (2000), Lee and Miller adopted the "simple and practical solution" of restricting the fitting period to start from 1950, so that the assumption of fixed b_x is better met. Booth et al. (2002) restricted the fitting period in order to improve total fit (see 'Linearity of $\hat{k}_t$ and choice of fitting period'), also better meeting the assumption of fixed b_x.

The limitation posed by fixed b_x is more significant in the context of long-run forecasting. To allow flexibility, Li et al. (2013) introduce a "rotation" of the b_x schedule to capture the slowing mortality improvement at younger ages and accelerating improvement at older ages. Specifically, a time-varying $b_{x,t}$ parameter is employed, which changes gradually from the initial SVD estimate of b_x to an assumed "ultimate" pattern at very high levels of life expectancy at birth. As $b_{x,t}$ changes, the corresponding k_t are re-estimated to maintain the life expectancy forecast of the LC method. A detailed illustration is provided using data series from 1950 for Japan and the U.S. This method has been adopted by the United Nations in their probabilistic population projections to derive age-specific death rates from Bayesian hierarchical forecasts of life expectancy (United Nations, 2022; Ševčíková et al., 2016). Based on sex-specific data for 28 countries of the European Union from 1950 to 2015, Vékás (2020) found that rotation is more prevalent for female mortality than for male mortality.

While b_x is generally positive at all ages, for short fitting periods, negative values can occur, reflecting particular historical experience at certain ages. Such values are counterbalancing in the normalization of b_x, leading to problems of interpretation. Death rates associated with negative b_x increase over time, and, as their perpetuation in the forecast is unrealistic, negative b_x values serve as a warning; in this case, a simple and practical solution is to increase the fitting period.

In their three-component smooth LC extension, Camarda and Basellini (2021) address the problem of simultaneously opposing changes at different ages (positive and negative b_x) by smoothing and decomposing the mortality age pattern into three independent additive components (childhood, early-adult, and senescent mortality), each described by the Poisson-LC with smoothing (see 'Smoothing'). Though each component-specific b_x is fixed, their combined effect allows for flexibility in mortality dynamics.

4.5. Linearity of $\hat{k}_t$ and choice of fitting period

Lee and Carter found the random walk with drift (RWD) time-series model to be most appropriate to describe and forecast the adjusted $\hat{k}_t$ for U.S. data. This model is linear and passes through the first and last points of the series. The applicability of the RWD was confirmed in the "universal pattern" of Tuljapurkar et al. (2000). This experience has been repeated many times and the

linear model is often assumed to be an integral part of the method, though other ARIMA models could be more appropriate (de Jong & Tickle, 2006; Lee & Carter, 1992). Indeed, use of the RWD regardless of the actual trend may lead to bias in the forecast.

The focus of the LC method is long-term forecasting, based on a long series of data. The longer the series, the less influential are (given) systematic departures from linearity and the greater the prospect of the RWD model being used. However, identifying the RWD as the best model in the statistical sense does not imply that observed departures from linearity are of no consequence. Rather, systematic departures are likely to result in the drift term being out of line with the recent observed trend in k_t, effectively introducing structural change in the first year of the forecast. Further, such departures inflate the variance of k_t, producing a wide prediction interval that can be expected to encompass future departures from linearity of similar magnitude to those in the past series. The choice of fitting period (as opposed to data availability) is thus an important criterion in producing the forecast (see also 'Fixed b_x').

Lee and Carter examined how the choice of the fitting period, given their 1989 jump-off year, affected their forecast. Their interpretation of the results (Table 3, Lee & Carter, 1992) focuses on the length of the fitting period, arguably leading to misinterpretation in terms of model stability (pp. 664–665), rather than on the influence of the first year of the fitting period (or starting year). Under the RWD model, for a given series with a fixed jump-off year, the starting year is pivotal in that, by truncating the series, it fully[7] determines the trend in k_t and its variance. These are the two defining features of the forecast. The influence of the starting year is in its position in the historical series (and, to a minor extent, its temporal random variation).

The possibility of purposively (and objectively) choosing the fitting period to achieve a better forecast was explored by Booth et al. (2002). Accepting the assumption of a linear trend, and noting that the starting year determines the drift (for a given jump-off year), these authors devised a statistical procedure for limiting the fitting period by bringing forward the starting year so as to optimize the total fit of the method (i.e. the LC model and the RWD model of $\hat{k}_t$ after implementing their second-step adjustment; see 'Adjusting $\hat{k}_t$'). In forecasting the non-linear Australian mortality decline, the authors showed that this procedure identified a relatively short fitting period as optimal, compared with the maximum period allowed by the data.

Others have also reduced the fitting period. It would appear that Tuljapurkar et al. (2000) chose 1950 as the starting year because of the limited availability of data for Japan. Lee and Miller (2001) then adopted 1950 in order to better meet the assumption of fixed b_x (see 'Fixed b_x'). While (Hyndman & Ullah, 2007) used data from 1899 to 2001 (and robust estimation), (Hyndman & Booth, 2008)

adopted a starting year of 1950 in order to "avoid difficulties with war years, the 1918 Spanish influenza pandemic, and structural change over the course of the twentieth century" (p. 329). Many authors omit any justification for the fitting period, despite its influence on an RWD forecast.

4.6. A single k_t

The LC model requires only the first term of the SVD, while the second and higher terms, which typically account for only a small percentage of total variability, together constitute the residual. The forecast is thus based on the first time index k_t, the linearity of which produces inflexibility. Given orthogonal SVD terms and the characteristic linearity of the first time index, higher terms will involve non-linear time indices; this information could potentially be used to improve forecast accuracy by adding a degree of flexibility, but is ignored. Moreover, employing a single time index results in mortality improvement factors by age being perfectly correlated (Cairns et al., 2008), which is unrealistic and undesirable. Several methods have been introduced to address this limitation.

Booth et al. (2002) explored the second and third terms of the SVD in order to gain a better understanding of age–time interactions, but did not incorporate them in forecasting. Motivated by the age-inflexibility of the LC model, Renshaw and Haberman (2003b) employed a second term to improve goodness-of-fit, and included it in forecasting. These authors demonstrated this approach in both the Gaussian (using the SVD) and Poisson (MLE) frameworks, with the estimated first and second time indices being forecast using appropriate ARIMA processes. In the Poisson framework, statistical criteria (such as the Bayesian information criterion) are employed to determine the optimal number of terms to be included. However, it should be borne in mind that a better goodness-of-fit does not necessarily produce greater forecast accuracy (Booth & Tickle, 2008). In a generalized approach employing functional principal components analysis of smoothed death rates and ARIMA time-series models, Hyndman and Ullah (2007) include a procedure for determining the number of terms to employ, based on minimizing the out-of-sample forecast error in age-specific rates. For French mortality (single years of age from 1899 to 2001), four terms were required. Hyndman and Booth (2008) note that six terms is the maximum likely to be needed.

