Noname manuscript No.
(will be inserted by the editor)

Mortality forecasting with Lee-Carter method: Adjusting for smoothing and lifespan disparity

the date of receipt and acceptance should be inserted later

Abstract Reliable mortality forecasts are essential part for health care policies in aging societies. Lee-Carter method and its later variants are widely accepted probabilistic approach for mortality forecasting due to simplicity and straightforward interpretation of the model parameters. This model assumes an invariant age component and linear time component for forecasting. We propose to apply the Lee-Carter method on smoothed mortality rates obtained by LASSO type regularization and hence to adjust the time component with observed lifespan disparity. Smoothing with lasso produces less error during fitting period compare to spline based smoothing techniques. As a more informative indicator of longevity, matching with lifespan disparity made the time component more reflective of the mortality improvement. The forecasts produced by the new method are more accurate during out-of-sample evaluation and provides optimistic forecast for many low-mortality countries.

Keywords Mortality forecasting · Mortality smoothing · Lifespan disparity · LASSO

1 Introduction

Mortality improvements are observed in all industrialized countries during twentieth century. Since the second half of the last century, the main factor driving continued gains in life expectancy in industrialized countries are highly attributable to reduction in death rates among the elderly (Booth et al. 2002). This produce a sarge in stochastic forecast of mortality as accurate forecasts of mortality and life expectancy are a core requirement for decision making in social, health-care and financial sectors. Fundamental changes of welfare policies depend highly on the accurate forecast of longevity. United Nations and several industrialized countries also adapted stochastic forecasting techniques in this context (Hyndman and Booth 2008). Several different approaches have been developed in stochastic mortality forecasting over time. Many of these techniques involve some forms of extrapolation, often utilizing standard time series methods. The most prominent method till now is that proposed by Lee and Carter (1992). This method is a simple but powerful extrapolative one, which captures the trend of mortality pattern through principal component techniques. Then it decomposes the differences of the observed log-mortality rates and their average into two parts; an invariant age component and a linear time component. Mortality forecast is done by standard time series forecasting on the time component; considering same structure of age-specific mortality level over time.

One limitation of this method is that its time component (presumably linear) does not perform well to capture the mortality trend for several countries. A simple modification was suggested by Lee and Carter (1992) to overcome this problem. They proposed to re-estimate the time component with observed total number of deaths. In another version of this model, re-estimation of the time component is proposed according to observed life expectancy at birth (Lee and Miller 2001). Booth et al. (2002) proposed to adjust the time component with respect to age distribution of deaths. Besides adjustment of time component, both of these later two approaches considered other different modifications to improve the forecast accuracy.

Age component can be inaccurate, as well, due to noise in the data and the presence of outlier (see Hyndman and Ullah 2007, for example). This issue lead many scholars to apply some smoothing techniques (Booth and Tickle 2008; Girosi and King 2006). Smoothing is particularly useful for Lee-Carter (LC) variants to increase precision. All LC type models estimate the product of time component (k_t) and age component (b_x) obtained from singular value decomposition of mortality matrix and then add it with observed average log-mortality rates (a_x) to obtain the fitted death rates. Although LC method assumes linear trend of the time component, still this is not completely linear even after all proposed adjustment policies. Similarly, the estimated age component have irregular trend in different parts of the life span. The product of these two parameters with irregular trend become more jagged over time affecting the estimated log-mortality rates obtained from the LC model (Girosi and King 2006). To illustrate this, the product of b_x and k_t for Swedish Females are plotted in Figure 1 where irregular trends are visible in earlier and later part of lifespan.

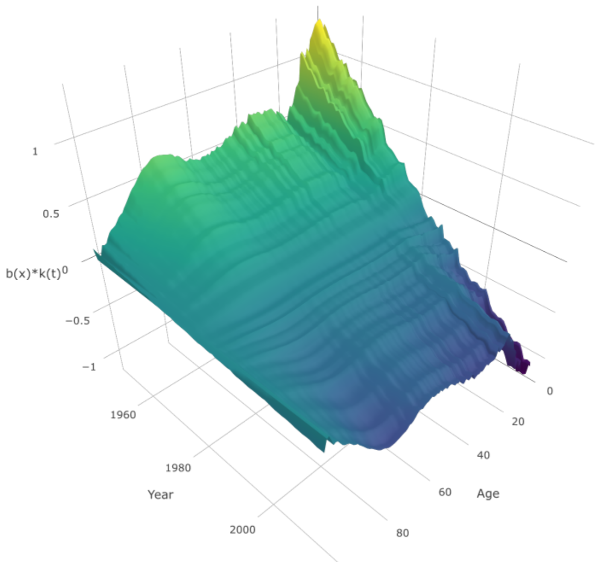


Fig. 1 Product of time component and age component for Swedish Females (1950:2016) from the fitted Lee-Carter model (Lee and Carter 1992).

To overcome this problem, different smoothing techniques can be applied in three different ways: (a) smoothing the observed mortality rates first and then fitting the model for forecasting; (b) smoothing over the fitted parameters and to make the forecast by standard time series techniques or, (c) applying smoothing on fitted/forecasted rates by the model. De Jong and Tickle (2006) combined spline smoothing and estimation via the Kalman filter to fit a generalized version of the Lee-Carter model, whereas Hyndman and Ullah (2007) proposed smoothing the mortality rates for each year using constrained regression splines prior to fitting a model using principal components decomposition following the functional data paradigm. Hyndman and Ullah (2007) smooths the mortality rates for each of the year considering all the available age-groups. This approach further extended the standard LC method for more than one principal component using functional data analysis and fitting the model using these smoothed mortality rate (Hyndman and Ullah 2007). Later two dimensional smoothing is proposed by Camarda et al. (2012) who considers both the age and period effect on mortality following the work of Currie et al. (2006). One of the shortcomings of these spline-based smoothing techniques is that they might over-smooth the mortality curves thus reducing the accuracy of smoothed data to be used for model fitting (Dokumentov et al. 2018). To address this issue, a smoothing technique needs to find an optimal balance between reducing noise and keeping accuracy of the observed data, particularly for mortality forecasting.

In this paper we propose to fit the LC model over smoothed mortality rates and to adjust the fitted time component of the model according to observed lifespan disparity (Vaupel and Canudas-Romo 2003; Zhang and Vaupel 2009). Instead of spline based smoothing techniques we smoothed the observed mortality rates first using *LASSO* type regularization (Dokumentov et al. 2018). Lasso smooths the data almost similarly like spline based techniques for all over the life span, whereas it produces less error to fit the data. Compare to age distribution of deaths, total number of deaths or life expectancy at birth; lifespan disparity is a more informative indicator of mortality improvement and longevity and it is more stable over time (Vaupel and Canudas-Romo 2003; Zhang and Vaupel 2009). Unlike life expectancy at birth, lifespan disparity can provide further information about shrinking or expansion of mortality and moreover it is an indicator of mortality shifting (Zhang and Vaupel 2009). Due to its stability over time, previous study considered $e_0^\dagger$ to evaluate forecast performances of different forecasting techniques rather than just considering the fitted mortality rates or life expectancy (Bohk-Ewald et al. 2017).

2 Methods

2.1 Spline based smoothing techniques

For one dimensional smoothing, Hyndman and Ullah (2007) applied weighted penalized regression splines independently for each year. This involves calculating a vector β which minimizes the following:

$$|w(y) - X\beta|^2 + \lambda^2 \beta^T D\beta. \tag{1}$$

Here, y is a vector of observations (mortality rates); X is a matrix representing linear spline bases; $D = diag(0, 0, 1, 1, \ldots 1)$ is a diagonal matrix; w is a vector of weights; and λ is the smoothing parameter. For smoothing mortality rates, observations in year t are given by $y_i = m_{x_{i,t}}$ for age group x_i years old (where $i = 0, \ldots, 110+$). The weights w_i are taken as the inverse of the estimated variances of y_i. Assuming the life table deaths follow a Poisson distribution, Hyndman and Ullah (2007) estimated the variance of y_i as $\sigma^2 \approx (Ex_{i,t}Mx_{i,t})^{-1}$ by Taylor series approximation. Here $Ex_{i,t}$ is the mid-year population of people aged x_i years in year t. Moreover these splines are constrained to ensure that the resulting function $f(x)$ is monotonically increasing for $x > c$ for some c (for example, 60 years). Hyndman and Ullah (2007) proposed to use a modified version of the method described by Wood

(1994) to implement this constraint.

The second approach is the two dimensional splines (Camarda et al. 2012). This method fits a two-dimensional P-spline model with equally-spaced B-splines along X and Y axes (age and calendar year respectively). The response variables must be a matrix of Poisson distributed death counts in this approach. For this splines offset can be provided, else all weights are assumed to be unity. In a Poisson regression setting applied to actual death counts, the offset will be the logarithm of the matrix of exposure population.

To smooth the mortality rates, this method utilizes a smoothing function which is the *Kronecker* product of B-spline basis over the two axes and includes a discrete penalization directly on the differences of the B-splines coefficients. The smoothing parameters λ are mainly used to tune the smoothness/accuracy of the fitted values (Currie et al. 2006). For optimizing the smoothing parameters, both AIC or BIC can be considered.

2.2 Smoothing by *LASSO*

For the proposed modifications in Lee-Carter (LC) methodology, we smooth the mortality rates before model fitting and we apply *LASSO* type regularization. Dokumentov et al. (2018) defined the lasso derived from a two-dimensional thin plate spline which is used to smooth the mortality rates considering age and period effects. For observed mortality rates y, age x and year t, the two dimensional thin plate spline is defined as a function $f(x,t)$ which minimizes the following:

$$\mathcal{J}\left(\{y_i\}_{i=i}^{n}, f\right) = \sum_{i=1}^{n} \left(y_i - f(x_i,t_i)\right)^2 + \lambda \int \left[\frac{\delta^2 f}{\delta x^2} + 2\frac{\delta^2 f}{\delta x \delta t} + \frac{\delta^2 f}{\delta t^2}\right] dxdt. \quad (2)$$

Here, $\lambda > 0$ is the smoothing parameter with $(x_i, t_i)_{i=1}^{n}$ knots (following, Wood 2006). The expression in equation (2) can be approximated by a sum if the knots form a regular grid. In that case, the second partial derivatives can be approximated also as linear combinations of function values at nearby knots. Denoting the mortality rates as a vector y as before (which is a two dimensional data packed as vector) and letting $\{f(x_i,t_i)\}_{i=1}^{n}$ as vector z, the expression of (2) can be re-written as,

$$\mathcal{J}(y,z) \approx ||y-z||_{L_2}^2 + \frac{\lambda}{n}\left(||D_{xx}z||_{L_2}^2 + ||D_{xt}z||_{L_2}^2 + ||D_{tt}z||_{L_2}^2\right), \quad (3)$$

where,

$$D_{xx} = \left\{\frac{\delta^2}{\delta x^2} f(x_i,t_i)\right\}_{i=1}^{n}, \; D_{xt} = \left\{\frac{\delta^2}{\delta x \delta t} f(x_i,t_i)\right\}_{i=1}^{n} \text{ and } D_{tt} = \left\{\frac{\delta^2}{\delta t^2} f(x_i,t_i)\right\}_{i=1}^{n}.$$

This expression of equation (3) can be approximated with thin plate spline computed in its knots taking the points of z for which right side of equation (3) is minimized. Following Schuette (1978), Dokumentov et al. (2018) replaced the L_2 norm by a L_1 norm to smooth with quintile Lasso. Schuette (1978) showed that L_1 norm is more robust than L_2 norm in presence of outliers, which occurs very often in case of mortality data. (Schuette 1978) proposed to use different λ coefficients before every derivative to adjust the influence of each derivative distinctly on smoothing. Hence the smoothing can be defined as,

$$Q(y) = \arg\min_{z} \{K(y, z)\}, \tag{4}$$

where, $K(y, z) = ||y - Mz||_{L_1} + \lambda_{xx} ||D_{xx}z||_{L_1} + \lambda_{xt} ||D_{xt}z||_{L_1} + \lambda_{tt} ||D_{tt}z||_{L_1}$.