Rather than incorporate the higher terms of the SVD, additional information can be explicitly modeled. In a novel approach by Renshaw and Haberman (2006) to improve the fit of the LC model, the first term of the SVD is used in a generalized LC model structure that also includes age-specific cohort effects. The cohort index, which was found to be non-linear for females and males in England and Wales, is forecast using an appropriate ARIMA model, giving additional insight into future mortality. Cairns et al. (2011, 2009) later found the parameter estimates of this cohort-based extension to lack robustness. Simpler model structures were proposed by Currie

---

[7] Except that slight differences in individual k_t values may result from SVDs based on different fitting periods.

(2006) and Haberman and Renshaw (2011) to improve the stability of parameter estimates, specifically by removing the age effects of the cohort index and also, in the Currie model, of the time index. Plat (2009) proposed a model that combines the best features of previous extensions, notably Currie (2006) and Renshaw and Haberman (2006), as well as additional time indices inspired by Cairns et al. (2006, 2009). The Plat model is illustrated using data for adult males in the U.S. (1961–2005), England and Wales (1961–2005), and the Netherlands (1951–2005). For a comprehensive treatment of the identifiability and robustness of age–period–cohort models, we refer the reader to Hunt and Blake (2020b); see also (Hunt & Blake, 2020a).

In the alternative approach of Camarda and Basellini (2021), full weight is given to the features of each of their three independent age-related components (see 'Fixed b_x'). This purposeful decomposition can be contrasted with the use of higher terms of the SVD, which represent residual fractions of features that could not be subsumed under the dominant first term, and which are assembled across disparate ages for which any similarity is coincidental rather than structural.

4.7. Smoothing

While the original LC forecast of U.S. mortality benefited from the stabilizing and smoothing effect of a large population analyzed by five-year age groups, it is nevertheless the case that the fitted and forecast death rates from the LC method can be jagged and erratic, with potential anomalies in rates at adjacent ages (particularly when using single years of age). This stems from irregularity in the a_x and especially the b_x schedules, with the effect of the latter being cumulated in the forecast over time. Girosi and King (2007, 2008) remarked that the LC forecast age-specific rates will always become less smooth and eventually unrealistic, that they will deviate from any given a_x, and that this occurs independently of the initial data. Different approaches have been proposed to overcome this limitation, almost all of which employ splines, including pre-estimation smoothing of rates, in-estimation smoothing of parameters, and post-estimation smoothing of parameters.

Hyndman and Ullah (2007) view the observed death rates in a given year as a realization of an underlying smooth function. Observed rates by single years of age are smoothed using weighted penalized regression splines with the constraint of monotonicity at older ages, and with weights derived by assuming an approximate binomial distribution for deaths. The resulting smooth functions are then used in functional principal components analysis to estimate model parameters. A similar approach is employed by Rabbi and Mazzuco (2021), who smooth the observed rates using LASSO-type regularization (rather than splines) to remove observational irregularities over both age and time before model estimation.

Several approaches integrate smoothing into the estimation procedure. de Jong and Tickle (2006) proposed the "LC(smooth)" extension, where the a_x and b_x schedules are smoothed using *B*-splines, thereby reducing the number of parameters to be estimated in their integrated approach (see 'The statistical framework'), and smoothness over time is obtained by adding smoothness to the last Kalman filter in the MLE procedure. Delwarde et al. (2007a) proposed smoothing the b_x schedule by including an additional term in the fitting algorithm that penalizes irregularity in b_x, with the smoothing parameter chosen by cross-validation. Currie (2013) extended this approach within the GLM framework to also penalize irregularity in a_x. These smoothing procedures were employed by Camarda and Basellini (2021) in the Poisson-LC models for their age-related mortality components. In-estimation smoothing of b_x through the use of splines is also employed in the "PL" model proposed by Renshaw and Haberman (2003b), in which k_t is linear. In a post-estimation smoothing approach, Renshaw and Haberman (2003b) smooth estimates of b_x derived in the Gaussian or Poisson framework, by using either cubic *B*-splines or natural cubic splines. Similarly, Renshaw and Haberman (2003c) smooth the estimated b_x related to the improvements factors using parametric functions.

Smoothing is particularly relevant for rates at the oldest ages where greater variability is problematic, stemming from higher rates (which are unbounded), smaller denominators and measurement error, including to the extent that observed rates may decline with age. Different approaches have been employed to address this data-quality issue. Lee and Carter used the Coale and Guo (1989) method to estimate rates at the oldest ages based on those at younger ages (see Section 2). Hyndman and Ullah (2007) assumed monotonicity. Camarda and Basellini (2021) include shape constraints in two of their component-specific a_x parameters: log-concaveness for the early-adult component and monotonicity for the senescent component.

4.8. Prediction interval

The most important advance made by the LC method is the probabilistic prediction interval. However, the interval of the original LC forecast of life expectancy at birth was considered too narrow: specifically, Alho (1992) found it "hard to defend" the exclusion of the 1918 Spanish influenza pandemic from the error estimation in the RWD model, noting that its inclusion would have widened the prediction interval of k_t by 57% in 2065. In their rejoinder, Lee and Carter acknowledged that there are good arguments for excluding and for including the pandemic, adding that their prediction intervals "do not reflect the possibilities of nuclear war or global environmental catastrophe" (Lee & Carter, 1992, p. 675).

As noted in Section 2, the variance of forecast mortality depends on the error in estimating the LC model and the error in modeling k_t. In calculating the prediction interval of k_t, Lee and Carter considered only the innovation (which dwarfs error in the drift), and derived prediction intervals for all mortality measures accordingly. However, this takes no account of the perfect correlation in changes in logged death rates by age. Ignoring correlations would likely lead to erroneous prediction intervals for life expectancy and other complex measures. Since

the analytical estimation of the correlations is complex, the preferred approach is to derive prediction intervals from the distributions of simulated forecast values, as the large number of randomly generated future mortality schedules takes care of correlations. In this approach, prediction intervals of different measures (death rates, life expectancies, etc.) are each independently determined by the percentiles of their own simulated distribution.

Hyndman and Ullah (2007) accounted for all sources of error in the total analytical forecast variance, including random error (which was removed prior to estimation). On finding that this variance does not accord with the empirical forecast variance obtained by simulation, Hyndman and Booth (2008) derived a method for adjusting the one-step analytical variance to match the one-step empirical variance.

Several efforts have been made to empirically estimate the variability of the LC parameters using the bootstrap. This approach addresses the difficulty in estimating the variance of complex non-linear functions of the LC and ARIMA parameters, which involve two very different sources of error. Brouhns et al. (2005) proposed a Poisson bootstrap approach, in which 1000 (say) fictitious death counts by age and year are each generated as realizations of a Poisson distribution with the parameter equal to the fitted deaths from the LC method, producing 1000 mortality surfaces. The Poisson-LC is fitted to each surface, providing 1000 new parameter estimates, the percentiles of which provide pointwise confidence and prediction intervals. Similarly, Koissi et al. (2006) introduced a residual bootstrap approach, where deviance or Pearson residuals from the fitted LC model are used to generate fictitious death counts. These are then employed to refit the LC model and generate new parameter estimates and mortality forecasts with associated percentiles. Li et al. (2009) generated bootstrap death counts from fitting the LC model in a negative binomial framework with age-specific overdispersion parameters, producing much wider prediction intervals for death rates and life expectancy than the Poisson-LC.

In one of the very few applications to populations with limited data, Li et al. (2004) apply the LC method in the context of unequally spaced mortality series, and provide new equations for estimating the RWD model.

4.9. Coherent mortality forecasts

Independent mortality forecasts for similar populations often produce long-run divergence in age-specific mortality and life expectancy. This feature does not accord with the observed global convergence in mortality since the middle of the twentieth century (Wilson, 2001). Consequently, the last two decades have witnessed a growing body of research on "coherent" mortality forecasting. In coherent forecasting, a group of populations is modeled and forecast simultaneously in such a way as to ensure the non-divergence of mortality levels in the long run. The rationale is to improve the mortality forecasts of individual populations by taking into account the mortality patterns of the group. Typical groups of populations used for this purpose include males and females in the

same country, geographic sub-national populations of a country, and several similar countries; populations by socioeconomic status and education also feature.

A very early coherent approach was proposed by Carter and Lee (1992) in modeling and forecasting U.S. mortality by sex. The authors proposed a joint-k model, concatenating data for the two sexes and estimating a single k_t (with sex-specific a_x and b_x). It was not until 2005 that a further coherent method appeared. Li and Lee (2005) developed the LC method to produce non-divergent forecasts for a group of populations and proposed a sequential procedure. First, they determined the common model by identifying the a_x and b_x schedules and k_t series for the group (using the Lee and Miller (2001) extension). Second, for each population, they captured population-specific features by fitting the LC model (without k_t adjustment) to the differences from the common model, obtaining population-specific parameters b_x and k_t. While the common k_t was forecast as in the LC method, the population-specific k_t were forecast using either a random walk (without drift) or a first-order autoregressive model. Both options ensured the non-divergence of mortality levels. Applications were shown for two-sex forecasts and for a group of 15 countries.

This Li and Lee coherent method, denoted LL, has also been used with mortality improvement factors for a group of five countries (Schinzinger et al., 2016). Kleinow (2015) modified the LL method to estimate population-specific time effects (rather than a common time effect), while retaining the common age effect for males in 10 countries. Enchev et al. (2017) embedded the LL method and several of its extensions within a maximum likelihood framework.

The LL method has been extended to include two (or more) dimensions of coherence. Van Baal et al. (2016) included sex-specific and sex-education-specific age and time factors, in addition to the common age and time factors. They illustrated this method with data for the population aged 65 years and older in the Netherlands, using non-linear time-series models, producing overall, sex-specific, and education-specific mortality forecasts. The latter indicated that educational differences in life expectancy will increase. Chen and Millossovich (2018) extended the LL method by including a common time factor for the total U.K. population, sex-specific period factors, and country-specific time factors. By introducing a Poisson framework, parameter estimation was achieved in a single step. Lyu et al. (2021) extended the LL method to incorporate coherence across both countries and causes of death using a two-step beta convergence test. Janssen et al. (2013) employed LL to incorporate coherence by sex and country in forecasting total and non-smoking-related mortality in the Netherlands (smoking-related mortality was forecast using an epidemiological model). A similar procedure was extended to additionally include alcohol- and obesity-related mortality (Janssen et al., 2021).

Bergeron-Boucher et al. (2017) adopted LL to produce a coherent forecasting method based on their CoDa extension using the life-table age-at-death distribution, denoted the "CoDa-coherent" model. Kjærgaard et al. (2020) extend the CoDa-coherent model by using common age parameters and multiple time parameters to

forecast mortality by socioeconomic status in Denmark, dividing the population into five subgroups using an affluence index. The resulting forecasts indicate that the large mortality differentials observed from 1985 to 2016 will persist into the future. Kjærgaard et al. (2019) propose two additional CoDa-coherent extensions to model and forecast several causes of death. Shang et al. (2022) extend CoDa to multivariate and multilevel functional time series methods for modeling and forecasting the two sexes. Functional principal components analysis with univariate time series models are used to derive life-table age-at-death distribution forecasts.

Hyndman et al. (2013) extended the functional data approach of Hyndman and Ullah (2007) to derive coherent forecasts of mortality for several populations based on product and ratio functions of smoothed rates. Respectively, these functions are the geometric mean of the several population rates and the ratio of individual-population rates to these mean rates. These are modeled using the LC model structure and functional principal components (see 'The statistical framework). The time indices for the product function components (several SVD terms) are forecast using appropriate ARIMA models, while the ratio function forecasts are restricted to stationary time series to ensure non-divergence. Shang (2016) proposes a multilevel functional principal components approach to model multiple sets of functions that may be correlated among groups. Age-, sex-, and state-specific mortality data for Australia are used to illustrate the method, with convergence in forecasts achievable by sex and/or state.

Coherence by sex is perhaps the most common application of coherent forecasting. de Jong et al. (2016) incorporated sex coherence by embedding the LC model in a complex number framework and explicitly modeling the relationship between the sexes. This "complex LC" extension is broadly comparable with independent models in terms of model fit, and its parameters are informative and interpretable. A disadvantage of the complex number framework is that only two populations can be modeled. The authors did not use their approach in forecasting. Li et al. (2015) incorporate coherence by sex into the seven model structures of Cairns et al. (2009). Using the product-ratio method, Booth (2020) showed that sex coherence improves forecast performance for male mortality but not for female mortality; sex differences in accuracy and bias are substantially reduced compared with independent forecasts, giving more balanced confidence in male and female forecasts.

Within the Poisson framework, Butt and Haberman (2009) introduced the "stratified-LC" model to analyze and forecast the mortality of population subgroups differentiated by an additional covariate (other than age and time). This is achieved by including in the LC model a subgroup-specific a_x^i parameter, which has a direct additive effect on the logged death rate. Villegas and Haberman (2014) model and forecast mortality in England by socioeconomic status defined by the quintiles of a deprivation index. Working within the LC parameterization, the authors propose the "LC relative" model, which considers subgroup mortality relative to total (reference) mortality. Their forecasts show a widening in relative terms of mortality differences between the highest and lowest quintiles.