Here, M will be an identity matrix as same number of knots and data points are considered which are positioned at same places (Dokumentov et al. 2018). Following Wood (2006), Dokumentov et al. (2018) stacked matrices M, $\lambda_{xx}D_{xx}$, $\lambda_{xt}D_{xt}$, $\lambda_{tt}D_{tt}$ on top of each other to give

$$R = [M', \lambda_{xx}D'_{xx'}, \lambda_{xt}D'_{xt'}, \lambda_{tt}D'_{tt}]' = \begin{bmatrix} I & I \; I \\ \lambda_{xx}D_{xx} & 0 \; 0 \\ \lambda_{xt}D_{xt} & 0 \; 0 \\ \lambda_{tt}D_{tt} & 0 \; 0 \end{bmatrix}. \tag{5}$$

In next step, the data vector y is extended by zeros until its length is equal to the number of rows in R. So, the extended y became: $y_{ext} = [y', 0']'$. The right hand side of equation (4) is then replaced with the equivalent expression,

$$K(y, z) = ||y_{ext} - Rz||_{L_1}. \tag{6}$$

Obtaining $Q(y)$ is then a quantile regression problem. Like spline based smoothing techniques, the smoothing from Lasso also depends on values of (flexibility of the smooth surface in age direction), λ_{xt} (flexibility of the smooth surface in age and time direction) and λ_{tt} (flexibility of the smooth surface in time direction). A subsequent problem to utilize quantile Lasso is to optimize $\lambda_{xx}, \lambda_{xt}$ and λ_{tt}, because the smoothness of the obtained mortality curves depend on these three parameters (also accuracy of Lasso). It is possible to have many local minima while dealing with mortality data. To overcome this problem, Dokumentov et al. (2018) used "Nelder-Mead"optimization[1] to get the global minima of the function. It should be noted that Lasso is not a smoothing technique but a fitting method. However, maximization of $K(y, z)$ is achieved conditioned on optimized values of λ_{xx}, λ_{xt} and λ_{tt} which regulates the smoothness of the regression solution.

[1] "Nelder-Mead"optimization is a simplex based optimization method, which is particularly suited in cases where the object function can have several local minima.

2.3 Estimation of model parameters: adjusting for lifespan disparity

After obtaining the smoothed mortality rates by lasso, we followed the standard LC methodology to obtain the initial estimate of the age and time component from the smoothed mortality data. The two-factor LC model is defined as,

$$\ln m_{x,t} = a_x + b_x k_t + \epsilon_{x,t}. \tag{7}$$

Here, $m_{x,t}$ is the central mortality rate at age x for year t; a_x represents the average of log-mortality at age x over time; b_x is the first principal component capturing relative change in the log-mortality rate at each age x; k_t represents the overall level of mortality in year t; and $e_{x,t}$ is the model residual. The constraints of the model are;

$$\sum_{x=0}^{p} b_x = 1 \text{ and, } \sum_{t=1}^{n} k_t = 0.$$

Our proposed modification on time component begins with estimation of the observed lifespan disparity (obtained from observed, non-smoothed mortality rates). Lifespan disparity illustrates the variation in the life span which is the differences in the length of life across the members of a population. To measure lifespan disparity, we utilize the definition of Vaupel and Canudas-Romo (2003) and Zhang and Vaupel (2009) where it is defined as average number of life years lost at birth. Symbolically,

$$e_0^{\dagger} = \frac{\int_0^{\omega} e_x d_x \ dx}{l_0} \approx \frac{\sum_0^{\omega} e_x l_x m_x}{l_0}; \tag{8}$$

where, ω is the maximum attainable age, d_x is the distribution of death and l_x is the number of people alive at age x (l_0 is the life table radix). Thus estimation of $e_0^{\dagger}$ is simple and straightforward. It can be easily implemented by the above expression with the assumption that deaths are Poisson distributed.

As mentioned before, we followed the same procedure of Lee and Carter (1992) to obtain the initial estimate of the model parameters except for everything is done on smoothed mortality rates. For estimation of the age and time component rank-1 approximation is considered only as it explains most of the variance (Lee and Carter 1992; Lee and Miller 2001). In order to minimize the sum of squares of the residuals, estimator of a_x is obtained by,

$$\hat{a}_x = \frac{1}{n} \sum_{t=t_1}^{t_n} \ln(m_{x,t}).$$

A new matrix $Z_{x,t} = \ln(m_{x,t}) - \hat{a}_x = b_x k_t$ is hence created to obtain the least square estimates of b_x and k_t. Singular Value Decomposition (SVD) is done on $Z_{x,t} = [\ln(m_{x,t}) - \hat{a}_x]$ to obtain the OLS estimate of LC model. SVD decomposes the $Z_{x,t}$ into the product of three matrices. Symbolically,

$$\text{SVD}(Z_{x,t}) = ULV' = L_1 U_{x_1} V_{t_1} + \dots L_n U_{x_n} V_{t_n};$$

from which we obtain the initial estimates of model parameters as,

$$\hat{k}_t = L_1 V_{t_1} \text{ and, } \hat{b}_x = U_{x_1}.$$

Lee and Carter (1992) proposed to conduct a second stage estimate of k_t by finding the value of k_t which, for a given population age distribution and previously estimated a_x and b_x produces exactly the observed number of total deaths for the fitting period of the model. We adjusted of the estimated k_t by solving the following equation:

$$e^{\dagger}_{0\ observed} = \sum_{0}^{\omega} \exp(\hat{a}_x + \hat{b}_x.k_{t\ adj}) e_x l_x / l_0. \tag{9}$$

The e_x and l_x of equation (9) are obtained from life table estimated from fitted $\hat{m}_x$. An *ARIMA(0,1,0)* with drift is then fitted for adjusted $\hat{k}_t$, from which forecast are done. Symbolically,

$$\hat{k}_t = c + \hat{k}_{t-1} + \xi_t; \tag{10}$$

where c is the drift term and ξ_t is the model residual. Previous studies applied different econometric models for forecasting, however, the best output was obtained for Random Walk with Drift (Hyndman and Ullah 2007). Following Lee and Miller (2001), we used the actual data for forecasting to avoid jump-off error; in this way we get more accurate forecast during out-of-sample evaluation and more optimistic forecast than that obtained from fitted data.

2.4 Errors in the mortality forecast

In proposed methodology, it is necessary to incorporate the smoothing error while estimating the forecast variance. Following Lee and Carter (1992), the forecast error for h year ahead forecast from base period t will be,

$$E_{x,t+h} = \alpha_x + (\hat{b}_x + \beta_x) u_{t+h} + \beta_x \hat{k}_{t+h} + \epsilon_{x,t+h} + \epsilon_s. \tag{11}$$

Lee and Carter (1992) defined α_x and β_x as errors in estimating the model parameters a_x and b_x respectively whereas u contains the errors due to innovations and errors in estimating the drift. For the proposed method; α_x, β_x and u_{t+h} will contain the individual level of error due to smoothing as well. We estimated the model parameters without smoothing the data and observed the slight effect of smoothing on trend of estimated parameters (with newly adjusted $\hat{k}_t$). Lee and Carter (1992) mentioned that the elements in equation (11) are uncorrelated from the results obtained by bootstrapping and informal experiments. Same concept is applicable for smoothing because the smoothing errors are very low and it is independently done before model fitting. As the error unexplained from the estimated parameters is attributable to smoothing effect, we added the last term as error from smoothing which is independent from other sources of errors during estimation of the model. This term is different than that of non-parametric approaches as those methods considered the smoothed mortality rates as functional form of age implied through smoothing (Hyndman and Ullah 2007). Consequently Lasso and the proposed adjustment on k_t reduces the overall variance of the fitted model. Hence, the variance of the forecast for modified Lee-Carter will be,

$$\sigma^2_{E_{x,t+h}} = \sigma^2_{\alpha_x} + \hat{b}^2_x \sigma^2_{u_{t+h}} + \sigma^2_{\beta_x}\left(\hat{k}^2_{t+h} + \sigma^2_{u_{t+h}}\right) + \sigma^2_{E_{x,t+h}} + \sigma^2_s. \tag{12}$$

During the estimation, the obtained variance for adjusted $\hat{k}_t$ found to be lower than previous model and smoothing also produced lower error.

2.5 Assessing the performance of the mortality forecasting techniques

To evaluate a forecast technique, we considered two criteria: the accuracy of forecast during out-of-sample evaluation period and how optimistic the forecast is in long run. We considered the following two measures for checking the forecast accuracy of mortality rates:

mean absolute forecast error,

$$\text{MAE} = \frac{1}{(p+1)q}\sum_{j=1}^{q}\sum_{x=0}^{p}\left|y_{x,j} - \widehat{y}_{x,j|j-h}\right|; \tag{13}$$

mean squared forecast error,

$$\text{MSE} = \frac{1}{(p+1)q}\sum_{j=1}^{q}\sum_{x=0}^{p}\left(y_{x,j} - \widehat{y}_{x,j|j-h}\right)^2; \tag{14}$$

and for life expectancy at birth, we considered the mean error of life expectancy,

$$\text{ME} = \frac{1}{q}\sum_{j=1}^{q}\left(e_{0,j} - \hat{e}_{0,j}\right). \tag{15}$$

Here $y_{x,j}$ represents the observed mortality rate for age x in year j and $\widehat{y}_{x,j}$ represents the forecast; $e_{0,j}$ represents the observed life expectancy at birth in year j and $\hat{e}_{0,j}$ represents the forecast. Unlike Shang et al. (2011) or Shang (2012), we choose mean squared forecast error over mean forecast error as measure of forecast accuracy. Mean forecast error of mortality rates could be misleading most of the times as it may conceal forecast inaccuracy due to the offsetting effect of large positive and negative forecast errors on very low error in forecasting. From the available mortality data, we used the data of last 10 years as the hold-out period for forecasting and the previous years as fitting period.

Using the data in the fitting period, we made the one-step-ahead and ten-step-ahead point forecasts, and determine the forecast accuracy by comparing the forecasts with the holdout data in the out-of-sample evaluation period. The analysis performed in this study are implemented by 'Demography' package of R for all of the LC and HU variants (including one dimensional smoothing). For two dimensional smoothing we used 'MortalitySmooth' package and for lasso we used the 'smoothAPC' package in R. The proposed method is implemented by modifying the existing function in 'Demography' and are given in appendix.

3 Results

3.1 Smoothing techniques

Since we smoothed the mortality rates first before model fitting, we compared the smoothing techniques before checking the model fitting. We compared the smoothed mortality rates obtained from Lasso with one dimensional spline (Hyndman and Ullah 2007) and two dimensional spline (Camarda et al. 2012). Accuracy of smoothing techniques has been evaluated in terms of lowest mean absolute error (MAE) and mean squared error (MSE). We did not consider BIC or other measures of goodness of fit as BIC requires a clearly defined likelihood function. Beside, this comparison on smoothing is performed considering only one necessary criterion useful for forecasting purpose. We wanted to identify the technique producing smoothed mortality surface with lowest error (with respect to observed data) considering lowest degrees of smoothing. We compared the results for all the low-mortality countries mentioned in Section 3 and Lasso provided most accurate mortality curves (with lowest errors). For illustration, the smoothed mortality rates from all three techniques for US Females are given in Figure 2 and the mortality surfaces are plotted in 3.

The mortality trends are regular for almost all of these low-mortality countries since the last half of twentieth century, as a result the smoothing techniques also performed well for these countries. We considered USA for this illustration because (i) it represents common natures of almost all aging societies (low infant mortality with similar pattern of rising trend in senescence mortality); (ii) presence of distinct accidental hump and (iii) distinctly high centenarian mortality rates in early 1950s. Thus, a smoothing technique preserving all of this characteristics will be preferable over the others in term of accuracy. To date, USA is one of the most populous countries having the lowest observed life expectancy at birth among the G-7 countries (HMD 2018), which is highly attributable to particular cause-specific distribution of deaths (Tuljapurkar et al. 2000). Compare to spline based smoothing techniques, the mortality surface is less smoothed for Lasso (Figure 2 and 3). Same pattern is observed for other lower mortality countries also. For US Females (and many other countries), an unexplainable second drop of mortality is visible in earlier

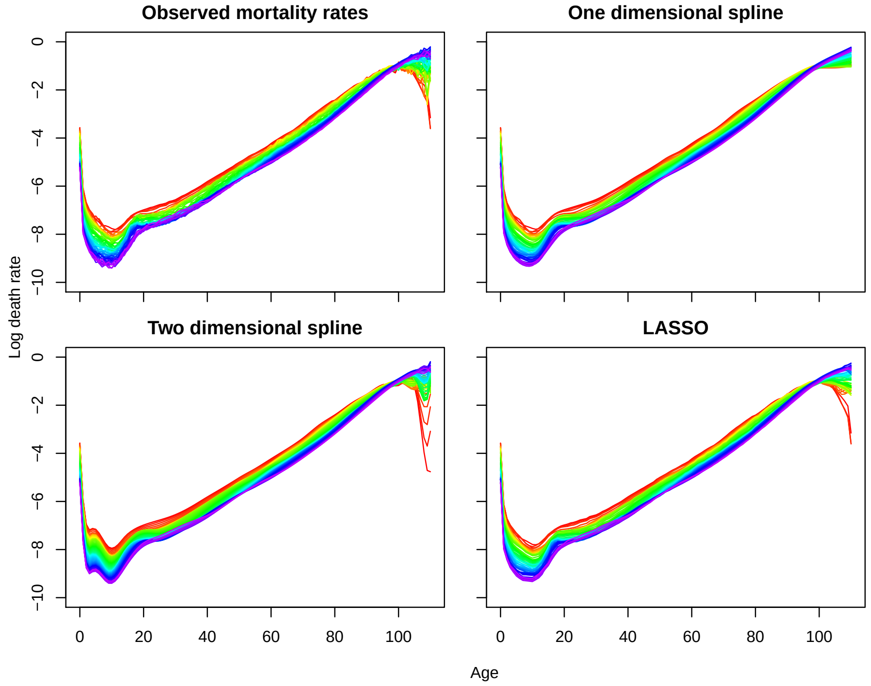


Fig. 2 Comparison of smoothing techniques for US Female mortality (1950:2016). Years are plotted using a rainbow palette so the earlier years are shown in red, followed by orange, yellow, green, blue and indigo with the most recent years plotted in violet.

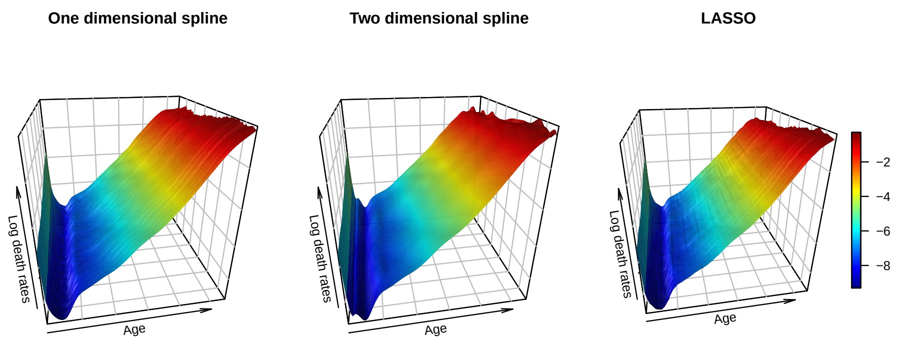


Fig. 3 Mortality surfaces from difference smoothing techniques for US Females (1950:2016).

life span in case of two dimensional spline (Figure 2). Previous studies mentioned that smoothing with two dimensional spline could be misleading for earlier part of lifespan as it might produce biased result for that (Dokumentov et al. 2018). To avoid this bias from infant and child mortality, starting the smoothing from age 10 or later could be a solution. However, this is not very relevant for mortality forecasting. The transformed error for all three smoothing techniques are illustrated in Figure 4 for the US Females.

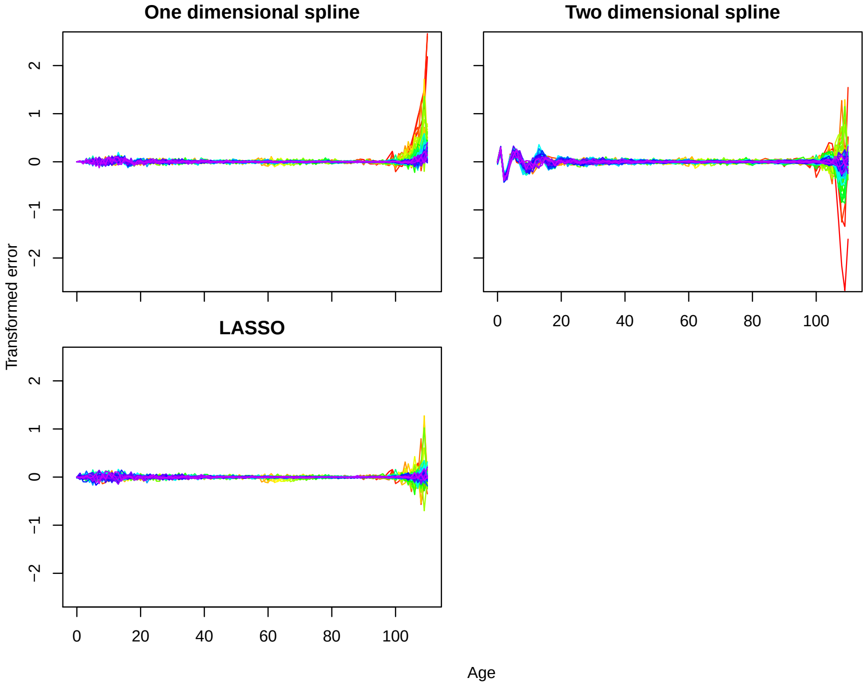


Fig. 4 Errors of smoothing techniques for female mortality of USA(1950:2016). Years are plotted using a rainbow palette as before.

Unlike two dimensional smoothing, Lasso and one dimensional smoothing technique performed well for earlier part of life. Presence of accidental hump is well-captured by Lasso compared to spline based techniques. Spline based smoothing are less affective for older ages. One dimensional smoothing could not capture the high centenarian mortality in beginning of 1950s, whereas two dimensional smoothing over-fitted that part. Better accuracy for Lasso can be explained from the context of smoothing. Lasso is more a fitting technique with more emphasizing on keeping the fitted values close to observed one. As a result, although the surface is less smoothed for Lasso, still it provides less error for fitting the mortality curves; which is not the case for spline based smoothing techniques. The accuracy of smoothing for US female mortality by these three techniques are summarized in Table 1. Summary for other countries are attached in Appendix for readers. The errors for mortality rates are magnified by 100 times to show the comparison more precisely.

Table 1 Accuracy of smoothing techniques for US Females (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	3.170	4.817	1.909
MSE(m_x)×100	0.953	1.162	0.215
ME(e_0)	-0.154	0.139	0.950
MAE(e_0)	0.002	0.087	0.016

For both absolute and squared error, lasso is more accurate than other two smoothing techniques. In addition, we reconstruct the life tables from smoothed mortality rates and from that we compared the fitted life expectancy at birth with observed one. Two dimensional smoothing has the lowest ME(e_0), although, looking more closely at age-specific pattern, error is high at earlier and later age groups–and they offset each other. From ME(e_0) reported in table 1 we also notice that LASSO produces more optimistic forecasts whereas one dimensional spline more pessimistic ones. Due to different pattern of outcome obtained from ME(e_0), we further examined the mean absolute error for life expectancy for more insight. The results showed highest accuracy for one dimensional smoothing followed by lasso and two dimensional smoothing. We did not consider MAE(e_0) for forecasting as it does not provide any additional information than that of ME(e_0) in case of forecasting.

3.2 Model fitting and forecast accuracy

We compared the findings of the proposed methodology with 4 LC variants and 3 non-parametric variants for all of the countries mentioned in Section 3. We denoted the proposed method by $\text{LC}_{e_0^\dagger}$ all over this section. For previous existing models, LC stands for the basic Lee and Carter (1992); LC_P stands for LC with Poisson regression without any adjustment for k_t (Brouhns et al. 2002); LM stands for modified LC proposed by Lee and Miller (2001) which adjust k_t according to life expectancy; BMS stands for modified LC model proposed by Booth et al. (2002) which adjust k_t according to observed age at death distribution and consider best fitting period; HU stands for the non-parametric approach proposed by Hyndman and Ullah (2007), HU_R stands for robust Hyndman and Ullah (2007); HU_W stands for weighted Hyndman and Ullah (2007). Among these variants, HU variants consider smoothed mortality rate for forecasting, where the mortality rates are smoothed using constraint regression splines (Hyndman and Ullah 2007). HU_R and HU_W are two different weighted versions of HU to overcome problem of outliers and giving more weight on recent years respectively. All of these existing methods use fitted data for mortality forecasting except for LM, which considers actual data. As non-parametric approaches are different than those of LC variants, we compared the individual parameters of $LC_{e_0^\dagger}$ with LC variants only. For illustration of the characteristics of the proposed methodology, we consider the Swedish female mortality in this section. We considered Sweden for consistent and stable trend of longevity since long. Beside, Sweden is well known for mortality data with highest quality (HMD 2018). The estimated a_x and b_x from LC variants and $LC_{e_0^\dagger}$ are plotted in figure 5. These two parameters are the same for previous LC variants due to same estimation technique.