5. Forecasting evaluations and comparisons

A crucial requirement of any forecast is accuracy. While ex post evaluations, of the type presented in Section 3, are of substantive interest, the more useful approach for comparing forecasting methods is out-of-sample evaluation. In such exercises, the fitting and forecast periods are both in the past, allowing for the immediate evaluation of the forecast against observations. Point and interval forecast accuracy and bias all have relevance, although often the focus is limited to point forecast accuracy.

In an early evaluation, Booth et al. (2005) compared the LC method (denoted LC in this section) with the Lee and Miller (2001) and Booth et al. (2002) extensions, denoted LM and BMS, respectively. The comparison showed that, based on 15-year forecasts for the sex-specific populations of 10 developed countries, LM and BMS are generally more accurate than LC, and that the lesser accuracy of LC is mainly due to the jump-off error and the longer fitting period, rather than the different k_t adjustment. Booth et al. (2006) extended this comparison to include the LC extensions of Hyndman and Ullah (2007), denoted HU, and de Jong and Tickle (2006), keeping the same populations and forecast period. This comparison found that all five extensions are more accurate than LC in terms of logged death rates (by up to 61%), but no significant differences were found among them in terms of the point forecast accuracy of life expectancy. While these evaluations are based on errors averaged over forecast horizons, the jump-off year was fixed such that the conclusions are limited to the specific forecast period.

Also using a fixed jump-off year, Rabbi and Mazzuco (2021) compared their extension, denoted RM, with LC, LM, BMS, Poisson-LC, HU, and two HU extensions over a 10-year forecasting period for females in 20 low-mortality countries. The authors show that their RM extension, together with HU and its extensions, produces more accurate death-rate forecasts than the other methods, and that their life expectancy forecasts are not systematically biased. Further, the number of years with an incorrect life expectancy interval forecast is no greater than for other methods. Camarda and Basellini (2021) compared their three-component smooth LC extension, denoted CB, with HU, the Poisson-LC with smooth parameters (Currie, 2013), and two methods that are not based on the LC framework. Comparing 10-year forecasts of sex-specific populations for five developed countries, the authors found the mortality forecasts of the LC-based methods to be less accurate than the non-LC based methods; moreover, the CB extension performed worse than HU and the Currie extension.

All of the above evaluations are limited by the use of a fixed jump-off year in defining the forecast period. More generalized evaluations make use of a rolling jump-off year so that the forecast horizon is no longer synonymous with the calendar year. Hyndman and Ullah (2007) compared their functional data forecasts with those of LC, LM, and BMS, using a rolling jump-off year with a forecast period of 20 years. Using French mortality data from 1899 to 2001 and jump-off years from 1958 to 1981, the authors demonstrated the better point forecast

accuracy of HU compared with LC and BMS for all forecast horizons and with LM for horizons greater than seven years. Bergeron-Boucher et al. (2020) used three different starting years (and a fixed jump-off year) to evaluate the length of the fitting period. In the context of the non-linear development of life expectancy in Denmark, the authors found that longer fitting periods, rather than shorter ones, would have provided more accurate forecasts. Booth (2020) used a rolling fitting period of fixed length to remove the effect of the fitting period in the evaluation of different reference standards for coherent forecasting.

Bergeron-Boucher and Kjærgaard (2022) used varying fitting and forecasting periods to evaluate the LC method for ages 65 years and older by sex in four countries where LC is used by the national statistical office. They found that LC underpredicts life expectancy at older ages, especially for males, because of the fixed b_x; further, a lack of robustness was indicated by the large variation in forecast errors. Accuracy and robustness improved by employing LL or the rotation of b_x, although neither consistently performed best. Dong et al. (2020) showed that, based on a variety of fitting and forecasting periods, their tensor approach substantially improved predictive performance, compared with independent forecasts, both for a group of ten European countries and for males and females within these individual countries.

Very recently, several studies have demonstrated the superior accuracy achieved by enhancing the LC model by machine learning methods. Levantesi and Pizzorusso (2019) showed that their LC model enhanced by machine learning improved forecast accuracy compared with the LC method (without k_t adjustment) in two forecasts of Italian male mortality. Nigri et al. (2019) used data by sex for six countries and fixed fitting and forecast periods to evaluate the performance of their RNN method and the best fitting ARIMA model (Hyndman & Khandakar, 2008) in forecasting the LC k_t series, finding the recurrent neural network (RNN) approach to be superior. Marino et al. (2022) showed that, for three countries by sex and two fitting periods with fixed forecast periods, their RNN LC approach improved point and interval forecast accuracy compared with the Poisson-LC.

The evaluation of coherent mortality forecasting methods against independent methods generally shows improvement. Hyndman et al. (2013) compared their coherent method with comparable independent functional forecasts of mortality for Sweden by sex and for Australia by state. The authors found their coherent forecasts to be at least as accurate as independent forecasts, with coherent forecast accuracy being less variable across sub-populations, which they see as a practical advantage. Booth (2020) used this method with a low-mortality standard and data for 21 countries by sex, demonstrating substantial gains in accuracy and bias, with reduced variability. Richman and Wüthrich (2021) showed that their neural networks approach improves 15-year forecasts, compared with both LC and LL forecasts, for several low-mortality countries. Enchev et al. (2017) evaluated the performance of the LL method and two extensions against the common-age-effect model (Kleinow,

2015). Using data from six countries and a fixed fitting and forecast period, they found that the common-age-effect model outperforms the other methods and that LL can suffer in terms of robustness when estimated by maximum likelihood.

Bergeron-Boucher et al. (2017) compared forecasts obtained from their independent and coherent CoDa extensions with forecasts from the LC and LL methods. Based on data for 15 Western European countries and a 15-year forecast period with a fixed jump-off year, this comparison shows that CoDa-coherent and LL are more accurate than their non-coherent counterparts, as expected, in terms of the age-at-death distribution and death rates, respectively, but that there is no difference between the two coherent extensions. In terms of bias in life expectancy at birth, the independent CoDa forecasts are the least biased, and while LC and LL tend to underpredict life expectancy, coherent CoDa tends to overpredict it.

Kjærgaard et al. (2019) used cancer mortality by sex in France and the Netherlands, with forecast horizons of 10 to 20 years and different jump-off years, to compare their two newly proposed CoDa-coherent extensions with the original CoDa-coherent and the LC method. The authors showed that point forecast accuracy improved in three out of the four populations using one of their proposed CoDa-coherent extensions. Kjærgaard et al. (2020) evaluated the independent and coherent CoDa extensions against the LC and LL methods with a rolling jump-off year for horizons of 5–11 years for Danish mortality by socioeconomic status. The CoDa-coherent model (using national mortality to estimate the common parameters) provides a more accurate forecast of life expectancy at age 50 years than the other methods. Using a single fitting period and a six-year forecasting period, Villegas and Haberman (2014) evaluated out-of-sample forecast accuracy for mortality by socioeconomic status in England, comparing their LC-relative extension with the Poisson-LC, the joint k-model (Carter & Lee, 1992), and the three-way-LC (Russolillo et al., 2011). The authors reported that, in terms of q_x, the LC-relative is more accurate than the other approaches examined.