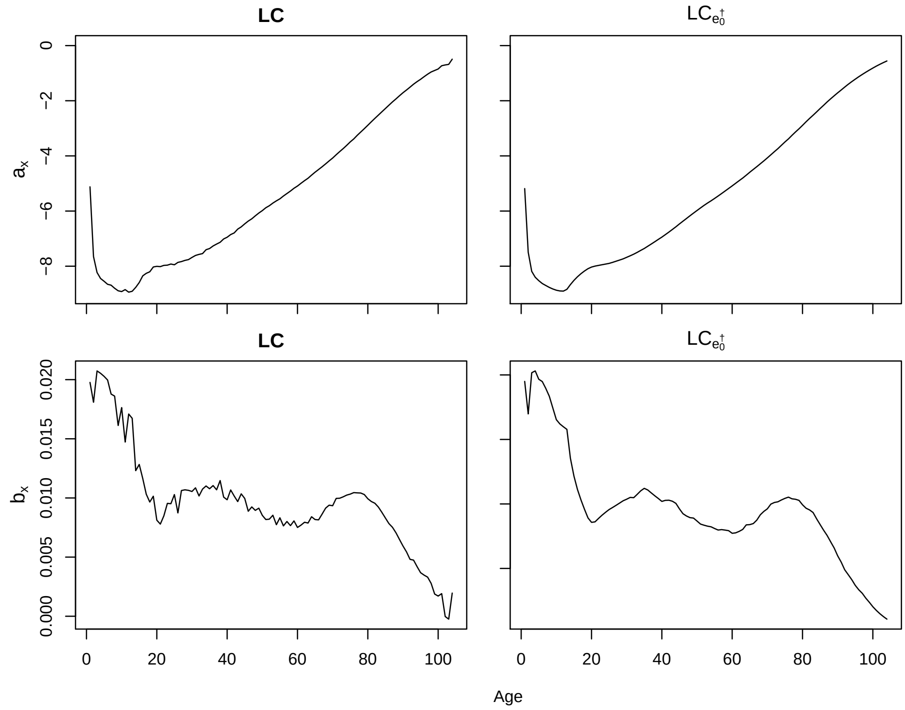


Fig. 5 Estimated a_x and b_x for Swedish Females (1950:2016) from LC variants and $LC_{e_0^\dagger}$. The estimated parameters from $LC_{e_0^\dagger}$ slightly different than that of LC variants due to application of smoothing technique prior to model fitting.

Clearly, the estimates and forecasts of k_t are different for all of these variants due to different adjustment methods and jump-off policy. Sweden has long and steady trend of mortality improvement which is also reflected after adjusting for lifespan disparity. The newly adjusted $\hat{k}_t$ has more regular and linear trend than estimates from earlier Lee-Carter methods. The trend of adjusted $\hat{k}_t$ along with forecast till 2050 by an *ARIMA (0,1,0)* for Swedish female is illustrated in Figure 6. The regularity of $\hat{k}_t$ in $\text{LC}_{e_0^\dagger}$ is attributable to both smoothing technique and new partial adjustment policy. Compare to the previous LC variants, the proposed $\hat{k}_t$ have less standard error which also consequently creates narrower confidence bound for forecast intervals later.

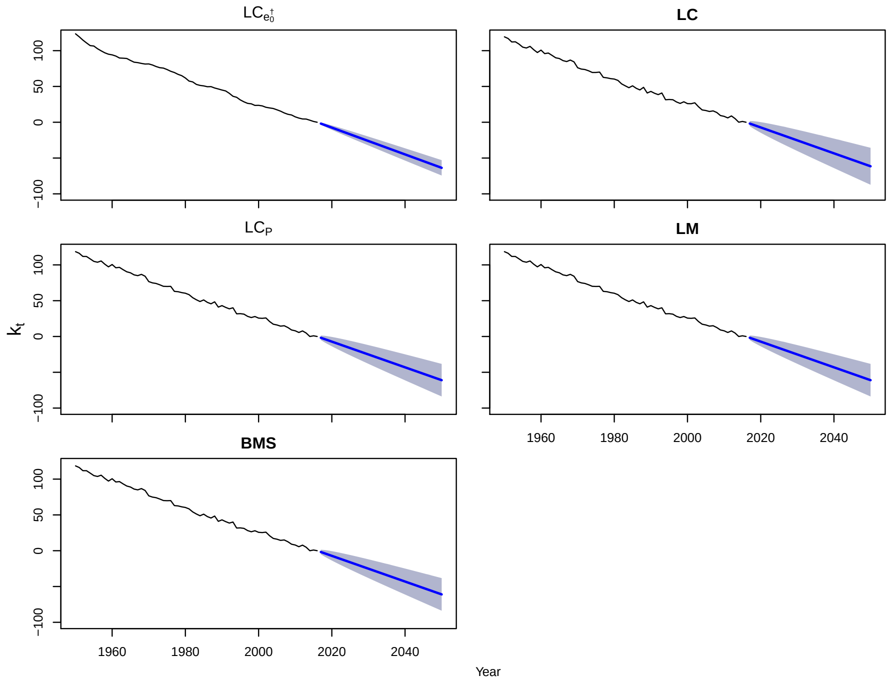
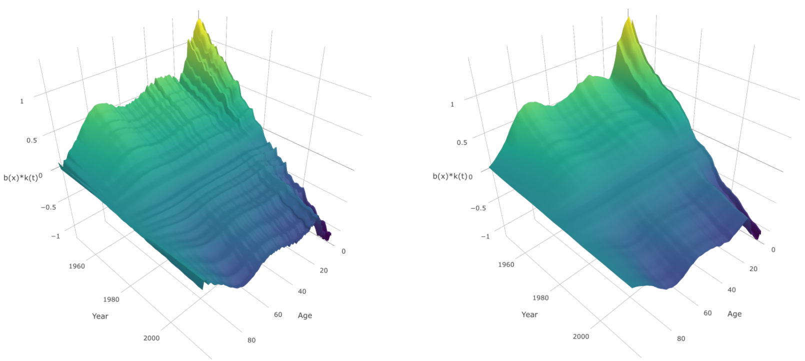


Fig. 6 Random walk with drift for estimated k_t of Swedish Females (1950:2016) for LC variants and $\text{LC}_{e_0^\dagger}$.

For Swedish female mortality, we used the life tables constructed up to age 103 years due to missing values at older ages (HMD 2018). For the trend of k_t shown in Figure 6, the model was fitted in the years 1950 to 2016 (the last available year during this analysis). Although we plotted the trend for 1950 to 2016, the best fitting period obtained for BMS was from 1978 to 2016. The product of $\hat{b}_x$ (obtained from smoothed mortality rates) with more regular $\hat{k}_t$ ultimately produce more smoothed surface than that observed for LC (Figure 1). The other LC variants also show almost similar surface like basic LC (Lee and Carter 1992). The new smooth surface of the product of $\hat{b}_x$ and $\hat{k}_t$ are given in Figure 7. For Swedish females, LC can explain 77.4% of observed variation whereas both $\text{LC}_{e_0^\dagger}$ and HU variants can explain 97.7% of that. Variance explained by the fitted models are given in Appendix. It should be noted that unlike rank-1 approximation of all LC variants and $\text{LC}_{e_0^\dagger}$, HU is estimated with 3 components (Hyndman and Ullah 2007).



(a) LC (b) $\mathrm{LC}_{e_0^\dagger}$

Fig. 7 Product of time component and age component for Swedish Females (1950:2016) using (a) LC (Lee and Carter 1992) and (b) proposed $\mathrm{LC}_{e_0^\dagger}$. (a) is illustrated with larger interface in Fig. 1.

Next, we compared the forecast accuracy of the proposed $\mathrm{LC}_{e_0^\dagger}$ with existing LC and HU variants. For out-of-sample evaluation of the methods, we fitted the model for all the available time period except for holding the last 10 years to compare with the forecast obtained by fitted model with respect to observed mortality rates. The forecast accuracy of all 8 models ($\mathrm{LC}_{e_0^\dagger}$, 4 LC variants and 3 HU variants) for the females of 20 low-mortality countries are illustrated in Figure 8, 9 and 10. Figure 8 showed the accuracy in terms of mean absolute errors (MAE) for mortality rates, Figure 9 is for mean squared errors (MSE) of mortality rates and mean error of obtained life expectancy at birth ($\mathrm{ME}(e_0)$) are presented in Figure 10. The detailed values are given in Appendix. For MAE and MSE, country-specific highest forecast accuracy for mortality rates can be identified from the points close to the x axis. Exceptionally high MAE and MSE are observed for Ireland and Japan in case of LC and $\mathrm{LC_P}$, whereas the other methods were almost close for them. We overlooked the results of $\mathrm{HU_W}$ for US Females for forecast accuracies as it showed decreasing trend of life expectancy forecast for US Females. The forecast of US female life expectancy was 80.88 years in 2050 for $\mathrm{HU_W}$, whereas the last observed e_0 was 81.38 years in 2016 (HMD 2018).

Figure 8 and 9 show that $\mathrm{LC}_{e_0^\dagger}$ together with (HU, $\mathrm{HU_R}$ and $\mathrm{HU_W}$ for some countries) the smallest- on average, forecast errors. However, Figure 10 reveals that HU and $\mathrm{HU_R}$ are method to have negative mean error of e_0, meaning that

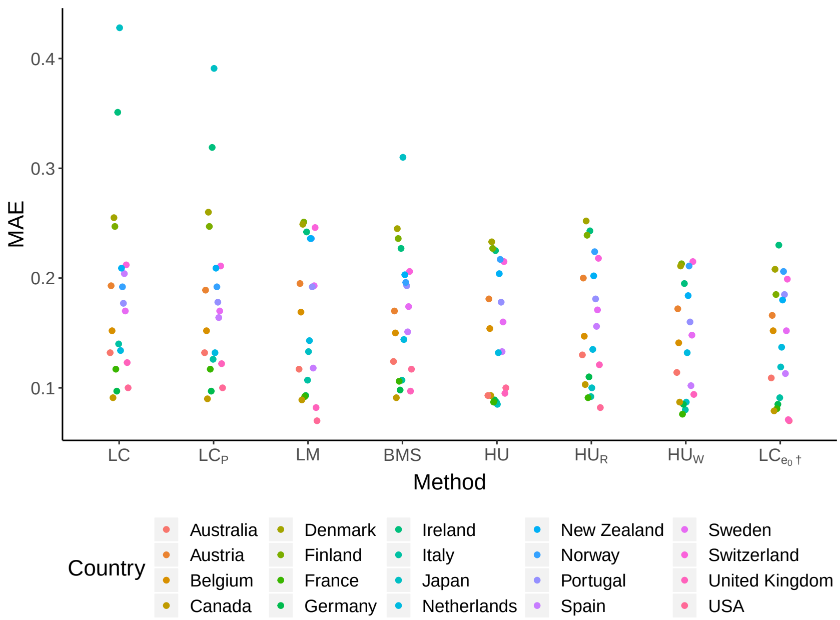


Fig. 8 MAE during out-of-sample evaluation period from different methods.

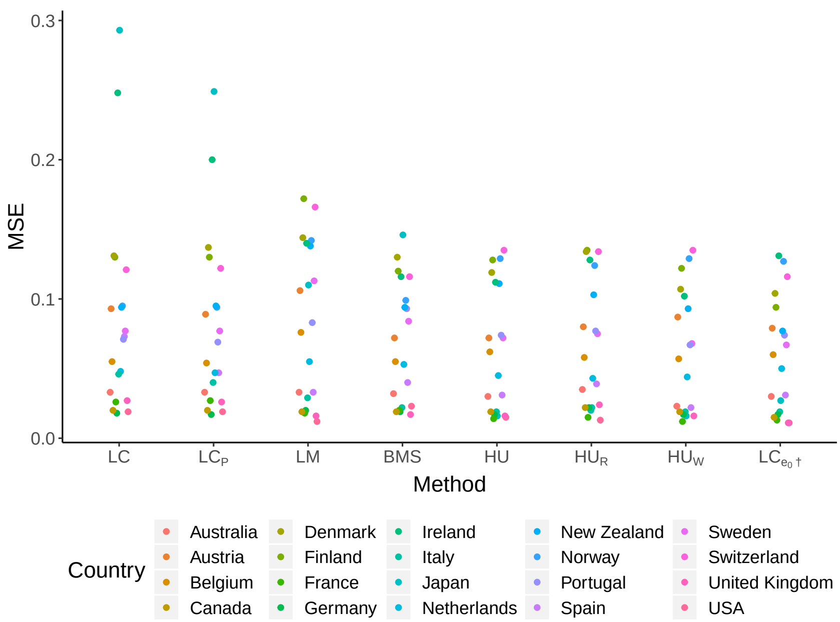


Fig. 9 MSE during out-of-sample evaluation period from different methods.

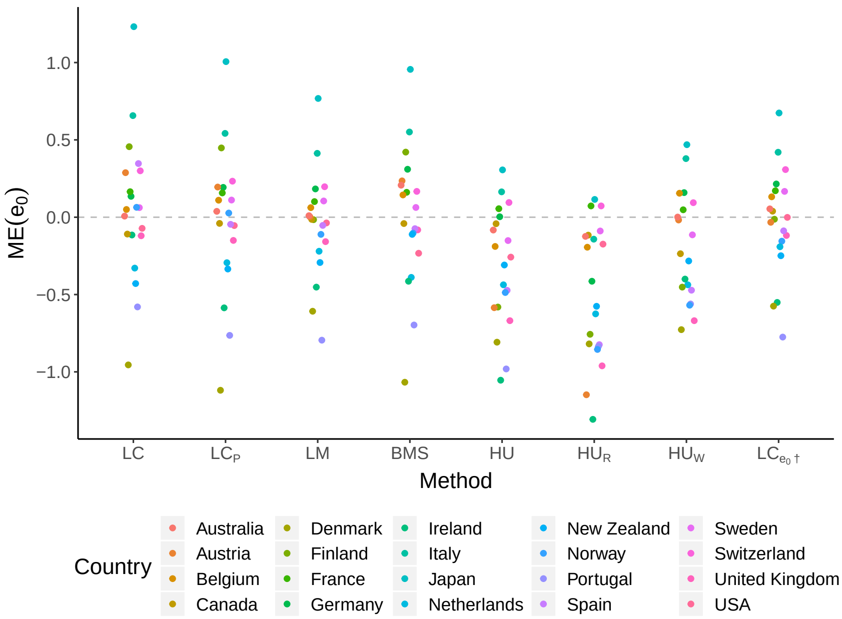


Fig. 10 $\mathrm{ME}(e_0)$ during out-of-sample evaluation period from different methods. The gray dashed line represents null error.

forecast of life expectancy is, on average, lower that the observed one. $\mathrm{LC}_{e_0^\dagger}$ and $\mathrm{HU_W}$ are more symmetric around null error, so they are more balanced for providing optimistic and pessimistic forecast of life expectancies. Previous studies also mentioned that none of the methods were uniquely best for all countries (Shang 2012).

Comparatively better performance of $\mathrm{LC}_{e_0^\dagger}$ are attributable to both of the proposed modifications. Beside application of Lasso as a more accurate smoothing technique, adjustment of time component according to lifespan disparity reflects more insight of the mortality trend of a population rather than total number of deaths or life expectancy at birth as it contains information of both remaining life expectancies and corresponding distribution of death.

3.3 Forecast of life expectancy

We compared the forecast of life expectancy at birth in 2050 obtained from all of these forecasting techniques for the low-mortality countries. The forecast of life expectancy at birth till 2050 for these low-mortality countries are summarized in Table 2. Country-specific highest forecast are marked using bold texts. $LC_{e_0^\dagger}$ was the forecasting technique producing the most optimistic forecasts of life expectancy at birth for 6 countries followed by BMS for 5 countries.

Table 2 Forecast of female life expectancy at birth in 2050 for 20 low mortality countries.

Country	LC	LC_P	LM	BMS	HU	HU_R	HU_W	$LC_{e_0^\dagger}$
Australia	89.24	89.26	89.24	**89.65**	89.28	88.66	89.42	89.43
Austria	88.95	88.88	88.92	**89.65**	88.56	89.07	88.51	88.95
Belgium	87.83	87.90	88.02	87.94	**88.42**	88.30	88.31	88.25
Canada	88.86	88.79	88.78	88.62	88.39	88.47	**89.19**	88.82
Denmark	86.81	86.78	**86.95**	86.51	85.91	86.93	86.21	86.75
Finland	89.25	89.36	89.51	88.96	88.01	88.43	89.22	**89.99**
France	90.26	90.34	90.45	90.54	90.61	90.55	90.13	**90.74**
Germany	87.59	87.76	87.74	**88.37**	87.87	87.28	87.53	87.82
Ireland	88.23	87.88	88.36	88.79	89.75	88.61	**89.82**	88.05
Italy	90.26	90.28	90.30	**90.80**	90.40	90.12	90.30	90.42
Japan	**93.72**	93.71	93.67	93.60	91.29	90.57	90.07	93.63
Netherlands	86.87	86.86	86.93	86.86	86.95	86.90	86.96	**87.22**
New Zealand	87.96	87.92	88.05	**88.65**	87.93	87.95	87.93	88.15
Norway	88.04	87.98	88.14	88.11	88.42	88.50	88.42	**88.55**
Portugal	88.41	88.50	88.43	89.40	86.24	**90.22**	86.24	88.60
Spain	90.66	90.27	90.57	90.22	89.93	89.81	**91.47**	90.53
Sweden	88.19	88.15	88.23	87.79	87.95	87.87	88.09	**88.38**
Switzerland	89.83	89.73	89.87	89.56	89.87	88.85	89.21	**90.06**
United Kingdom	87.12	87.16	87.16	87.54	**88.06**	87.13	87.67	87.36
USA	**85.68**	85.42	85.30	84.66	84.08	84.15	81.38	85.12

To illustrate the forecast of life expectancy at birth more clearly, the prediction interval of e_0 till 2050 are plotted in Figure 11 for Swedish Females.

Although, HU_W was more accurate than $LC_{e_0^\dagger}$ for Swedish Females during out-of-sample evaluation, the mortality forecast for Swedish Females were more optimistic from $LC_{e_0^\dagger}$ than HU_W (Table 2). Due to different weighting techniques used in HU variants, the forecasts are also different for each of them. For mortality forecast only $LC_{e_0^\dagger}$ and LM consider actual data over fitted one. Due to lack of smoothing and different adjusting of time component, the forecast of LM have more jagged pattern than any other methods.

To compare the interval forecast of e_0, we applied the existing semi-parametric bootstrapping technique proposed by Hyndman and Booth (2008). This technique was mainly designed for non-parametric forecasting obtained from functional data analysis (Hyndman and Ullah 2007). Following Hyndman and Booth (2008), simulated forecasts of log-mortality rates are obtained by adding disturbances to the forecast of $\hat{k}_t$ which are then multiplied by the fixed age component $\hat{b}_x$. Hence the life expectancies are estimated for each set of simulated log-mortality rates. Prediction intervals are then constructed from the

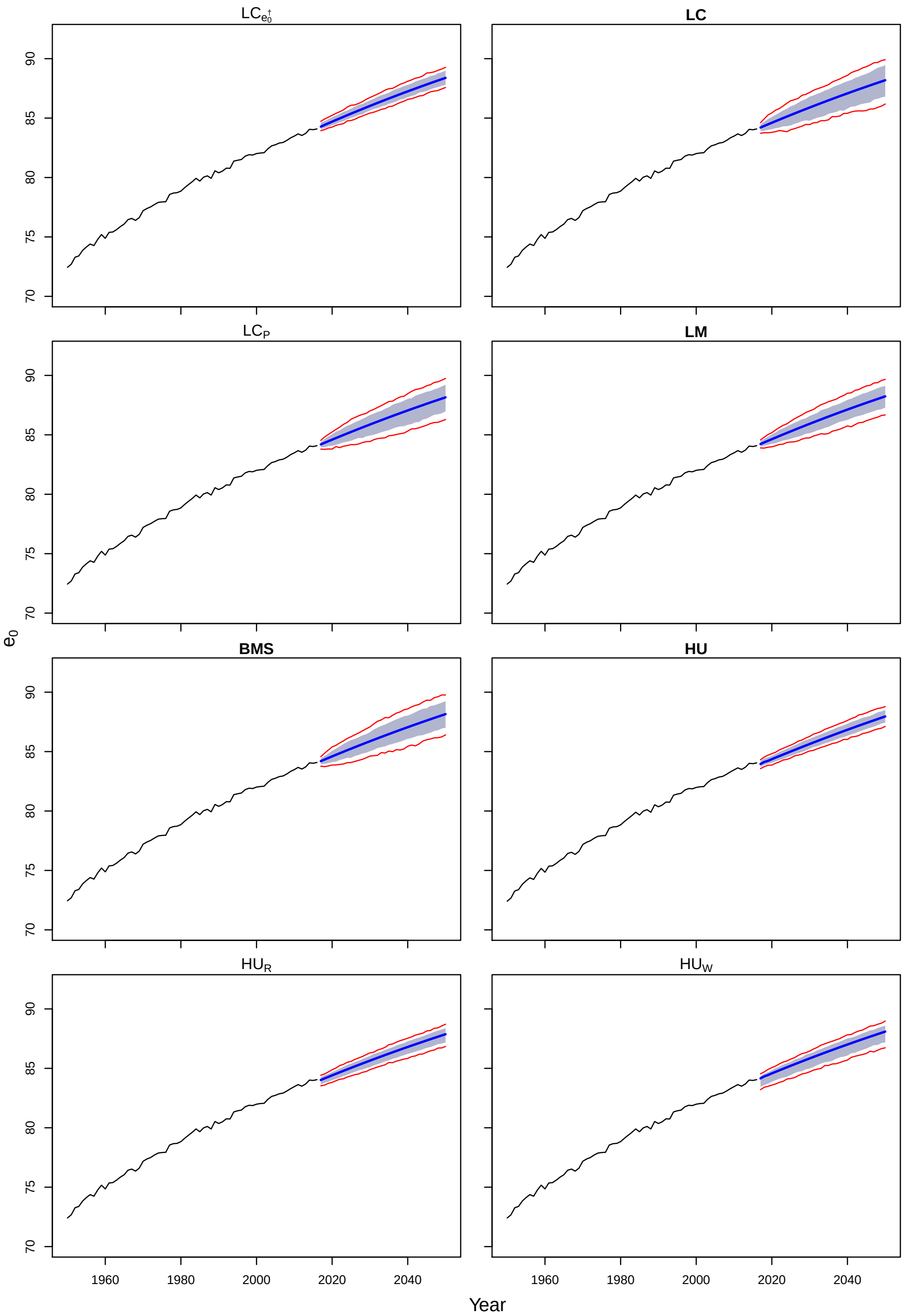


Fig. 11 Prediction interval of Swedish female e_0 by different methods till 2050. The blue area represents 80% prediction interval and red lines are for 95% prediction interval.

percentiles of the simulated life expectancies. Due to lower standard error obtained for newly adjusted $\hat{k}_t$, the interval forecast obtained from $\mathrm{LC}_{e_0^\dagger}$ are narrower than previous LC variants. Previous studies mentioned this problem (i.e., the prediction intervals of the LC type models are too narrow) which may lead to underestimate the coverage probability (see Lee and Carter 1992; Shang 2012, for example). Interval forecasting thus remained as a common issue for LC type models that has room for improvement.