Fewer efforts have been made to consider the accuracy of prediction intervals in the evaluation and comparison of the mortality forecasts of different methods. Dowd et al. (2010) compared the point and interval forecast accuracy of six extensions: LC; Renshaw and Haberman (2006), denoted RH; an age-period-cohort approach (Currie, 2006); the Cairns et al. (2006) method, denoted CBD; and two CBD extensions. Using male mortality at age 65–84 years in England and Wales, the authors concluded that all models (except RH) "perform well most of the time" and "there is relatively little to choose between these models" (p. 294). Bergeron-Boucher et al. (2017) noted that CoDa-coherent prediction intervals are wider than the LL intervals. Analyzing period and cohort data by sex in two countries, with a rolling jump-off year, Shang et al. (2022) found that in most cases their proposed CoDa extensions outperformed other single- and multi-population models (including LL and the product-ratio method).

A more thorough comparison of mortality forecasts by Shang et al. (2011) considered 10 LC extensions, 14

developed countries, and the assessment of both point and interval forecast accuracy. Mortality forecasts of up to 30 years were computed using a rolling jump-off year, and evaluated based on a horizon of one year. The most accurate point forecasts of logged death rates were obtained by HU extensions, and the least biased by the LM method. All 10 extensions were found to underestimate the variability of death rates, with all HU extensions being more accurate. Most of these studies employ coverage probabilities to evaluate prediction intervals. However, as pointed out by Alho and Spencer (2005, p. 248), empirical validation of prediction intervals is problematic because of correlations in the data across age and/or time, and potentially across sex/country.

6. Discussion

In 1982, Nathan Keyfitz stated that "for forecasting one would like the succession of life tables for a given population over past times to be representable by points moving in a simple way through a parameter space – ideally in a straight line – over a succession of calendar years. The simpler the curve, the more realistic is likely to be its projection into the future" (Keyfitz, 1982, p. 329). Ten years later, Lee and Carter provided a remarkable response to Keyfitz's vision by summarizing the development of life tables in a simple time index. This review, celebrating the thirtieth anniversary of its publication, confirms that the LC method represents a significant methodological breakthrough in mortality forecasting that revolutionized the field and spawned a great deal of further research.

That research can be divided into two broad approaches to addressing the limitations of the LC method (detailed in Section 4), given that the LC model lacks a clear underlying stochastic process. On the one hand, a small body of research has retained the Gaussian framework implicit in the SVD, working primarily with logged death rates; we label this the "Gaussian approach". On the other hand, a much larger body has focused on improving the statistical framework, working with death counts (and including population size as an offset), and deriving ML estimates of the LC parameters, usually in a Poisson setting; we label this the "Poisson approach". We discuss these two approaches separately before addressing the issues and features common to them.

The Gaussian approach retains the stepped characteristic of the LC method, usually beginning with parameter estimation using SVD. The second-step adjustment of k_t to match total deaths is an ad hoc procedure to improve model fit, necessitated by the homoscedasticity assumption of the Gaussian framework (see 'Adjusting $\hat{k}_t$'). Efforts to improve the method while retaining this approach led to alternative ad hoc procedures, matching life expectancy at birth, age-specific deaths, and, latterly, lifespan disparity. The need for a second-step adjustment is eliminated in the functional data approach by pre-estimation smoothing of rates under the assumption of a binomial or Poisson distribution for death counts, thereby accounting for heteroscedasticity and better meeting the homoscedasticity assumption. Within this Gaussian approach, three-dimensional data (age, time, and an additional variable) are modeled using tensor decomposition.

This broad approach also encompasses LC modeling and forecasting of the age-at-death distribution, the probability of death, and survival functions. Coherent versions of all these approaches are included.

The Poisson approach explicitly addresses the lack of a clear stochastic process underlying the LC model. Given the inherent heteroscedasticity of death counts, an appropriate way to model mortality is to consider observed deaths (at given age and time) as a realization of a Poisson distribution, with mean equal to the product of the LC-modeled death rate and population exposure. By embedding the LC model within a Poisson framework, heteroscedasticity of the mortality process is accounted for, removing the need for the adjustment of k_t. In the Poisson framework, ML estimates of the LC parameters are derived either by the Newton–Raphson algorithm or, equivalently, by a series of GLMs (in which the LC model is linearized with respect to the LC parameters). Smoothing of a_x and b_x (and, indirectly, logged rates) over age, or age and time, is incorporated simultaneously with their estimation by penalizing the likelihood. Overdispersion is accommodated either by assuming a negative binomial distribution or by a quasi-likelihood estimation procedure. Coherent forecasting is included in this approach. Further, the approach can be used to model the probability of death, with MLE based on the binomial distribution.

In both approaches, the features (and limitations) of the LC model remain. The only attempt to explicitly address the limitation of fixed b_x introduces a specifically designed rotation of b_x over the forecasting period. This imposes a gradual convergence of b_x to an assumed ultimate age pattern of change at a given life expectancy, and therefore introduces an element of subjectivity in an extrapolative setting. Further, the corresponding adjustment of forecast k_t, to maintain the accepted (decelerating) life expectancy forecast of the LC method, means that its linearity is lost. Additionally, in the commonly used RWD model, the k_t value for the starting year of the fitting period in relation to the historical series is of considerable significance. The importance of the starting year is intrinsic to the RWD, this being the only time-series model that does not give diminishing weight to earlier points. Efforts to optimize the starting year (under the assumption of linear k_t) have been based on goodness-of-fit and have not considered the effect on the forecast.

Smoothing is also a feature of both approaches. Smoothing is of particular importance when forecast rates are of interest, but is less relevant when the focus is the point forecast of life expectancy at birth because of cumulation across age. Smoothing the mortality age pattern has the major advantage of avoiding jaggedness in forecast rates, but has the potential disadvantage of reducing historical variability, which would likely lead to prediction intervals that are too narrow (this also occurs when five-year age groups are used). One existing solution is to explicitly account for the random variability in the data and to incorporate this in the prediction interval.

The choice of fitting period, common to both approaches, can have considerable impact on the forecast and warrants serious consideration. On the one hand, a longer period limits overfitting and captures broader variation

in mortality trends; on the other hand, a shorter period better captures the most recent trend, which may be desirable when distant and recent past trends differ. It is of note that some comparative evaluations of the accuracy of different methods are based on multiple out-of-sample forecasts using a rolling starting year and/or a rolling jump-off year, with the purpose of either removing (or averaging) their effect or examining the effect of the length of the fitting period (see Section 5). To our knowledge, purposeful averaging of the starting year (as opposed to averaging the length of the fitting period) has not been a feature of actual LC-based forecasts, though it could easily be accomplished. When other (non-RWD) time-series models are employed, diminishing weight assigned to more distant data greatly reduces the influence of the starting year. Further, in the case of functional forecasting, a weighted functional principal components algorithm is used to achieve the same result.

Much of the research, on the Poisson approach in particular, is directed towards improving goodness-of-fit. The use of more than one term to describe the dynamics of change affords a degree of flexibility in the fitted and forecast age pattern of change through the combination of the several b_x schedules and k_t series. However, this flexibility can be associated with a lack of robustness. Other efforts to improve fit include the integrated estimation of the LC parameters and time-series model via Bayesian methods or the Kalman filter. Alternatively, the inclusion of additional parameters has involved adding a cohort effect, incorporating parameters estimated using machine learning techniques, and age-related decomposition with the consequent use of multiple LC models. Nonetheless, a better model fit does not generally foreshadow a more accurate forecast (Booth & Tickle, 2008). This partially explains why, for example, the accuracy of European demographic forecasts in the first decade of this century did not improve over those of the past (Keilman, 2008). A more fruitful avenue for further research is likely to be found in forecast evaluation rather than goodness-of-fit.

The success of the LC method can be attributed not only to its simplicity and objectivity, but also to its accessibility. The adoption of the method is largely dependent on the availability of open-access software. From the beginning, software was readily available: Lee and Carter used GAUSS to calculate the SVD and RATS for the time-series estimation and forecast. Over the following 30 years, somewhat obscure dedicated routines within standard statistical software have gradually developed to become the basis of structured and user-friendly packages. Notable in the Stata environment is Wang (2001), but most contributions have been made in the R open software environment. Initially released in 2010, the R package demography (Hyndman et al., 2019) represents one of the earliest and most widely employed tools for implementing the LC method. This package includes LC extensions to the Gaussian approach, e.g. with and without k_t adjustment, the LM and BMS extensions, as well as the HU and product-ratio methods. However, this package uses a specific object type, which limits its ease of use in combination with other R functions. A second R package, StMoMo (Villegas et al., 2018), includes LC extensions to

the Poisson approach, such as the Poisson-LC and RH models, as well as other models employed in actuarial science (e.g. CBD), and allows users to define their own model. A third package, ilc, implements different extensions to the LC model (within both Gaussian and Poisson approaches) and includes the stratified-LC model (Butt & Haberman, 2009). Finally, the code for other LC extensions is available in papers and as supplementary materials in public repositories (see, e.g., Camarda & Basellini, 2021; Currie, 2016; Rabbi & Mazzuco, 2021).

While the LC method and its accessibility opened up the field for further research, it has yet to be fully embraced by practitioners. What is seen as simple by some is complex for others, and the method remains esoteric for many. Its broad adoption in the actuarial profession is supported by the provision of user-friendly tools by actuarial institutes to their members. The absence of such support to demographic and statistical practitioners in general (but see Booth et al., 2014) is undoubtedly one reason why so few national statistical offices have adopted LC forecasting in the context of mortality or population projections, especially as the tradition of heeding expert opinion is firmly entrenched. It is a sad indictment of the world of applied demography – and of our success in communicating new approaches – that these methods are not more widely embraced. A greater effort to promulgate the principles of probabilistic forecasting and teach LC-based methods is essential to their adoption, not only among demographers but also among planners and other users. The United Nations has accepted the challenge of forecast uncertainty, adopting Bayesian hierarchical methods for population projections, with the LC method used to derive forecast death rates from projected life expectancy. This sets an excellent example for national statistical offices to follow. While we enthusiastically support the adoption of stochastic forecasting methods, and specifically endorse the LC approach, more detailed recommendations are beyond the scope of this review.

To conclude, the introduction of the LC method provided a simple yet powerful data-driven approach that both encompassed the novel idea of the probabilistic forecast and challenged the authority of expert opinion. There can be no doubt that the LC method broke new ground. What does the future hold for its role in mortality forecasting? We identify several directions that warrant further research. First, the significant limitation of fixed b_x remains and, as a defining feature of the LC model, does not easily lend itself to a statistical solution. The idea of rotation, used in the Gaussian approach, may offer an avenue for further research. Second, the potential gains in accuracy and insight offered by cohort data have barely been explored in forecasting and warrant further attention (for a recent exercise, see Basellini & Camarda, 2022). Third, in coherent forecasting, the impact of the standard on forecast accuracy and bias, including their variability across populations, warrants further investigation, with a view to increasing confidence in the forecasts. Fourth, in order to better differentiate the value of different LC extensions (and other methods), greater attention needs to be paid to rigorous and more comprehensive comparative research, not only with respect to model fit but

also – and more importantly – with respect to forecast performance. Crucially, both point and interval forecast accuracy and bias should be evaluated: the common practice of evaluating only the point forecast against only the LC method (often imprecisely defined) is inadequate. Fifth, in the Poisson approach, exposures are typically (if not always) taken as given; considering exposures as endogenous would be a welcome advancement of the estimation procedure. Sixth, in line with forecast averaging for evaluation purposes, the performance of actual forecasts may be improved through averaging over starting and jump-off years; research in this area may also lead to increased confidence in forecast performance. Further, insofar as in mortality forecasting the uncertainty within a given model is generally as great as the uncertainty among models (Richards & Currie, 2009), model averaging may be a fruitful avenue of research (see, e.g., Shang, 2012).

Last, it is abundantly clear, writing in the midst of the Covid-19 pandemic, that the theoretical basis of the LC method – that the past is a good guide to the future – can no longer be blithely assumed, at least for short term forecasts after 2019. The major challenge for any extrapolative methodology – dealing with short-term shocks or structural change – has assertively raised its head in mortality forecasting. With the benefit of hindsight, the influence of the previous pandemic on the forecast could be removed through modeling, but in the present situation, hindsight is not at hand. Capturing the effect of the pandemic in forecasting mortality, other than its inclusion in the fitting period, arguably represents a pressing need for further research. However, current data offer very limited indication of the duration or ongoing development of the pandemic's effects on mortality, and cannot anticipate future medical breakthroughs or further novel disease. Indeed, in the words of Lee and Carter (p. 665): "We must keep in mind, however, that no amount of analysis of the past can assure us that dramatic structural changes will not occur in the future".