4 Conclusion

We introduced the application of lifespan disparity in ground of mortality forecasting in this article. The concept is adopted along with smoothing on widely used mortality forecasting technique of Lee and Carter (1992). We utilized lasso type regularization instead of contemporary spline based smoothing techniques in this approach. Lasso can effectively adjust the irregular pattern of mortality improvement over time but still produces lower error while smoothing the observed data. We partially adjust the time component according to lifespan inequality which gives more information about longevity and mortality transition of a population. We compared the proposed method with different variants of Lee-Carter method, including variants incorporating smoothing as well. The proposed methodology generates more accurate forecast than most of the existing variants for many countries. Some of these methods are often criticized for pessimistic forecast of life expectancy during out-of-sample evaluation, whereas the proposed method performed reasonably well in this context. In addition, more optimistic forecasts were obtained from the proposed method for several countries compare to the existing methods.

Although the results are promising for the proposed methodology, there are still some issues that deserve further investigation. For starter, interval forecast of life expectancy at birth is narrow for several of the populations in case of $\mathrm{LC}_{e_0^\dagger}$. This is an old criticism regarding Lee-Carter type models (Hyndman and Booth 2008). In the proposed model it happened due to application of smoothing and new adjustment technique which made the time component more regular than the previous approaches. As a consequence it reduced the variance of the ARIMA model. In addition, variance of the model is lower in the proposed method, which also affects the interval forecast. Adding more components may provide a solve in this context, however, it is not always unique (see Figure 11 for an example). Thus, it remains as a trade-off between accuracy and width of interval forecast. Secondly, we considered only rank-1 approximation for model fitting and obtained almost similar level of explained variation as of the higher order approximation. It should be noted that, the higher order terms are also not free from invariance problem and thus those model also contain constraints for parameter estimation (see Hyndman and Ullah 2007, for example). Another limitation of the current study is that we could not consider the findings with many existing LC variants. Among them we did not consider Li et al. (2013) as many populations do not fulfill the primary condition of age rotation. Also, we did not compare the models where smoothing are applied in different fashion than our one: smoothing the estimated parameters or smoothing the fitted mortality rates.

From the findings of current study, we may suggest some future scope of further development in Lee-Carter framework. The first issue will be to overcome to old problem of invariant b_x in Lee-carter framework. One possible solution to do that is to adopt Bayesian approaches on parameter estimation. We proposed the method for single population, extension may be useful for coherent mortality forecasting as well. Besides, we did not consider the period and cohort effect separately during smoothing and model fitting; controlling the cohort effect may produce more accurate forecast for some countries. We considered the simplest definition of lifespan disparity for adjusting the time component, other definitions or other measures of lifespan disparity may provide different results. Due to availability of mortality data with good quality, only low mortality countries are considered in this study. Application on different mortality regime than that used in this study may provide more insight in this line of research as well.

Appendix

Table 3 Accuracy of smoothing techniques for Females of Australia (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	7.475	8.888	5.817
MSE(m_x)×100	1.524	2.246	1.068
ME(e_0)	-0.020	-0.003	0.005
MAE(e_0)	0.020	0.144	0.036

Table 4 Accuracy of smoothing techniques for Females of Austria (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	9.100	10.177	8.862
MSE(m_x)×100	2.784	3.070	2.479
ME(e_0)	-0.025	0.001	0.027
MAE(e_0)	0.025	0.119	0.102

Table 5 Accuracy of smoothing techniques for Females of Belgium (1950:2015).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	9.012	10.106	5.727
MSE(m_x)×100	2.582	2.874	1.453
ME(e_0)	-0.020	0.004	0.018
MAE(e_0)	0.020	0.132	0.031

Table 6 Accuracy of smoothing techniques for Females of Canada (1950:2011).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	5.911	7.269	6.080
MSE(m_x)×100	0.985	1.402	0.926
ME(e_0)	-0.011	-0.001	0.017
MAE(e_0)	0.011	0.084	0.086

Table 7 Accuracy of smoothing techniques for Females of Denmark (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	12.208	12.804	12.018
MSE(m_x)×100	4.680	4.528	4.194
ME(e_0)	-0.045	-0.002	0.010
MAE(e_0)	0.045	0.121	0.109

Table 8 Accuracy of smoothing techniques for Females of Finland (1950:2015).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	12.367	13.062	12.310
MSE(m_x)×100	4.746	4.809	4.368
ME(e_0)	-0.048	-0.002	0.060
MAE(e_0)	0.048	0.127	0.138

Table 9 Accuracy of smoothing techniques for Females of France (1950:2015).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	3.835	5.422	2.722
MSE(m_x)×100	0.523	0.908	0.314
ME(e_0)	-0.002	-0.001	-0.006
MAE(e_0)	0.003	0.145	0.016

Table 10 Accuracy of smoothing techniques for Females of Germany (1956:2015).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	4.148	5.551	2.083
MSE(m_x)×100	0.573	0.955	0.189
ME(e_0)	-0.012	-0.015	0.007
MAE(e_0)	0.011	0.103	0.014

Table 11 Accuracy of smoothing techniques for Females of Ireland (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	15.618	16.033	12.685
MSE(m_x)×100	6.114	5.876	4.211
ME(e_0)	-0.110	-0.006	0.078
MAE(e_0)	0.110	0.177	0.168

Table 12 Accuracy of smoothing techniques for Females of Italy (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	4.260	6.072	3.910
MSE(m_x)×100	0.682	1.184	0.515
ME(e_0)	-0.003	-0.001	0.010
MAE(e_0)	0.003	0.169	0.086

Table 13 Accuracy of smoothing techniques for Females of Japan (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	3.411	4.912	2.853
MSE(m_x)×100	0.601	0.897	0.425
ME(e_0)	-0.003	-0.005	0.005
MAE(e_0)	0.003	0.149	0.017

Table 14 Accuracy of smoothing techniques for Females of Netherlands (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	6.220	7.560	3.712
MSE(m_x)×100	1.335	1.646	0.725
ME(e_0)	-0.010	-0.001	0.003
MAE(e_0)	0.011	0.102	0.047

Table 15 Accuracy of smoothing techniques for Females of New Zealand (1950:2013).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	15.052	15.656	14.523
MSE(m_x)×100	5.881	5.742	4.981
ME(e_0)	-0.082	-0.004	0.001
MAE(e_0)	0.082	0.181	0.178

Table 16 Accuracy of smoothing techniques for Females of Norway (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	11.801	12.976	11.541
MSE(m_x)×100	4.687	4.663	4.045
ME(e_0)	-0.040	-0.002	0.035
MAE(e_0)	0.040	0.106	0.078

Table 17 Accuracy of smoothing techniques for Females of Portugal (1950:2015).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	7.908	9.029	5.400
MSE(m_x)×100	1.853	2.124	1.020
ME(e_0)	-0.024	0.008	0.035
MAE(e_0)	0.023	0.279	0.061

Table 18 Accuracy of smoothing techniques for Females of Spain (1950:2014).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	7.095	7.990	5.523
MSE(m_x)×100	2.837	1.979	1.022
ME(e_0)	-0.018	-0.003	0.057
MAE(e_0)	0.018	0.203	0.167

Table 19 Accuracy of smoothing techniques for Females of Sweden (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	10.408	11.281	10.716
MSE(m_x)×100	3.479	3.585	3.354
ME(e_0)	-0.029	-0.001	0.060
MAE(e_0)	0.029	0.097	0.102

Table 20 Accuracy of smoothing techniques for Females of Switzerland (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	11.088	11.794	10.051
MSE(m_x)×100	4.102	4.028	3.269
ME(e_0)	-0.038	-0.002	0.015
MAE(e_0)	0.038	0.119	0.041

Table 21 Accuracy of smoothing techniques for Females of United Kingdom (1950:2016).

Measure of accuracy	1 dim Smoothing	2 dim Smoothing	*LASSO*
MAE(m_x)×100	4.169	5.706	4.403
MSE(m_x)×100	0.531	0.948	0.544
ME(e_0)	-0.007	-0.006	0.006
MAE(e_0)	0.007	0.124	0.111

Table 22 Variance explained by the fitted models for the low mortality countries.

Country	LC	LC_P	LM	BMS	HU	HU_R	HU_W	$LC_{e_0^\dagger}$
Australia	0.890	0.890	0.890	0.795	0.986	0.880	0.994	0.956
Austria	0.854	0.854	0.854	0.542	0.983	0.836	0.993	0.975
Belgium	0.831	0.831	0.831	0.549	0.983	0.813	0.992	0.922
Canada	0.901	0.901	0.901	0.888	0.988	0.889	0.994	0.980
Denmark	0.681	0.681	0.681	0.644	0.971	0.640	0.989	0.942
Finland	0.774	0.774	0.774	0.355	0.977	0.748	0.990	0.958
France	0.939	0.939	0.939	0.938	0.992	0.956	0.996	0.960
Germany	0.957	0.957	0.957	0.934	0.983	0.952	0.997	0.972
Ireland	0.735	0.735	0.735	0.560	0.984	0.703	0.993	0.909
Italy	0.955	0.955	0.955	0.925	0.994	0.957	0.996	0.977
Japan	0.955	0.955	0.955	0.955	0.997	0.976	0.996	0.965
Netherlands	0.870	0.870	0.870	0.870	0.983	0.874	0.983	0.929
New Zealand	0.639	0.639	0.639	0.447	0.959	0.548	0.959	0.942
Norway	0.654	0.654	0.654	0.349	0.943	0.600	0.943	0.906
Portugal	0.905	0.905	0.905	0.852	0.992	0.906	0.992	0.946
Spain	0.915	0.915	0.915	0.910	0.991	0.886	0.996	0.945
Sweden	0.774	0.774	0.774	0.774	0.977	0.712	0.990	0.977
Switzerland	0.791	0.791	0.791	0.613	0.985	0.778	0.994	0.942
United Kingdom	0.920	0.920	0.920	0.888	0.987	0.915	0.995	0.965
USA	0.916	0.916	0.916	0.863	0.985	0.862	0.993	0.941

Note: Country-specific available years and age-groups are considered for model fitting.

Table 23 MAE of the forecast methods during out-of-sample evaluation period.