CRediT authorship contribution statement

Ugofilippo Basellini: Conceived the study, Designed research, Performed analysis, Writing – review & editing, Writing - original draft. **Carlo Giovanni Camarda:** Conceived the study, Designed research, Writing – review & editing. **Heather Booth:** Designed research, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alho, J. M. (1992). Modeling and forecasting U.S. mortality: Comment. *Journal of the American Statistical Association*, *87*(419), 673–674.

Alho, J. M., & Spencer, B. D. (2005). *Statistical demography and forecasting*. Springer.

Basellini, U., & Camarda, C. G. (2022). Lee-Carter cohort mortality forecasts. In *PAA 2022 annual meeting*. Atlanta, Georgia.

Bell, W. R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, *13*(3), 279–303.

Bell, W. R., & Monsell, B. (1991). Using principal components in time series modeling and forecasting of age-specific mortality rates. In *Proceedings of the American statistical association, social statistics section* (pp. 154–159). US Bureau of the Census Suitland.

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

Bergeron-Boucher, M.-P., & Kjærgaard, S. (2022). Mortality forecasting at age 65 and above: An age-specific evaluation of the Lee-Carter model. *Scandinavian Actuarial Journal*, *2022*(1), 64–79.

Bergeron-Boucher, M.-P., Kjærgaard, S., Oeppen, J., & Vaupel, J. W. (2019). The impact of the choice of life table statistics when forecasting mortality. *Demographic Research*, *41*, 1235–1268.

Bergeron-Boucher, M.-P., Kjærgaard, S., Pascariu, M. D., Aburto, J. M., Alvarez, J.-A., Basellini, U., Rizzi, S., & Vaupel, J. W. (2020). Alternative forecasts of danish life expectancy. In S. Mazzuco, & N. Keilman (Eds.), *Developments in demographic forecasting* (pp. 131–151). Cham: Springer International Publishing.

Booth, H. (2020). Coherent mortality forecasting with standards: Low mortality serves as a guide. In S. Mazzuco, & N. Keilman (Eds.), *Developments in demographic forecasting* (pp. 153–178). Cham: Springer International Publishing.

Booth, H., Hyndman, R. J., & Tickle, L. (2014). Prospective life tables. In A. Charpentier (Ed.), *Computational actuarial science with R* (pp. 319–344). Chapman & Hall/CRC.

Booth, H., Hyndman, R. J., Tickle, L., & De Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, *15*, 289–310.

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*(1–2), 3–43.

Booth, H., Tickle, L., & Smith, L. (2005). Evaluation of the variants of the Lee-Carter method of forecasting mortality: a multi-country comparison. *New Zealand Population Review*, *31*(1), 13–34.

Brass, W. (1971). On the scale of mortality. In W. Brass (Ed.), *Biological aspect of demography*. London: Taylor & Francis.

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics*, *42*, 693–734.

Brouhns, N., Denuit, M., & Van Keilegom, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, *3*, 212–224.

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

Butt, Z., & Haberman, S. (2009). ilc: a collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms. *Technical report*, London, UK: Faculty of Actuarial Science & Insurance, City University London.

Cairns, A. J., 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*(4), 687–718.

Cairns, A. J., Blake, D., & Dowd, K. (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, *2008*(2–3), 79–113.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics & Economics*, *48*(3), 355–367.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & 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.

Camarda, C. G., & Basellini, U. (2021). Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, *37*(3), 569–602.

Carter, L. R., & Lee, R. D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, *8*(3), 393–411.

Chen, R. Y., & Millossovich, P. (2018). Sex-specific mortality forecasting for UK countries:a coherent approach. *European Actuarial Journal* , *8*, 69–95.

Coale, A., & Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index*, 613–643.

Currie, I. D. (2006). Smoothing and forecasting mortality rates with P-splines. Talk given at the Institute of Actuaries, June 2006. Available at https://www.actuaries.org.uk/system/files/documents/pdf/currie0.pdf.

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

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

Currie, I. D. (2020). Constraints, the identifiability problem and the forecasting of mortality. *Annals of Actuarial Science*, *14*(2), 537–566.

Czado, C., Delwarde, A., & Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics & Economics*, *36*(3), 260–284.

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

Delwarde, A., Denuit, M., & Partrat, C. (2007). Negative binomial version of the Lee–Carter model for mortality forecasting. *Applied Stochastic Models in Business and Industry*, *23*(5), 385–401.

Deprez, P., Shevchenko, P. V., & Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal* , *7*(2), 337–352.

Dong, Y., Huang, F., Yu, H., & Haberman, S. (2020). Multi-population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, *2020*(8), 754–775.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2010). Backtesting stochastic mortality models: An ex post evaluation of multiperiod-ahead density forecasts. *North American Actuarial Journal*, *14*(3), 281–298.

Enchev, V., Kleinow, T., & Cairns, A. J. G. (2017). Multi-population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, *4*, 319–342.

Girosi, F., & King, G. (2007). Understanding the Lee-Carter mortality forecasting method. *Technical report*, RAND Corporation.

Girosi, F., & King, G. (2008). *Demographic forecasting*. Princeton University Press.

Goldman, N., & Lord, G. (1986). A new look at entropy and the life table. *Demography*, *23*(2), 275–282.

Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics & Economics*, *48*(1), 35–55.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Human Mortality Database (2022). University of california, berkeley (USA) and Max Planck institute for demographic research (Germany). Available at www.mortality.org or www.humanmortality.de. Data retrieved on June 27, 2022.

Hunt, A., & Blake, D. (2020a). Identifiability in age/period mortality models. *Annals of Actuarial Science*, *14*(2), 461–499.

Hunt, A., & Blake, D. (2020b). Identifiability in age/period/cohort mortality models. *Annals of Actuarial Science*, *14*(2), 500–536.

Hyndman, R. J., & Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, *24*(3), 323–342.

Hyndman, R. J., Booth, H., Tickle, L., & Maindonald, J. (2019). Demography: Forecasting mortality, fertility, migration and population data. R package version 1.22.

Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, *50*(1), 261–283.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *26*(3), 1–22.

Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*, 4942–4956.

Janssen, F. (2018). Advances in mortality forecasting: introduction. *Genus*, *74*(1).

Janssen, F., Bardoutsos, A., El Gewily, S., & De Beer, J. (2021). Future life expectancy in Europe taking into account the impact of smoking, obesity, and alcohol. In B. Nicolau, E. Franco, U. Basellini, & C. Payne (Eds.), *ELife*, *10*, Article e66590.

Janssen, F., van Wissen, L. J. G., & Kunst, A. E. (2013). Including the smoking epidemic in internationally coherent mortality projections. *Demography*, *50*(4), 1341–1362.

de Jong, P., & Tickle, L. (2006). Extending Lee-Carter mortality forecasting. *Mathematical Population Studies*, *13*, 1–18.

de Jong, P., Tickle, L., & Xu, J. (2016). Coherent modeling of male and female mortality using Lee–Carter in a complex number framework. *Insurance: Mathematics & Economics*, *71*, 130–137.

de Jong, P., Tickle, L., & Xu, J. (2020). A more meaningful parameterization of the Lee–Carter model. *Insurance: Mathematics & Economics*, *94*, 1–8.

Kannisto, V., Lauritsen, J., Thatcher, A. R., & Vaupel, J. W. (1994). Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review*, *20*, 793–810.

Keilman, N. (2008). European demographic forecasts have not become more accurate over the past 25 years. *Population and Development Review*, *34*(1), 137–153.

Keyfitz, N. (1982). Choice of function for mortality analysis: Effective forecasting depends on a minimum parameter representation. *Theoretical Population Biology*, *21*(3), 329–352.

Kjærgaard, S., Ergemen, Y. E., Bergeron-Boucher, M.-P., Oeppen, J., & Kallestrup-Lamb, M. (2020). Longevity forecasting by socio-economic groups using compositional data analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *183*(3), 1167–1187.

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*(5), 1351–1370.

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics & Economics*, *63*, 147–152.

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

Ledermann, S., & Breas, J. (1959). Les dimensions de la mortalité. *Population (French Edition)*, 637–682.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality [with rejoinder]. *Journal of the American Statistical Association*, *87*(419), 659–671.

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

Gómez de León, J. (1990). Empirical EDA models to fit and project time series of age specific mortality rates. *Technical report 50*, Oslo: Statistics Norway, Research Department.

Levantesi, S., & Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks*, *7*(1), 26.

Li, J. S.-H., Hardy, M. R., & Tan, K. S. (2009). Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. *Astin Bulletin*, *39*(1), 137–164.

Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, *42*(3), 575–594.

Li, N., Lee, R. D., & Gerland, P. (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality-decline for long-term projection. *Demography*, *50*, 2037–2051.

Li, N., Lee, R. D., & Tuljapurkar, S. (2004). Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review*, *72*, 19–36.

Li, J. S.-H., Zhou, R., & Hardy, M. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics & Economics*, *63*, 121–134.

Lyu, P., De Waegenaere, A., & Melenberg, B. (2021). A multi-population approach to forecasting all-cause mortality using cause-of-death mortality data. *North American Actuarial Journal*, *25*(sup1), S421–S456.

Marino, M., Levantesi, S., & Nigri, A. (2022). A neural approach to improve the Lee-Carter mortality density forecasts. *North American Actuarial Journal*, 1–18.

McNown, R., & Rogers, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography*, *26*(4), 645–660.

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., & Perla, F. (2019). A deep learning integrated Lee–Carter model. *Risks*, *7*(1).

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

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

Pollard, J. H. (1987). Projection of age-specific mortality rates. *Population Bulletin of the United Nations*, *21–22*, 55–69.

Rabbi, A. M. F., & Mazzuco, S. (2021). Mortality forecasting with the Lee–Carter method: Adjusting for smoothing and lifespan disparity. *European Journal of Population*, *37*, 97–120.

Rau, R., Soroko, E., Domantas, J., & Vaupel, J. W. (2008). Continued reductions in mortality at advanced ages. *Population and Development Review*, *34*, 747–768.

Renshaw, A., & Haberman, S. (2003). Lee–Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *52*(1), 119–137.

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

Renshaw, A., & Haberman, S. (2003). On the forecasting of mortality reduction factors. *Insurance: Mathematics & Economics*, *32*, 379–401.

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

Richards, S. J., & Currie, I. D. (2009). Longevity risk and annuity pricing with the Lee-Carter model. *British Actuarial Journal*, *15*(2), 317–343.

Richman, R., & Wüthrich, M. V. (2021). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, *15*(2), 346–366.

Russolillo, M., Giordano, G., & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, *2011*(2), 96–117.

Schinzinger, E., Denuit, M. M., & Christiansen, M. C. (2016). A multivariate evolutionary credibility model for mortality improvement rates. *Insurance: Mathematics & Economics*, *69*, 70–81.

Shang, H. L. (2012). Point and interval forecasts of age-specific life expectancies: A model averaging approach. *Demographic Research*, *27*, 593–644.

Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method. *The Annals of Applied Statistics*, *10*(3), 1639–1672.

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

Shang, H. L., Haberman, S., & Xu, R. (2022). Multi-population modelling and forecasting life-table death counts. *Insurance: Mathematics & Economics*.

Stoeldraijer, L., van Duin, C., van Wissen, L., & 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*(13), 323–354.

Stoeldraijer, L., van Duin, C., van Wissen, L., & Janssen, F. (2018). Comparing strategies for matching mortality forecasts to the most recently observed data: exploring the trade-off between accuracy and robustness. *Genus*, *74*(1), 1–20.

Tabeau, E. (2001). A review of demographic forecasting models for mortality. In E. Tabeau, A. van den Berg Jeths, & C. Heathcote (Eds.), *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective* (pp. 1–32). Dordrecht: Springer Netherlands.

Tuljapurkar, S., Li, N., & Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. *Nature*, *405*(6788), 789–792.

United Nations (2022). World Population Prospects 2022: Methodology of the United Nations population estimates and projections. Department of Economic and Social Affairs, Population Division. UN DESA/POP/2022/TR/NO. 4.

Van Baal, P., Peters, F., Mackenbach, J., & Nusselder, W. (2016). Forecasting differences in life expectancy by education. *Population Studies*, *70*(2), 201–216.

Vaupel, J. W. (1986). How change in age-specific mortality affects life expectancy. *Population Studies*, *40*(1), 147–157.

Vaupel, J. W., & Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography*, *40*, 201–216.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454.

Vékás, P. (2020). Rotation of the age pattern of mortality improvements in the European Union. *Central European Journal of Operations Research*, *28*(3), 1031–1048.

Villegas, A. M., & Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal*, *18*(1), 168–193.

Villegas, A. M., Kaishev, V. K., & Millossovich, P. (2018). StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software*, *84*(3), 1–38.

Ševčíková, H., Li, N., Kantorova, V., Gerland, P., & Raftery, A. E. (2016). Age-specific mortality and fertility rates for probabilistic population projections. In R. Schoen (Ed.), *The springer series on demographic methods and population analysis: Vol. 39. Dynamic demographic analysis* (pp. 285–310). Springer.

Wang, D. (2001). Modeling mortality data using the Lee-Carter model. *Stata Technical Bulletin*, *10*(57).

Wilmoth, J. R. (1990). Variation in vital rates by ages, period and cohort. *Sociological Methodology*, *20*, 295–335.

Wilmoth, J. R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. *Technical report*, Department of Demography, University of California, Berkeley.

Wilson, C. (2001). On the scale of global demographic convergence 1950–2000. *Population and Development Review*, *27*(1), 155–171.