Country	LC	LC_P	LM	BMS	HU	HU_R	HU_W	$LC_{e_0^\dagger}$
Australia	0.132	0.132	0.117	0.124	**0.093**	0.130	0.114	0.109
Austria	0.193	0.189	0.195	0.170	0.181	0.200	0.172	**0.166**
Belgium	0.152	0.152	0.169	0.150	0.154	0.147	**0.141**	0.152
Canada	0.091	0.090	0.089	0.091	0.093	0.103	0.087	**0.079**
Denmark	0.255	0.260	0.249	0.245	0.233	0.252	0.211	**0.208**
Finland	0.247	0.247	0.251	0.236	0.227	0.239	0.213	**0.185**
France	0.117	0.117	0.092	0.106	0.087	0.091	**0.076**	0.081
Germany	0.097	0.097	0.093	0.098	0.089	0.110	**0.085**	**0.085**
Ireland	0.351	0.319	0.242	0.227	0.225	0.243	**0.195**	0.230
Italy	0.140	0.126	0.107	0.107	0.087	0.092	**0.080**	0.091
Japan	0.428	0.391	0.133	0.310	**0.085**	0.100	0.087	0.119
Netherlands	0.134	**0.132**	0.143	0.144	0.132	0.135	**0.132**	0.137
New Zealand	0.209	0.209	0.236	0.203	0.204	0.202	0.184	**0.180**
Norway	**0.192**	**0.192**	0.236	0.196	0.217	0.224	0.211	0.206
Portugal	**0.177**	0.178	0.192	0.193	0.178	0.181	**0.160**	0.185
Spain	0.204	0.164	0.118	0.151	0.133	0.156	**0.102**	0.113
Sweden	0.170	0.170	0.193	0.174	0.160	0.171	**0.148**	0.152
Switzerland	0.212	0.211	0.246	0.206	0.215	0.218	0.215	**0.199**
United Kingdom	0.123	0.122	0.082	0.097	0.095	0.121	0.094	**0.071**
USA	0.100	0.100	**0.070**	0.117	0.100	0.082	-	**0.070**

Table 24 MSE of the forecast methods during out-of-sample evaluation period.

Country	LC	LC_P	LM	BMS	HU	HU_R	HU_W	$LC_{e_0^\dagger}$
Australia	0.033	0.033	0.033	0.032	0.030	0.035	**0.023**	0.030
Austria	0.093	0.089	0.106	**0.072**	**0.072**	0.080	0.087	0.079
Belgium	0.055	**0.054**	0.076	0.055	0.062	0.058	0.057	0.060
Canada	0.020	0.020	0.019	0.019	0.019	0.022	0.019	**0.015**
Denmark	0.131	0.137	0.144	0.130	0.119	0.134	0.107	**0.104**
Finland	0.130	0.130	0.172	0.120	0.128	0.135	0.122	**0.094**
France	0.026	0.027	0.018	0.020	0.014	0.015	**0.012**	0.013
Germany	0.018	**0.017**	0.020	0.019	**0.017**	0.022	**0.017**	**0.017**
Ireland	0.248	0.200	0.140	0.116	0.112	0.128	**0.102**	0.131
Italy	0.046	0.040	0.029	0.022	**0.019**	0.020	**0.019**	**0.019**
Japan	0.293	0.249	0.110	0.146	**0.016**	0.022	**0.016**	0.027
Netherlands	0.048	0.047	0.055	0.053	0.045	**0.043**	0.044	0.050
New Zealand	0.094	0.095	0.138	0.094	0.111	0.103	0.093	**0.077**
Norway	0.095	**0.094**	0.142	0.099	0.129	0.124	0.129	0.127
Portugal	0.071	0.069	0.083	0.093	0.074	0.077	**0.067**	0.074
Spain	0.073	0.047	0.033	0.040	0.031	0.039	**0.022**	0.031
Sweden	0.077	0.077	0.113	0.084	0.072	0.075	0.068	**0.067**
Switzerland	0.121	0.122	0.166	**0.116**	0.135	0.134	0.135	**0.116**
United Kingdom	0.027	0.026	0.016	0.017	0.016	0.024	0.016	**0.011**
USA	0.019	0.019	0.012	0.023	0.015	0.013	-	**0.011**

Table 25 ME(e_0) of the forecast methods during out-of-sample evaluation period.

Country	LC	LC_P	LM	BMS	HU	HU_R	HU_W	$LC_{e_0^\dagger}$
Australia	0.007	0.038	0.010	0.207	-0.083	-0.124	0.001	0.054
Austria	0.288	0.195	**0.003**	0.235	-0.585	-1.148	-0.018	-0.033
Belgium	**0.050**	0.110	0.062	0.144	-0.189	-0.194	0.155	0.132
Canada	-0.109	-0.040	-0.014	-0.041	-0.042	-0.116	-0.236	**0.039**
Denmark	-0.955	-1.119	-0.608	-1.067	-0.808	-0.819	-0.727	**-0.575**
Finland	0.456	0.448	-0.016	0.421	-0.581	-0.757	-0.452	**-0.013**
France	0.165	0.157	0.101	0.161	0.055	0.073	**0.048**	0.171
Germany	0.135	0.193	0.183	0.310	**0.003**	-0.414	0.159	0.215
Ireland	**-0.115**	-0.586	-0.452	-0.414	-1.054	-1.307	-0.400	-0.551
Italy	0.657	0.542	0.413	0.551	0.164	**-0.142**	0.379	0.420
Japan	1.232	1.006	0.768	0.956	0.306	**0.115**	0.469	0.674
Netherlands	-0.329	-0.294	-0.220	-0.389	-0.437	-0.625	-0.437	**-0.191**
New Zealand	-0.429	-0.335	-0.293	**-0.111**	-0.309	-0.576	-0.283	-0.249
Norway	0.064	**0.027**	-0.111	-0.103	-0.487	-0.855	-0.569	-0.155
Portugal	**-0.580**	-0.764	-0.795	-0.697	-0.981	-0.836	-0.561	-0.775
Spain	0.347	**-0.046**	-0.053	-0.074	-0.473	-0.824	-0.472	-0.088
Sweden	**0.062**	0.111	0.105	0.063	-0.151	-0.089	-0.114	0.167
Switzerland	0.300	0.233	0.197	0.167	0.095	**0.073**	0.094	0.308
United Kingdom	-0.120	-0.150	-0.158	**-0.082**	-0.669	-0.961	-0.669	-0.118
USA	-0.072	-0.054	-0.037	-0.233	-0.258	-0.174	-	**-0.001**

R-codes for the proposed model

```
library(demography)
library(smoothAPC)
library(forecast)
library(ftsa)


data #demogdata object for observed mortality rates#


years <- data$year

#estimation  of the observed lifespan disparity#
ltjd<-print(lifetable(data, series="female",
max.age = max(data$age)),type = c("period"))

eddgg<-sapply(ltjd, FUN = function(x) sum(x$ex*x$lx*x$mx))

overall_edgr <- cbind.data.frame("years"=years,"edagger"=eddgg)
plot(overall_edgr)

#estimation of fitted lifespan disparity#
```

```
#basic Lee-Carter model is fitted on smoothed mortality data#
# ddata1 is the demogdata object containing pre-smoothed #
# mortality rates using Lasso#

ln.female <- lca(ddata1, max.age= max(ddata1$age),
adjust = "none", interpolate=TRUE)
a1j<-ln.female$ax
b1j<-ln.female$bx
k1j<-ln.female$kt

length(a1j)
length(b1j)
length(k1j)

mx1j<-exp(b1j*k1j[1]+a1j)

# for getting fitted mortality rates from estimated parameters #
# of LC model#
fitmx <- function (kt,ax,bx,transform=FALSE)
{
clogratesfit <- outer(kt, bx)
logratesfit <- sweep(clogratesfit,2,ax,"+")
if(transform)
return(logratesfit)
else
return(exp(logratesfit))
}


wwmxj<-fitmx(k1j,a1j,b1j)
wmxj<-t(wwmxj)
plot(wmxj)


ages1 <- ddata1$age
years1 <- ddata1$year
fDx1 <- matrix(wmxj,
ncol = length(years1),
nrow = length(ages1))
fNx1 <- usa$pop$female

lcfdata1 <- demogdata(data = fDx1, pop = fNx1, ages = ages1,
 years = years1, type = "mortality", label = "label", name = "female")

plot(lcfdata1)
```

```
#extracting component of e-dagger from fitted LC to use later #
fitswed <- print(lifetable(lcfdata1,years = ddata1$year,
ages = ages1,max.age = max(ddata1$age)),type = c("period"))

years<-1950:2016


fitted_edgr <- sapply(fitswed, FUN = function(x) sum(x$ex*x$lx*x$mx))

#extracting lx, ex from life table obtained from fitted LC #
llxx <- matrix(unlist(lapply(fitswed, FUN = function(x) x[, "lx"])),
nrow = length(years), ncol = length(ages1), byrow = TRUE)


eexx <- matrix(unlist(lapply(fitswed, FUN = function(x) x[, "ex"])),
nrow = length(years), ncol = length(ages1), byrow = TRUE)

mmxx <- matrix(unlist(lapply(fitswed, FUN = function(x) x[, "mx"])),
nrow = length(years), ncol = length(ages1), byrow = TRUE)


#the function for lc_edagger#

lcadagger<-function (data, series = names(data$rate)[1],
 years = data$year, ages = data$age, max.age = max(data$age),
 adjust = c("edagger", "none"), chooseperiod = FALSE, minperiod = 20,
 breakmethod = c("bai",
"bms"), scale = FALSE, restype = c("logrates", "rates",
"deaths"), interpolate = TRUE)
{
if (class(data) != "demogdata") {
stop("Not demography data")
}
if (!any(data$type == c("mortality", "fertility"))) {
stop("Neither mortality nor fertility data")
}
is.el <- function(el,set)
{
is.element(toupper(el),toupper(set))
}

# Compute expected age from single year mortality rates
get.e0 <- function(x,agegroup,sex,startage=0)
{
lt(x, startage, agegroup, sex)$ex[1]
}
```

```
# Replace zeros with interpolated values
fill.zero <- function(x,method="constant")
{
tt <- 1:length(x)
zeros <- abs(x) < 1e-9
xx <- x[!zeros]
tt <- tt[!zeros]
x <- stats::approx(tt,xx,1:length(x),method=method,f=0.5,rule=2)
return(x$y)
}

adjust <- match.arg(adjust)
restype <- match.arg(restype)
breakmethod <- match.arg(breakmethod)
data <- extract.ages(data, ages, combine.upper = FALSE)
if (max.age < max(ages))
data <- extract.ages(data, min(ages):max.age, combine.upper = TRUE)
startage <- min(data$age)
get.series <- function(data,series)
{
if(!is.el(series,names(data)))
stop(paste("Series",series,"not found"))
i <- match(toupper(series),toupper(names(data)))
return(as.matrix(data[[i]]))
}

mx <- get.series(data$rate, series)
pop <- get.series(data$pop, series)
startyear <- min(years)
stopyear <- max(years)
if (startyear > max(data$year) | stopyear < min(data$year))
stop("Year not found")
startyear <- max(startyear, min(data$year))
if (!is.null(stopyear))
stopyear <- min(stopyear, max(data$year))
else stopyear <- max(data$year)
id2 <- stats::na.omit(match(startyear:stopyear, data$year))
mx <- mx[, id2]
pop <- pop[, id2]
year <- data$year[id2]
deltat <- year[2] - year[1]
ages <- data$age
n <- length(ages)
m <- sum(id2 > 0)
```

```
edgr<-overall_edgr$edagger
mx <- matrix(mx, nrow = n, ncol = m)
if (interpolate) {
mx[is.na(mx)] <- 0
if (sum(abs(mx) < 1e-09, na.rm = TRUE) > 0) {
warning("Replacing zero values with estimates")
for (i in 1:n) mx[i, ] <- fill.zero(mx[i, ])
}
}
mx <- t(mx)
mx[mx == 0] <- NA
logrates <- log(mx)
pop <- t(pop)
deaths <- pop * mx
ax <- apply(logrates, 2, mean, na.rm = TRUE)
if (sum(ax < -1e+09) > 0)
stop(sprintf("Some %s rates are zero.\n Try reducing the maximum age
 or setting interpolate=TRUE.", data$type))
clogrates <- sweep(logrates, 2, ax)
svd.mx <- svd(clogrates)
sumv <- sum(svd.mx$v[, 1])
bx <- svd.mx$v[, 1]/sumv
kt <- svd.mx$d[1] * svd.mx$u[, 1] * sumv
ktadj <- rep(0, m)
logdeathsadj <- matrix(NA, n, m)
z <- log(t(pop)) + ax
x <- 1:m
ktse <- stats::predict(stats::lm(kt ~ x), se.fit = TRUE)$se.fit
ktse[is.na(ktse)] <- 1
agegroup = ages[4] - ages[3]

edgr<-overall_edgr$edagger

fitmx <- function (kt,ax,bx,transform=FALSE)
{
# Derives mortality rates from kt mortality index,
# per Lee-Carter method
clogratesfit <- outer(kt, bx)
logratesfit <- sweep(clogratesfit,2,ax,"+")
if(transform)
return(logratesfit)
else
return(exp(logratesfit))
}
#to obtaining root, following function from demography package is used#
findroot <- function(FUN,guess,margin,try=1,...)
```

```
{
# First try in successively larger intervals around best guess
for(i in 1:5)
{
rooti <- try(stats::uniroot(FUN,
interval=guess+i*margin/3*c(-1,1),...),silent=TRUE)
if(class(rooti) != "try-error")
return(rooti$root)
}
# No luck. Try really big intervals
rooti <- try(stats::uniroot(FUN,
interval=guess+10*margin*c(-1,1),...),silent=TRUE)
if(class(rooti) != "try-error")
return(rooti$root)

# Still no luck. Try guessing root using quadratic approximation
if(try<3)
{
root <- try(quadroot(FUN,guess,10*margin,...),silent=TRUE)
if(class(root)!="try-error")
return(findroot(FUN,root,margin,try+1,...))
root <- try(quadroot(FUN,guess,20*margin,...),silent=TRUE)
if(class(root)!="try-error")
return(findroot(FUN,root,margin,try+1,...))
}

# Finally try optimization
root <- try(newroot(FUN,guess,...),silent=TRUE)
if(class(root)!="try-error")
return(root)
else
root <- try(newroot(FUN,guess-margin,...),silent=TRUE)
if(class(root)!="try-error")
return(root)
else
root <- try(newroot(FUN,guess+margin,...),silent=TRUE)
if(class(root)!="try-error")
return(root)
else
stop("Unable to find root")
}

quadroot <- function(FUN,guess,margin,...)
{
x1 <- guess-margin
x2 <- guess+margin
```

```
y1 <- FUN(x1,...)
y2 <- FUN(x2,...)
y0 <- FUN(guess,...)
if(is.na(y1) | is.na(y2) | is.na(y0))
stop("Function not defined on interval")
b <- 0.5*(y2-y1)/margin
a <- (0.5*(y1+y2)-y0)/(margin^2)
tmp <- b^2 - 4*a*y0
if(tmp < 0)
stop("No real root")
tmp <- sqrt(tmp)
r1 <- 0.5*(tmp-b)/a
r2 <- 0.5*(-tmp-b)/a
if(abs(r1) < abs(r2))
root <- guess+r1
else
root <- guess+r2
return(root)
}

# Try finding root using minimization
newroot <- function(FUN,guess,...)
{
fred <- function(x,...){(FUN(x,...)^2)}
junk <- stats::nlm(fred,guess,...)
if(abs(junk$minimum)/fred(guess,...) > 1e-6)
warning("No root exists. Returning closest")
return(junk$estimate)
}
if (adjust == "edagger") {


Fundg<-function(p,bx,ax,edgr,llxxi,eexxi){
edgr - sum(exp(ax + bx*p)*llxxi*eexxi)
}
for (i in 1:m) {
if (i == 1)
guess <- kt[1]
else guess <- mean(c(ktadj[i - 1], kt[i]))
ktadj[i] <- findroot(Fundg, guess = guess, margin = 3 *
ktse[i], edgr=edgr[i] , llxxi=llxx[i,],
eexxi=eexx[i,],ax = ax, bx = bx)
logdeathsadj[,i]<-z[,i]+bx*ktadj[i]
}
}
else if (adjust == "none")
```

```
ktadj <- kt
else stop("Unknown adjustment method")
kt <- ktadj

#the following part is needed in case of choosing best fitting#
#period#
if (chooseperiod) {
if (breakmethod == "bai") {
x <- 1:m
bp <- strucchange::breakpoints(kt ~ x)$breakpoints
bp <- bp[bp <= (m - minperiod)]
bestbreak <- max(bp)
return(lca(data, series, year[(bestbreak + 1):m],
ages = ages, max.age = max.age, adjust = adjust,
interpolate = interpolate, chooseperiod = FALSE,
scale = scale))
}
else {
RS <- devlin <- devadd <- numeric(m - 2)
for (i in 1:(m - 2)) {
tmp <- lcadagger(data, series, year[i:m], ages = ages,
max.age = max.age, adjust = adjust, chooseperiod = FALSE,
interpolate = interpolate, scale = scale)
devlin[i] <- tmp$mdev[2]
devadd[i] <- tmp$mdev[1]
RS[i] <- (tmp$mdev[2]/tmp$mdev[1])
}
bestbreak <- order(RS[1:(m - minperiod)])[1] - 1
out <- lcadagger(data, series, year[(bestbreak + 1):m],
ages = ages, max.age = max.age, adjust = adjust,
chooseperiod = FALSE, interpolate = interpolate,
scale = scale)
out$mdevs <- ts(cbind(devlin, devadd, RS), start = startyear,
deltat = deltat)
dimnames(out$mdevs)[[2]] <- c("Mean deviance total",
"Mean deviance base", "Mean deviance ratio")
return(out)
}
}
logfit <- fitmx(kt, ax, bx, transform = TRUE)
if (restype == "logrates") {
fit <- logfit
res <- logrates - fit
}
else if (restype == "rates") {
fit <- exp(logfit)
```

```
res <- exp(lograte) - fit
}
else if (restype == "deaths") {
fit <- exp(logfit) * pop
res <- deaths - fit
}
residuals <- fts(ages, t(res), frequency = 1/deltat,
 start = years[1], xname = "Age", yname = paste("Residuals",
 data$type,
"rate"))
fitted <- fts(ages, t(fit), frequency = 1/deltat, start = years[1],
xname = "Age", yname = paste("Fitted", data$type, "rate"))
names(ax) <- names(bx) <- ages
if (scale) {
avdiffk <- -mean(diff(kt))
bx <- bx * avdiffk
kt <- kt/avdiffk
}
deathsadjfit <- exp(logfit) * pop
drift <- mean(diff(kt))
ktlinfit <- mean(kt) + drift * (1:m - (m + 1)/2)
deathslinfit <- fitmx(ktlinfit, ax, bx, transform = FALSE) *
pop
dflogadd <- (m - 2) * (n - 1)
mdevlogadd <- 2/dflogadd * sum(deaths * log(deaths/deathsadjfit) -
(deaths - deathsadjfit))
dfloglin <- (m - 2) * n
mdevloglin <- 2/dfloglin * sum(deaths * log(deaths/deathslinfit) -
(deaths - deathslinfit))
mdev <- c(mdevlogadd, mdevloglin)
output <- list(label = data$label, age = ages, year = year,
mx = t(mx), ax = ax, bx = bx, kt = ts(kt, start = startyear,
deltat = deltat), residuals = residuals, fitted = fitted,
varprop = svd.mx$d[1]^2/sum(svd.mx$d^2), y = fts(ages,
t(mx), start = years[1], frequency = 1/deltat, xname = "Age",
yname = ifelse(data$type == "mortality", "Mortality",
"Fertility")), mdev = mdev)
names(output)[4] <- series
output$call <- match.call()
names(output$mdev) <- c("Mean deviance base", "Mean deviance total")
output$adjust <- adjust
output$type <- data$type
return(structure(output, class = "lca"))
}
```

```
#to fit the model#

funmod2 <- lcadagger(ddata1, ages=ddata1$age,
max.age = max(ddata1$age), adjust = "edagger", interpolate=TRUE)

# to see the fitted parameters, variation explained#
funmod2

#for forecast t years ahead mortality forecast#
forecast.female <- forecast(funmod2, h=t, jumpchoice = "actual")
plot(forecast.female, main=expression(LC[e[0]^"\u2020"]))


# t years ahead 95% prediction interval for life expectancy #
# at birth #
fun2<- e0(forecast(funmod2, level=95, max.age=max(ddata1$age),
jumpchoice = "actual", h=t), PI=TRUE)

plot(fun2, main=expression(LC[e[0]^"\u2020"]))
```

References

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan disparity as an additional indicator for evaluating mortality forecasts. *Demography*, 54(4):1559–1577.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336.

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

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

Camarda, C. G. et al. (2012). Mortalitysmooth: An r package for smoothing poisson counts with p-splines. *Journal of Statistical Software*, 50(1):1–24.

Currie, I. D., Durban, M., and Eilers, P. H. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):259–280.

De Jong, P. and Tickle, L. (2006). Extending lee–carter mortality forecasting. *Mathematical Population Studies*, 13(1):1–18.

Dokumentov, A., Hyndman, R. J., and Tickle, L. (2018). Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat*, 7(1):e199.

Girosi, F. and King, G. (2006). *Demographic Forecasting*. Cambridge University Press, Cambridge.

HMD (2018). *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).

Hyndman, R. J. and 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. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(4):537–549.

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

Li, N., Lee, R., and Gerland, P. (2013). Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, 50(6):2037–2051.

Schuette, D. R. (1978). A linear programming approach to graduation. *Transactions of Society of Actuaries*.

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

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the g7 countries. *Nature*, 405(6788):789.

Vaupel, J. W. and Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of nathan keyfitz's 90th birthday. *Demography*, 40(2):201–216.

Wood, S. N. (1994). Monotonic smoothing splines fitted by cross validation. *SIAM Journal on Scientific Computing*, 15(5):1126–1133.

Wood, S. N. (2006). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

Zhang, Z. and Vaupel, J. W. (2009). The age separating early deaths from late deaths. *Demographic Research*, 20:721–730.